Essays on Exchange Rate Pass Through

Lu Han

Faculty of Economics
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Downing College

April 2017

I would like to dedicate this thesis to my loving parents, Liping Zhang and Zhiguo Han, and my grandma Xiuying Han.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 60,000 words including appendices, bibliography, footnotes, tables and equations.

Lu Han
April 2017

Acknowledgements

First and foremost, I am deeply indebted to my supervisor Giancarlo Corsetti. There is no doubt that without him none of these three chapters would exist. Two years ago, I secretly started a project (without asking for his permission) that was clearly beyond my ability to manage. The project aims to bridge the gap between existing theoretical and empirical works on exchange rate pass through using machine learning algorithms. Sooner I realised, each progress I had achieved only made my next step even more difficult. There was a long period in which every day I could not stop thinking to stop working on the project. I am thankful to him who saw the potential value of my project and persistently encouraged me to continue my work. I am grateful for his guidance and not giving up on me during the most difficult time of my research period. His attitude towards a myriad of things pertaining not only to academic life but also life in general has inspired me immeasurably. I am eternally grateful for all his help and guidance.

I am extremely grateful to my research advisor and co-author Meredith Crowley. I am thankful for the time she has spent listening to me; teaching me how to clarify my thoughts and problems; helping me to practice and prepare my presentations again and again. I am genuinely motivated by her earnest attitude towards research and thankful for her decisive contributions during our collaboration process. I would like to thank my co-author, Huasheng Song, for discussions on China-related problems and providing the excellent research environment during the period I visited Zhejiang University. I am grateful to the grant provided by Cambridge Endowment for Research in Finance to fund the purchase of the data used in my third chapter.

I am thankful to Tiago Cavalcanti, Petra Geraats, Pontus Rendahl, Coen Teulings and other faculty members for insightful discussions and advice. Especially, I would like to thank Alexey Onatskiy for his suggestions and comments on algorithm related problems in my third chapter.

I would like to thank the support of my fellow PhD colleagues at Cambridge. Especially, my close friend and flatmate Girish Bahal who accompanied me during most of my struggling periods. Thanks go to River Chen, Daniel Goyal, Alice Kuegler, Frederico Lima, Kai Liu, Simon Lloyd, Zhaoyang Liu, Anand Shrivastava, Jonathan Smith, Riccardo Trezzi, and Chen Wang for their generous help and valuable comments during the period of my PhD study. Special thanks for delicious chicken soups from Haihan Tang and instant noodles from Jasmine Xiao.

I am grateful to my undergraduate mentor Michael Ben-Gad without whose encouragement I would have never started my PhD journey in economics.

Last, I would like to thank my parents for all their unconditional love and encouragement, spiritual and financial supports.

Abstract

This dissertation contributes to the theoretical and empirical understandings of international transmissions of exchange rate shocks. It consists of three chapters.

The first chapter extends Corsetti and Dedola (2005) and further allows for competitions in retail networks. In the model, there are four types of firms interacting with each other including retailing manufacturers, non-retailing manufacturers, specialised retailers and nontradable good producers. The equilibrium depends on the interaction among these four types of firms, which leads to a dynamic and incomplete exchange rate pass through (ERPT) depending on the firms' share of retail networks. With the standard calibration, the model can generate a high (4-5) long-run trade elasticity without conflicting with a low (0.5-1) short-run elasticity, suggesting that the dynamics of retail networks offer a potential explanation of the trade elasticity puzzle.

Chapter 2 investigates the ERPT of Chinese exporters. We propose an estimator that utilises orthogonal dimensions to control for unobserved marginal costs and estimate destination specific markup adjustments to bilateral and multilateral exchange rate shocks. Our estimates suggest that the cost channel accounts for roughly 50% of conventional EPRT estimates. We offer new channels of heterogeneity in firms' pricing behaviour and provide supporting evidence on the international pricing system.

Chapter 3 aims to bridge the gap between theoretical and empirical works on ERPT. I propose a machine learning algorithm that systematically detects the determinants of ERPT. The proposed algorithm is designed to work directly with highly disaggregated firm-level customs trade databases as well as publicly available commodity trade flow datasets. Tested on the simulated data from a realistic micro-founded multi-country trade model, my algorithm is proven to have accuracies around 95% and 80% in simple and complex scenarios respectively. Applying the algorithm to China's customs data from 2000 to 2006, I document new evidence on the nonlinear relationships among market structures, unit value volatility and ERPT.

Table of contents

List of figures

List of tables

Chapter 1

Optimal Price Setting and Exchange Rate Pass Through - The Role of the Retail Capacity

This paper studies the role played by the dynamics of retail capacity in determining exchange rate pass through (ERPT), trade elasticity and volatility of import price. As shown by Corsetti and Dedola (2005), local cost components in the form of distribution may contribute to explaining incomplete pass through, as they tend to make the trade elasticity depend on the relative distribution margin. I model distribution as an investment in retail capacity, inducing a difference between short and long run demand elasticities. The slow adjustment of retail capacity offers a possible explanation for the "international trade elasticity puzzle". The model can generate a high (4-5) long-run trade elasticity without conflicting with a low (0.5-1) short-run elasticity.

1.1 Introduction

Two important puzzles have been shaping the debate in open macro in recent years. One is the so-called elasticity puzzle: macro and short-run estimates tend to show that trade elasticities are very low.[1] The other one concerns the determinants of the local currency price stability of imports, i.e., the low degree of exchange rate pass through. In this paper, I underline the importance of the dynamics of retail capacity for understanding exchange rate pass through and trade elasticity and show how these two puzzles can be explained in a unified framework.

This paper builds on Corsetti and Dedola (2005) and models the local component as the result of optimal investment decisions on building local retail networks. I consider a market structure with three types of producers and a single type of distributing firms. The first type of producers, denoted as "retailing manufacturers", sell directly to home and foreign consumers by building their own local retail networks. The second type of producers, denoted as "non-retailing manufacturers", sell their products to local retailers. The third type of producers specialise in local nontradable goods and no distribution is needed. Local distributors, denoted as "local retailers", buy a range of home and foreign tradable goods, set the retail price for each product and adjust their investment in retail networks according to all products they sell.

The perceived demand elasticity is firm-specific, depending on a firm's share of retail networks. A larger share enables the firm to access a larger consumer base, reduce its per unit distribution cost and lower its distribution margin. By selling to local retailers, non-retailing manufacturers benefit from the large retail network accumulated by local retailers at the cost of the additional markup charged by local retailers. Non-retailing manufacturers are not in control of the retail network. They price their products taking decisions made by local retailers as given, resulting in a high ERPT. Retailing manufacturers internalise their investments of retail networks. These firms have two means to boost their sales. At the equilibrium, they trade off between a more costly method which leads to a long run gain in sales by investing in their retail networks and a less costly method by temporarily cutting their prices. At their optimal choice, ERPT is relatively low compared to that for non-retailing manufacturers. Aggregate ERPT and trade elasticity depend on firms' share of retail networks and the corresponding vertical market structure. In this aspect, my model

[1] The quantitative analyses of international real business cycle models and international trade models have made significant progress over the last two decades. Nonetheless, the trade literature [e.g. Eaton and Kortum (2002) and Alvarez and Lucas (2007)] needs to set a high Armington (1969) elasticity of substitution between imports and home produced tradable goods to capture the long-run adjustment of intensive and extensive margins. In contrast, international real business cycle models, Backus, Kehoe, and Kydland (1994) and Heathcote and Perri (2002) for example, need a small elasticity to generate observed volatility of terms of trade and the negative correlation between trade balance and terms of trade. This tension is referred to as the "international elasticity puzzle".

contributes to the literature on how vertical or horizontal market structures affect prices and ERPT[2] by adding a novel aspect on how firms compete for retail resources.

The difference between short-run and long-run trade elasticities can be explained through two channels. The first channel is similar to the mechanism of Drozd and Nosal (2012)[3] where changes in consumption are restricted by the firm's retail capacity in the short run, and trade elasticity goes up as this restriction is released gradually in the long run. All quantity adjustments are driven by price changes in the short run, while in the long run the retail capacity extends gradually and the relative consumption shifts more. In the second channel, changes in retail capacity influence the per unit cost to distribute a product and alter the wedge between consumer and producer prices (the distribution margin), resulting in dynamics of the firm's elasticity of demand. In contrast to intuitive judgements, the model shows that the short-run difference in demand elasticity between home and foreign products is larger with the second channel. As the retail capacity adjusts in the long run, this difference becomes smaller, and the trade elasticity goes up gradually. Combining these two channels, the model can generate a high (4-5) long-run elasticity of substitution between home and foreign products without conflicting a the low (0.5-1) short-run elasticity. While most trade literature attributes the difference between short-run and long-run trade elasticities to extensive margin movements, my model emphasises the rule of intensive margins.

As an extension of the analysis, I investigate the role of retail capacity in explaining the connection between price volatility and ERPT. Empirical findings by Berger and Vavra (2013) document a positive relationship between price volatility and pass through[4]. I show analytically and quantitatively that this empirical pattern is consistent with differences in retail capacity and distribution margin across firms. In the short run where the retail capacity fails to adjust, ERPT of a firm is positively correlated with the volatility of its import price if (a) the size of idiosyncratic shocks is large and (b) its distribution margin is low. An increase in average distribution margin reduces the correlation between price volatility and ERPT. This correlation also depends on whether the exporter is in control of its local retail network. Simulation results suggest that the change of distribution margin contributes little to the

[2]Section 1.2 reviews the related literature.

[3]Drozd and Nosal (2012) provide a possible explanation of the elasticity puzzle based on slow adjustment of consumer list. Due to the adjustment cost on the investment of the market share, the producer cannot raise their market share to the steady state level immediately and the consumer list thus does not update much. This gives a low trade elasticity in the short run because the change of relative price is high and that of relative retail quantity is low. In the long run, the consumer list adjusts and trade elasticity increases. However, in their model, the consumer demand is always equal to the consumer list holding by the producer and the change of prices has no influence on consumer's choice. The equilibrium retail price is determined through the inverse demand function once the demand is chosen. My model differs from theirs in the sense that the consumer demand in my model is jointly determined by the producer's retail price and its retail capacity. The producer can attract more consumers through either a lower price or marketing activities that extend its retail capacity.

[4]New empirical findings on the relationship among firm level characteristics, market structure, price volatility and ERPT are presented in my third chapter.

variance of the price for those firms that are not in control of their local retail networks. In contrast, the variance of retail capacity explains 40% of price volatility for those firms directly controlling their retail networks. Furthermore, aggregate shocks are important in explaining the price volatility of retailing manufacturers but not that of non-retailing manufacturers.

The rest of the paper is organised as follows. Section 1.2 provides a brief literature review on research modelling local component and strategic interactions. Section 1.3 presents new empirical evidences on distribution margin and ERPT. Section 1.4 introduces the basic model. Section 1.5 extends the basic model and discusses the role of retail capacity in explaining the link among trade elasticity, price volatility and ERPT. Section 1.6 concludes.

1.2 Literature Review

Discussion on international relative prices using two-country open-economy DSGE models can be traced back to the early 1990s. At that time, a large number of international macroeconomic models imply the law of one price holds for tradable goods sold across countries and ERPT is perfect [e.g. Backus et al. (1992), Obstfeld and Rogoff (1995) and Stockman and Tesar (1995)]. However, the increasing evidence suggesting the failure of the law of one price and the purchasing power parity gives rise to a new group of studies characterising the "pricing to market" behaviour of firms[5]. The term “pricing to market” was first used in Krugman (1986) to indicate that markets are segmented and firms discriminate and charge different markups across countries[6]. In line with Krugman (1986), a large number of literature emphasises the pricing to market behaviour and attributes the incomplete pass through to local nominal rigidities [e.g. Betts and Devereux (2000)]. However, to generate the observed low pass through, they need to set a very high degree of nominal rigidities. Moreover, their models give wrong implications of the relationship between the terms of trade and the exchange rate as critiqued by Obstfeld and Rogoff (2000).

Later studies started to consider reasons apart from the nominal rigidity. These studies can be mainly divided into three groups. The first group emphasises the importance of the local component of consumer price.[7] The existence of the local component drives a wedge between consumer and producer prices, making the elasticity of demand an increasing function of the producer price and resulting in incomplete EPRT at both producer and consumer prices. The second group focuses on variable demand elasticities and markups under different market structures. The difference in the price elasticity of demand across countries motivates producers to charge different markups. The firm’s optimal markup

[5] Rogoff (1996) provides a very good literature survey on the failure of the law of one price and the purchasing power parity.

[6] The segmentation of markets may come from various reasons which imply that the cost of arbitrage is not zero. For example, in Goldberg and Verboven (2005)’s survey on the car market, manufacturers prevent local dealers from exporting the car to other countries by threatening to withdraw their franchise.

[7] These studies can also be viewed as literature characterising vertical interactions.

often depends on its horizontal or vertical interaction with other firms. As a result, the import price changes less than one to one with the change in the firm's marginal cost and the nominal exchange rate. The third group studies the exchange rate pass through at the border and has developed models in which exporters can choose whether to price their exports in home currency or in foreign currency, knowing that price updates will be subject to certain frictions[8]. A number of factors such as the local monetary policy stability, the market share of exporters and the exchange rate regime play a crucial role in this choice.

The importance of local components to price discrimination across markets was first pointed out by Corsetti and Dedola (2005). Intuitively, local components produce a wedge between the consumer price and the import price at the border and reduce the sensitivity of consumption to the producer price at the consumer level. The price elasticity of demand with respect to the producer price is a function of distribution margin. As different distribution margins imply distinct demand elasticities across markets, it is optimal for producers to discriminate and charge different prices across countries. An increase in the wholesale price of the firm lowers the distribution margin and increases the demand elasticity. As a result, the desired markup narrows. The optimal wholesale price moves less than one to one in response to a exchange rate shock and ERPT is incomplete. The observation that ERPT at consumer prices is lower compared to ERPT at import prices can be easily explained using this setting. Therefore, models with local components can generate observed low ERPT without assuming huge local nominal rigidities. In addition to incomplete pass through, Corsetti et al. (2008a) based on this setting can generate high volatility of real exchange rates and low volatility of terms of trade relative to real exchange rates, which standard business cycle models fail to capture.

This explanation is highly appreciated by empirical studies. Goldberg and Verboven (2001) conduct a empirical analysis for automobile retail prices in five European countries. They attribute around 38% pass through of the nominal exchange rate to local distribution cost. Hellerstein (2008) builds a structural model to fit panel data in beer industry. He finds the optimal markup adjustment of producer and the existence of local distribution cost equally explain the price stickiness. He argues that foreign exporters bear a greater cost in response to an appreciation of home exchange rate compared to local producers and local retailers.

The intuition of the local component of price can be applied to other contexts. For example, Alessandria (2009) develops a consumer searching model where the searching process is costly but gives the possibility of buying goods at a lower price. The real consumer price can be viewed as producer price with an additive component, i.e. the corresponding searching cost. Following the same logic, a higher searching cost leads to a higher markup and ERPT is incomplete due to the additive term to producer prices. Similarly, Drozd and

[8] Since my theoretical models are not directly linked to the optimal currency choice, the detailed literature review of this group is omitted in the literature review section.

Nosal (2012) assume that producers and local retailers conduct a Nash bargaining with total profit of sales. As a result, exporter's wholesale price P_x denoted in the home currency equals a local component $\eta \varepsilon P_c^*$ plus a portion of the marginal cost where η is the fixed bargaining power for the local retailer.

Research modelling different market structures and ERPT can be categorised into horizontal and vertical interactions. In terms of horizontal interactions, Dornbusch (1987) explains the incomplete ERPT by considering oligopolistic markets in which the optimal adjustment of the markup depends on market structures and the underlying curvature of the demand curve. Atkeson and Burstein (2007) present a simple Ricardian model of international trade where exporters charge a price equal to the marginal cost of their local competitors and thus the ERPT is incomplete. De Blas and Russ (2015) generalise Atkeson and Burstein (2007)'s Bertrand competition setting with multi-countries in which the degree of the price rigidity and the incomplete ERPT depend on the distribution of markups which is in turn determined by the number of competitors.

Among the literature studying horizontal interactions, Atkeson and Burstein (2008) appears to be a very successful explanation. They extend Dornbusch (1987) to capture the fluctuations in the international relative prices. They consider a market with a continuum of sectors and a finite number of differentiated products under each sector. The elasticity of demand across sectors is assumed to be smaller than the elasticity within the same sector. The firm chooses its retail price to maximise its profit subject to the inverse demand function and takes into account that the sector demand will be affected by its price. The demand elasticity is decreasing in the market share. Therefore, a firm with a high market share in its sector assigns a higher weight to the low substitutable competitors across industries and thus has a lower elasticity. Correspondingly, the optimal markup is increasing in the market share. In addition, any price changes of a firm in the industry will change the aggregate industry price and its market share. The elasticity of demand alters accordingly and ERPT is incomplete. They argue that pricing to market behaviour is heterogeneous across firms. The within-sector cost dispersion is central in their paper in the sense that it pins down the distribution of markups and thus the pricing behaviour. In the equilibrium, only large firms choose pricing to market.

Amiti et al. (2014) confirm this result using the data of Belgian exporters. They find that exporters with larger market share in the destination market have a lower pass through. Moreover, a higher import intensity of inputs lowers the response to exchange rate shocks. That is, shocks of exchange rates are partially offset by the adverse adjustment of marginal cost due to the change of input prices. They find that the distribution of importers is quite skewed in the sense that large importers are also large exporters. The effect of a depreciation of home currency with respect to all trade partners on export prices is partially offset by the increase in the cost of buying inputs for exporters. They construct a model by combining

Atkeson and Burstein (2008) with Halpern et al. (2011) and attribute half of incomplete ERPT to varying markups with the change of the marginal cost accounting for another half.

Corsetti et al. (2007) explicitly model strategic vertical interactions among upstream and downstream firms, explore the possible interactions between optimal price setting and nominal rigidities and study their implications on optimal monetary policy. In their model, upstream producers exercise their monopoly power and set different prices for downstream retailers at home and abroad. Local monopolistic downstream firms using one intermediate traded good to produce nontradable final goods then sell the good to the consumer. The nominal rigidity is modelled using Calvo pricing, where only a fraction of upstream producers and downstream retailers can change their price. Therefore, downstream retailers face different marginal costs depending on whether their upstream producer changes price. In addition, upstream producers updating their price will need to consider that only a fraction of downstream firms buying their products will also reoptimise in the same period.

The optimal price depends explicitly on the demand elasticity. The perceived demand elasticity is market-specific, depending on differences in industry-specific inflation rates and the degree of price dispersions in the local market. Therefore the deviation from law of one price comes not only from the nominal rigidity but also from the vertical interaction among upstream and downstream monopolists. Furthermore, the vertical interaction among firms with sticky prices lowers the demand elasticity. Nominal rigidities at the retail level do not necessarily lower the equilibrium reaction of final prices to exchange rate movements due to the strategic substitutability between upstream and downstream firms. They show analytically that upstream nominal rigidities lead to a lower short-run ERPT irrespective of vertical interactions. Nevertheless, downstream nominal rigidities induce a larger price response to exchange rate changes because of strategic substitutability. Under reasonable calibration, the effect of upstream nominal rigidities dominates the effect of downstream ones and the pass through is incomplete.

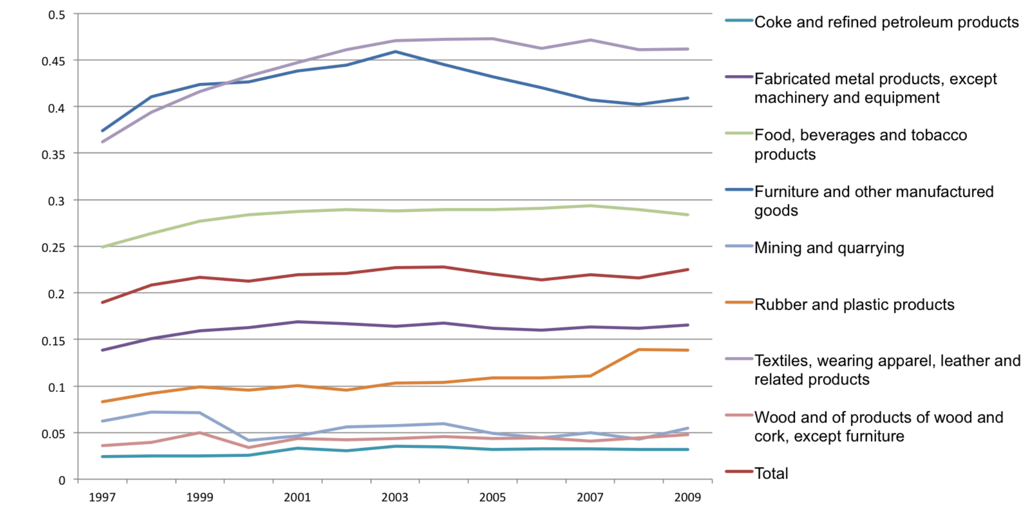
1.3 Empirical Evidence on Distribution Margin

As an increasing number of papers draw attention to the importance of distribution cost in interpreting EPRT, understanding the property of this local component is vital to the analysis. This section provide new evidence on distribution margin across industries and countries.

The distribution margin is estimated based on the Supply Table at current prices of the National Accounts from the Eurostat database. The data is available at the industry level at annual frequency from 1995 to 2010 for most European countries. However, this database is largely incomplete such that the relevant data for distribution margin are unavailable for most countries during the reporting periods with the exception of the year 2008. Industries are categorised by classification of products by activities (CPA) system. 65 industries including 22 goods sectors and 43 service sectors are reported in the dataset.

The dataset reports, among others, "the distribution margin and trade cost" and "the total supply at purchasers' prices". Following Goldberg and Campa (2010), the distribution margin is calculated as "the distribution margin and trade cost" divided by "the total supply at purchasers' prices". The calculated distribution margin for the UK from 1997 to 2010 is shown in Table A in the appendix. As expected, the calculated distribution margin for service sectors is close to zero. Thus, only distribution margins for 22 goods sectors are presented[9].

Fig. 1.1 Estimated distribution margins of selected industries in the United Kingdom

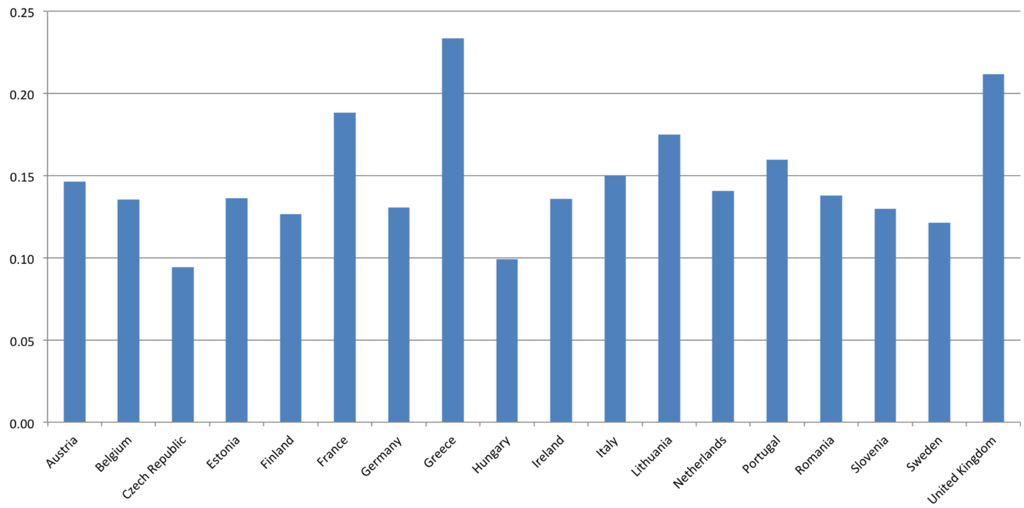


Source: Own Calculations. Data from the "Supply Table at Current Price" under the National Account in the EUROSTAT database. Industries are categorized under the statistical classification of products by activity (CPA).

As can be seen from figure 1.1, there is strong evidence on heterogeneity in distribution margins across industries. Final goods show relative bigger distribution margins compared to intermediate products and raw materials. Furnitures, textiles and wearing apparels have a high distribution margin which ranges from 0.36 to 0.47. Food, beverages and tobacco products have a distribution margin slightly above the average (denoted with "Total"). Basic energy products, such as coke and refined petroleum products, and mining and quarrying products have a relatively low distribution margin. Although the distribution share differs substantially across industries, the annual average distribution margin within the industry is relatively stable with only a small increase trend during the period from 1997 to 2010. The aggregate distribution share starts from 18.9% in 1997 and reaches 22.4% in 2010.

[9]The aggregate distribution share (denoted as "Total") is calculated by dividing the sum of trade and distribution costs for 22 goods sectors by the sum of their total supply at purchasers' prices.

Fig. 1.2 Estimated aggregate distribution margins across countries in 2008



Source: Own Calculations. Data from the "Supply Table at Current Price" under the National Account in the EUROSTAT database. Industries are categorized under the statistical classification of products by activity (CPA).

The estimated distribution margin for 18 European countries in 2008 are presented in Table B in the appendix. Figure 1.2 visualises the aggregate distribution margin in these countries. The estimated distribution margins of these 18 countries show strong heterogeneity across industries. A more important finding is that even in the same industry, there exists huge differences in distribution margins across countries. For instance, the most astonishing distribution margin difference lies in fish and fishing products, where the lowest distribution margin is 6.79% in Germany and highest distribution margin is as high as 61.00% in Romania. In addition, the magnitude of the difference in the industrial distribution margin across countries varies for different industries. The distribution margin difference (highest minus lowest) for food, beverages and tobacco products is only 12.92%, which is much smaller than that for fish and fishing products. The third observation is that the difference of distribution margin at the aggregate level is smaller across countries compared to the industry level. As shown in figure 1.2, the aggregate distribution margin of most countries lies in the range from 10% to 20%. The difference between the highest aggregate distribution margin 23.37% (Greece) and the lowest 9.43% (Czech Republic) is only 12.94%.

The differences in distribution margins may be caused by various reasons. According to the data from the Supply Table, the costs incurred by the distribution sector could be divided into two parts, namely the transportation cost to the retail store and the cost involved in operating and sustaining the retail network. Distribution costs may differ for various products due to distinct inherent qualities and even for the same product due to different geographies and business structures across countries. The operating cost may differ due to country and industry specific situations, such as the related tax rate and economic conditions.

In order to investigate the variance and correlation of distribution margin cross industries, series of monthly prices are needed. Since the Input-Output table is available on an annual basis, I use sector retailer price margin series from U.S. Bureau of Labour Statistics.[10] The price margin is calculated as the current selling price minus the current acquisition price. Table 1.C.5 in the appendix presents variance calculated from percentage change of the price margin index. These variances demonstrate a high degree of heterogeneity. The volatility of the aggregated retail margins (0.0019) is much smaller than that for individual sectors. However, the across sector covariance matrix of price margin changes does not show any clear pattern.

To sum up, three patterns can be extracted from data. First, distribution margins differ greatly across industries within a country. Second, while the difference in distribution margin is relatively large for the same industry across countries, this difference is much smaller at the aggregate level. Third, the variance of the retail price margin shows high degree of heterogeneity.

1.3.1 The empirical relationship between ERPT and distribution margin

In this subsection, I provide a preliminary empirical investigation on the relationship between distribution margin and ERPT.

I run regressions to estimate ERPT across countries and industries following the regression specification of Goldberg and Campa (2005), Goldberg and Campa (2006) and Mumtaz et al. (2008):

$$\triangle \log(MP_t^{i,d}) = C^{i,d} + \beta^{i,d} \triangle \log(NER_t^{i,d}) + \gamma^{i,d} \triangle \log(MC_t^{*,i,d}) + \sum_{\iota=1}^{3} \chi_{\iota}^{i,d} \triangle \log(MP_{t-\iota}^{i,d}) + \varepsilon_t^{i,d}$$

where MP denotes the import price, NER denotes the bilateral nominal exchange rate, MC denotes proxies for the marginal cost, and the superscripts i,d represent the industry and destination countries respectively. I estimate sectoral level bilateral ERPT of imports from the United Kingdom to 10 European countries including Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Portugal, Spain and Sweden.

The import prices are approximated by monthly industrial level bilateral import unit value index[11] from the United Kingdom to the destination country. The marginal costs are approximated by the sectoral export unit value index from the destination country to the United Kingdom[12]. Data are available at a monthly frequency from January 1995 to

[10] US BLS started to release monthly sector level of retailer price indices from Jan 2009. Please see the following link for detailed description: http://www.bls.gov/opub/btn/volume-1/pdf/wholesale-and-retail-producer-price-indexes-margin-prices.pdf

[11] The unit value index is less preferable than the price index. However, due to data availability, the unit value index is often used in empirical estimations [see Goldberg and Campa (2006)].

[12] Empirical ERPT studies face problems of finding a good proxy for marginal cost. My second chapter discusses this issue and provides a sequential fixed effect estimator to address this problem.

December 2004 at COMEXT (Eurostat). The regression sample periods range from 1995:01 to 2001:12 for all countries except for Sweden where the full sample period (from 1995:01 to 2004:12) is used.[13]

The parameter for short-run ERPT is given by $\beta^{i,d}$ and the long-run ERPT can be calculated as $\frac{\beta^{i,d}}{1-\sum_{l=1}^{3}\chi_l^{i,d}}$. Short-run ERPT estimates are reported in Table C in the appendix. Among 200 short-run ERPT estimates, 73 and 50 estimates reject the t-test with null hypothesis of zero at 10% and 5% significance level respectively. The adjusted R-squared ranges from 0.097 to 0.785 with a large proportion of adjusted R-squared lying in the range between 0.2 and 0.4.

Third, using the estimated short-run ERPT and estimated distribution margin, I run simple OLS regressions to examine their relationship. The relevant distribution margin data used are estimated distribution margins across countries in 2008. The regression equation is specified as below:

$$ERPT_{i,d} = \underset{(0.070757)}{0.443983} - \underset{(0.394291)}{0.886} * DM_{i,d} + \varepsilon_{i,d}$$

or

$$ERPT_{i,d} = \underset{(0.111251)}{0.103349} - \underset{(0.052347)}{0.099} * \log(DM_{i,d}) + \varepsilon_{i,d}$$

After matching the estimated ERPT with the distribution data available, 143 observations are used in the above regressions. Both equations give a significant negative relationship between ERPT and the distribution margin. The estimated coefficients on distribution margin for the first and second equation reject the null hypothesis of zero at 5% and 10% significance level respectively. Standard errors are presented in parentheses.

Admittedly, these two regressions are free from measurement error and omitted variable bias only under the following three conditions. First, the industry-level distribution margin is country specific and not influenced by other variables such as volatility of exchange rate and price levels. Second, the industrial distribution margin in the year 2008 represents the average distribution margin in the period from January 1995 to December 2001[14]. Third, distribution margin does not differ between import goods and local produced goods[15].

As a robustness check, I calculate the correlation between the distribution margin and ERPT using ERPT estimations of Goldberg and Campa (2005) and Goldberg and Campa (2006). Goldberg and Campa (2005) estimate ERPT for goods of five categories namely food, energy, raw materials, manufacturing and non-manufacturing using quarterly data for 23 OECD countries from 1975Q1 to 2003Q4. Goldberg and Campa (2006) report ERPT

[13] Augmented Dickey Fuller tests with a time trend are performed for time series of each country and each industry in the sample. At the 5% significance level, I cannot reject the null hypothesis of the existence of a unit root for 290 of total 370 series including the import prices, bilateral exchange rates and proxies of marginal costs. The number of lags for import prices is determined by the Akaike Information Criterion.

[14] As shown in figure 1.2, the distribution margin at industry level is relatively stable over time. Even if there is a measurement error, this error is unlikely to be correlated with ERPT.

[15] As the imported goods are distributed using the local distribution sector, this is likely to be true. However, I haven't found relevant empirical studies for reference.

for products classified by 1-digit Standard International Trade Classification (SITC) during the period from 1989:m1 to 2001:m3. I reconcile the distribution margin data available to fit their classifications[16] and find negative correlations for both pairs with –0.111 and –0.068 respectively.[17]

1.4 The Basic Model

This section uses a simplified model to introduce my key settings and explain the two channels driving the difference between short run and long run trade elasticities. The world economy consists of two countries, home and foreign, of the same size. The representative household in the home country consumes three types of goods, namely home tradable goods $C(h)$, foreign tradable goods $C(f)$, and nontradable goods $C(n)$. I restrict my analysis to two types of firms, namely retailing manufacturers and non-tradable goods producers. A richer and more realistic model is studied in section 1.5.

I model the building of retail networks as an investment in the retail capacity k and embed the effect of changing retail networks through the following CES aggregator, i.e.,

$$C_{H,t} \equiv \left[\int_0^1 \left(\frac{k_t(h)}{K_{H,t}} \right)^{\frac{1}{\theta_H \theta_K}} C_t(h)^{\frac{\theta_H - 1}{\theta_H}} dh \right]^{\frac{\theta_H}{\theta_H - 1}}, \quad C_{F,t} \equiv \left[\int_0^1 \left(\frac{k_t(f)}{K_{F,t}} \right)^{\frac{1}{\theta_H \theta_K}} C_t(f)^{\frac{\theta_H - 1}{\theta_H}} df \right]^{\frac{\theta_H}{\theta_H - 1}}$$

$$K_{H,t} \equiv \left[\int_0^1 k_t(h)^{\frac{1}{\theta_K}} dh \right]^{\theta_K}, \quad K_{F,t} \equiv \left[\int_0^1 k_t(f)^{\frac{1}{\theta_K}} df \right]^{\theta_K}$$

where θ_H is elasticity of substitution among home tradable goods and the elasticity of substitution among foreign tradable goods is assumed to be the same as its home counterpart, i.e., $\theta_F = \theta_H$. θ_K reflects the intensity and effectiveness of retail capacity accumulation on consumer's preference. If θ_K is equal to unity, aggregate consumption is just a capacity weighted average of consumption of single products. A higher θ_K indicates that the marginal return of consumption demand in investing retail capacity is low.

The idea of this setting is that a firm should be able to affect consumer's demand through a channel other than price. The investment in the retail network can be viewed as a proxy for all local promotion efforts which increase the demand without affecting the retail price and quality of the product. For example, the producer can advertise their product, which affects the consumer's preference *ex ante*[18]. A higher retail capacity can be interpreted as more retail

[16] The reconciliation of Goldberg and Campa (2005) follows their "Appendix Table 3: Industry Names". Due to the availability of distribution data, only 12 countries are left in the dataset. The concordance between SITC and CPA for Goldberg and Campa (2006)'s estimations is done by using the Combined Nomenclature (CN) reported by Eurostat.

[17] 49 and 130 observations are used in the first and second calculation respectively.

[18] Many empirical marketing studies have made it clear that advertising and marketing can change the consumer's preference on top of the price and quality of the product. See Baker, Hutchinson, Moore, and Nedungadi (1986) and Borzekowski and Robinson (2001) for examples.

stores, which makes the product more accessible and reduces the consumer's searching cost. A higher retail capacity may be viewed as a strategy to increase the effectiveness of matching. In the context of Alessandria (2009), it increases the consumer's probability of getting a price quote for the firm. After the purchase of the product, it could be an investment that improves customer services, which affects the demand of the product next period for a reason similar to the deep habit by Ravn, Schmitt-Grohé, and Uribe (2006). Mechanically, it provides a simple setting which bypasses the difficulty of using inequality constraints to restrict the demand below the retail capacity.

The aggregation between home and foreign goods can be defined as usual:

$$C_{T,t} \equiv \left[(S_H)^{\frac{1}{\rho}} C_{H,t}^{\frac{\rho-1}{\rho}} + (S_F)^{\frac{1}{\rho}} C_{F,t}^{\frac{\rho-1}{\rho}} \right]^{\frac{\rho}{\rho-1}}$$

where ρ is the elasticity of substitution between home and foreign tradable goods, and the shares of the retail network $(S_{H,t}, S_{F,t})$ are given by

$$S_{H,t} = \frac{(K_{H,t})^{\frac{1}{\theta_K}}}{(K_{H,t})^{\frac{1}{\theta_K}} + (K_{F,t})^{\frac{1}{\theta_K}}}, \quad S_{F,t} = \frac{(K_{F,t})^{\frac{1}{\theta_K}}}{(K_{H,t})^{\frac{1}{\theta_K}} + (K_{F,t})^{\frac{1}{\theta_K}}}$$

With the retail network share in the CES aggregator, the change in demand of home tradables, $C_{H,t} = S_{H,t} \left(\frac{P_{H,t}}{P_t} \right)^{-\rho} C_{T,t}$, is driven by two forces. In addition to the conventional mechanism via price, the change in the retail network share plays a nontrivial role. The consumer's demand from the CES aggregation without the retail network share, $\left(\frac{P_t(h)}{P_{H,t}} \right)^{-\theta_H} C_{H,t}$, can be regarded as the potential demand. Without any retail capacity ($S_{H,t} = 0$), the producer cannot sell anything to the consumer. The classical market share measure, which is defined as $\frac{P_H C_H}{P_T C_T}$, can be viewed as a weighted average between the retail network share and the conventional price driven market share.

I take the conventional setting that nontradable goods do not need to go through the distribution process and producers sell directly to buyers.

$$C_{N,t} \equiv \left[\int_0^1 C_t(f)^{\frac{\theta_N - 1}{\theta_N}} dh \right]^{\frac{\theta_N}{\theta_N - 1}}, \quad C_t \equiv \left(C_{T,t}^{\frac{\phi-1}{\phi}} + C_{N,t}^{\frac{\phi-1}{\phi}} \right)^{\frac{\phi}{\phi-1}}$$

The corresponding price aggregators can be derived as

$$P_{H,t} = \left[\int_0^1 \left(\frac{k_t(h)}{K_{H,t}} \right)^{\frac{1}{\theta_K}} p_t(h)^{1-\theta_H} dh \right]^{\frac{1}{1-\theta_H}}, \quad P_{F,t} = \left[\int_0^1 \left(\frac{k_t(f)}{K_{F,t}} \right)^{\frac{1}{\theta_K}} p_t(f)^{1-\theta_H} df \right]^{\frac{1}{1-\theta_H}}$$

$$P_{T,t} = \left[S_{H,t} P_{H,t}^{1-\rho} + S_{F,t} P_{F,t}^{1-\rho} \right]^{\frac{1}{1-\rho}}, \quad P_t = \left[P_{T,t}^{1-\phi} + P_{N,t}^{1-\phi} \right]^{\frac{1}{1-\phi}}$$

1.4.1 The exporter's problem

The production function of producers is assumed to be linear in labour L and productivity Z, i.e., $Y = ZL$. As this production function has constant returns to scale, the pricing problem of the producer of home tradables for the domestic and foreign markets can be analysed separately.

In the foreign market, the home producer h chooses its price denominated in local currency $P_t^*(h)$ and its investment in the local retail network $i_t^*(h)$ to maximise its expected total profit subject to its marginal cost $MC_{H,t}$, its demand given by the CES aggregator $D_t^*(h)$, and the accumulation process of retail capacity (1.3).[19]

$$\max_{P_t^*(h), i_t^*(h)} \quad E_t \sum_{\iota=t}^{\infty} Q_{t,t+\iota} \left\{ \left[\varepsilon_\iota P_\iota^*(h) - MC_{H,\iota} \right] D_\iota^*(h) - \left[i_\iota^*(h) \right] \varepsilon_\iota P_{N,\iota}^* \right\}$$

subject to

$$D_t^*(h) = \left(\frac{k_t^*(h)}{K_{H,t}^*} \right)^{\frac{1}{\theta_K}} \left(\frac{P_t^*(h)}{P_{H,t}^*} \right)^{-\theta_H} C_{H,t}^* \tag{1.1}$$

$$MC_{H,t} = \frac{W_t}{Z_{H,t}} \tag{1.2}$$

$$k_{t+1}^*(h) = i_t^*(h) + (1-\delta) k_t^*(h) - \chi\left(k_t^*(h)\right) D_t^*(h) \tag{1.3}$$

The component $\chi(k_t^*(h)) D_t^*(h)$ in equation (1.3) is the second key feature of my model. The idea is to model the effectiveness of distribution as a function of existing retail capacity. For each product being sold, $\chi(k_t^*(h))$ units of nontradable goods are required to distribute the product to the consumer. The function $\chi(k) : [0,\infty) \to [0,\infty)$ measures the efficiency of distribution and is assumed to be decreasing in the retail capacity at a decreasing rate.[20] This setting penalises a firm that attempts to temporarily sell at a quantity above its retail capacity. Instead of shutting down the price channel in the short run and letting the demand be completely determined by the retail capacity as in Drozd and Nosal (2012), my setting enables a short-run boost in sales by cutting prices and leaves the firm to trade off between the high cost on distribution and the short-run boost in sales.

Substituting the investment constraint into the objective function, the optimisation problem can be rewritten as:

$$\max_{P_t^*(h), k_{t+1}^*} \quad E_t \sum_{\iota=t}^{\infty} Q_{t,t+\iota} \left\{ \begin{array}{c} \left[\varepsilon_\iota P_\iota^*(h) - MC_{H,\iota} - \chi\left(k_\iota^*(h)\right) \varepsilon_\iota P_{N,\iota}^* \right] \left(\frac{k_\iota^*(h)}{K_{H,\iota}^*} \right)^{\frac{1}{\theta_K}} \left(\frac{P_\iota^*(h)}{P_{H,\iota}^*} \right)^{-\theta_H} C_{H,\iota}^* \\ - \left[k_{\iota+1}^*(h) - (1-\delta) k_\iota^*(h) \right] \varepsilon_\iota P_{N,\iota}^* \end{array} \right\}$$

[19] The foreign variables are denoted with asterisk and all prices are denominated in the local currency.

[20] In this analysis, I choose an analytically convenient functional form for $\chi(.)$: $\chi(k) = \frac{\varphi}{k} + \vartheta$, $\chi'(k) = -\frac{\varphi}{k^2}$.

The wholesale price $\overline{P}^*_{H,t}$ is defined as the consumer price $P^*_{H,t}$ net of the per unit distribution cost $\chi(k^*_{H,t})P^*_{N,t}$:

$$\overline{P}^*_t(h) = P^*_t(h) - \chi(k^*_t(h))P^*_{N,t}$$

As in Corsetti and Dedola (2005), the demand elasticity with respect to the wholesale price $\varepsilon_{D^*_t(h),\overline{P}^*_t(h)}$ is increasing in the wholesale price and decreasing in the distribution margin $\delta_t(h) \equiv \frac{\chi(k^*_t(h))P^*_{N,t}}{P^*_t(h)}$. This satisfies the sufficient condition for incomplete pass through proposed by Marston (1990). As the wholesale price goes up, the optimal markup decreases due to a higher price elasticity of demand.

$$\varepsilon_{D^*_t(h),\overline{P}^*_t(h)} = -\frac{\partial D^*_t(h)}{\partial \overline{P}^*_t(h)} \frac{\overline{P}^*_t(h)}{D^*_t(h)} = \theta_H(1 - \delta^*_t(h))$$

The optimal price is given by

$$P^*_{H,t} = \frac{\theta_H}{\theta_H - 1}\left[\frac{MC_{H,t}}{\varepsilon_t} + \chi(k^*_{H,t})P^*_{N,t}\right]$$

The optimal markup for retail price $P^*_{H,t}$ is given by $\frac{\theta_H}{\theta_H-1}\left[1 + \frac{\varepsilon_t\chi(k^*_{H,t})P^*_{N,t}}{MC_{H,t}}\right] = \frac{\overline{P}^*_t(h)+\chi(k^*_t(h))P^*_{N,t}}{MC_{H,t}}$ and is decreasing in retail capacity. The analytical ERPT can be derived as

$$-\frac{\frac{MC_{H,t}}{\varepsilon_t}}{\frac{MC_{H,t}}{\varepsilon_t} + \chi(k^*_{H,t})P^*_{N,t}} \tag{1.4}$$

Both wholesale price and retail price move less than one to one to exchange rate shock due to the additive local component[21]. Given the optimal price, the demand elasticity with respect to retail capacity can be derived as follows:

$$\varepsilon_{D^*_t(h),k^*_t(h)} = \frac{1}{\theta_K} + \theta_H \frac{\chi(k^*_{H,t})P^*_{N,t}}{\frac{MC_{H,t}}{\varepsilon_t} + \chi(k^*_{H,t})P^*_{N,t}}$$

The first part represents the marginal benefit from owning a higher retail market share, while the second part represents the gain in demand by having a lower price. Note that the second part is a function of the distribution margin and the elasticity of substitution among home tradable products. If the price competition is intense (θ_H is high), it is optimal for the firm to invest more in retail capacity and lower its prices. Nevertheless, a higher level of retail capacity weakens the benefit of investing per effect of a lower distribution margin. Although it is clear that ERPT is an increasing function of the retail capacity holding, the

[21]Note that the elasticity of price with respect to marginal cost is the same expression but with different signs. A home appreciation lowers ε_t and increases the price of home product denominated in foreign currency and an increase in marginal cost has the same effect if the retail capacity $k^*_{H,t}$ remains constant.

direction of retail capacity adjustment depends on the types of shocks and choices of other firms.

Compared to Corsetti and Dedola (2005), the size of this component is governed by the firm's choices. The optimal investment is governed by the following expression:

$$\varepsilon_t P^*_{N,t} = E_t Q_{t,t+1} \left\{ \begin{bmatrix} \frac{\frac{1}{\theta_H - 1}\left[MC_{H,t} + \varepsilon_t \chi(k^*_{H,t}) P^*_{N,t}\right]}{\theta_K k^*_{t+1}(h)} \\ -\chi'\left(k^*_{t+1}(h)\right) \varepsilon_{t+1} P^*_{N,t+1} \end{bmatrix} D^*_{t+1}(h) + (1-\delta)\varepsilon_{t+1} P^*_{N,t+1} \right\} \tag{1.5}$$

This equation states that the producer chooses to invest in the distribution capacity until the marginal cost of retail capacity (the left hand side) equals the expected marginal benefit in the future. For the right hand side, $\frac{\frac{1}{\theta_H - 1}\left[MC_{H,t} + \varepsilon_t \chi(k^*_{H,t}) P^*_{N,t}\right]}{\theta_K k^*_{t+1}(h)}$ reflects the per unit marginal gains from an increase in demand due to a higher market share and $-\chi'\left(k^*_{t+1}(h)\right) \varepsilon_{t+1} P^*_{N,t+1} D^*_{t+1}(h)$ represents the marginal benefit per unit from being more efficient in distribution. Under reasonable assumptions on the distribution efficiency function $\chi(k)$, it can be shown that these two terms, which are multiplied by the demand $D^*_{t+1}(h)$, are both decreasing in the retail capacity. $(1-\delta)\varepsilon_{t+1} P^*_{N,t+1}$ is simply the value of the invested retail capacity next period after the depreciation. For each individual firm, the price of nontradable goods is exogenous. An increase in k^*_{t+1} reduces the first two expressions in the brackets but increases the demand via a lower price and a higher retail market share. There is a trade off between the decrease in the per unit benefits and the increase in the units sold.

1.4.2 The consumer's problem

The home representative household chooses the optimal consumption C, labour supply L, money holding M, international bonds holding B_H, B_F to maximise his lifetime expected utility[22]:

$$U_t = E_t \sum_{\iota=t}^{\infty} \beta^{\iota - t} \left[\frac{C_\iota^{1-\sigma}}{1-\sigma} + \xi \frac{\left(\frac{M_{\iota+1}}{P_\iota}\right)^{1-\sigma}}{1-\sigma} + \alpha \frac{(1-L_\iota)^{1-\nu}}{1-\nu} \right]$$

subject to

$$\begin{aligned} M_{t+1} + B_{H,t+1} + \varepsilon_t B_{F,t+1} \leq {} & M_t + (1+i_t) B_{H,t} + (1+i^*_t)\, \varepsilon_t B_{F,t} \\ & + W_t L_t - T_t - P_t C_t + \int_0^1 \pi_t(h)\, dh + \int_0^1 \pi_t(n) \end{aligned}$$

[22]As the utility function is separable in real money balances and consumption, money demand is determined residually. Therefore, the parameter in front of the real money balances does not affect the result of my analysis, thus ξ is arbitrarily set equal to 1.

where σ measures the degree of risk aversion on consumption; ν measures the disutility of labour; β is the discount factor and the risk aversion parameter on the real money holding is set equal to σ. Two international bonds $B_{H,t+1}$ and $B_{F,t+1}$ are denominated in home and foreign currency respectively. The representative household owns all home firms and receives the profit from all home producers $\int_0^1 \pi_t(h)\,dh + \int_0^1 \pi_t(n)\,dn$.

The government spending is assumed to be 0 such that

$$M_t - M_{t-1} + T_t = 0$$

All seigniorage revenues are rebated to households through lump-sum taxes. Throughout the analysis, I assume that monetary authorities adopt a strict inflation targeting such that nominal price changes are equivalent to CPI based real price changes.

1.4.3 Calibration

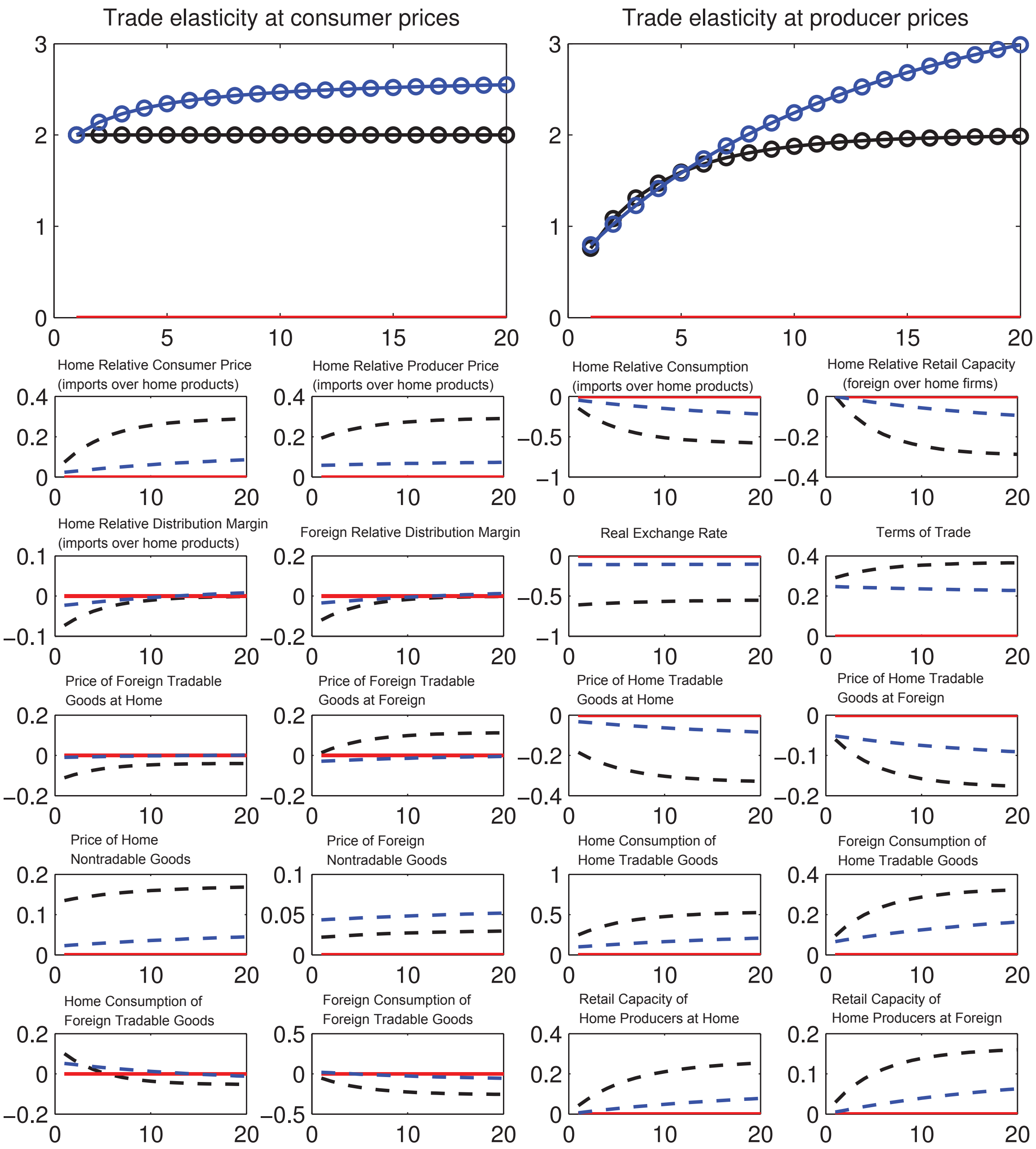
The remaining settings of the basic model and their associated equilibrium conditions are presented in appendices 1.A and 1.B.[23] The model is simulated under symmetric conditions where the value of foreign parameters is assumed to be the same as their home counterparts. A summary of calibration is available in table 1.B.3. The calibration of most parameters $(\theta_H, \theta_N, \sigma, \beta, \delta, \nu)$ follows directly from Corsetti, Dedola, and Leduc (2008a). The retail share aggregation factor θ_K is the key parameter which controls the size of the market share channel. I set θ_K equal to 1.4 such that, with a trade elasticity ρ equal to 2, the benchmark model specified in section 1.5 produces a short-run elasticity (0.5-1) consistent with the business cycle literature and a long-run elasticity (4-5) consistent with the empirical estimations of Marquez (1990) and Simonovska and Waugh (2014). The elasticity of substitution between tradable and nontradables is set to be 0.74 based on the estimation of Mendoza (1991). The value of the friction parameter γ on the capacity adjustment does not change the direction of impulse responses but slows down the capacity adjustment. I choose γ such that the average adjustment cost is around 0.5% of the current retail capacity.

1.4.4 IRFs and trade elasticity

Figure 1.3 presents impulse responses to a one percent positive permanent productivity shock on home tradable goods. Upon the shock, the marginal cost of home tradable goods decreases. The price of home tradable goods lowers less than one percent due to the existence of the distribution cost. As the productivity shock is permanent, expected lower price means a higher expected future consumer demand. This leads to additional marginal benefits from investing in the retail capacity. Producers of home tradable goods act according to equation

[23] The basic model is a special case of the generalised model where the proportion of retailing manufacturers is 1.

Fig. 1.3 In response to a 1 percent positive permanent productivity shock on home tradable goods



Note: The model with and without using the retail network CES aggregator are denoted with blue and black lines respectively. Elasticities are measured in the absolute value. All other impulse responses are measured in percentage deviations from the steady state. The horizontal axis denotes quarters after the shock.

(1.5) and increase their retail capacity gradually. The price of home tradable goods at the foreign market follows a similar pattern, but at a lower magnitude due to the appreciation of the exchange rate[24]. The change in monetary stance is reflected by the price of nontradable goods, since no distribution service is required for nontradable goods and the producers charge a constant markup over its marginal cost. As there is no shock to the productivity of nontradable goods, the rises in their prices in both countries simply reflect the increase of nominal wages. Intuitively, monetary policy in the home country inflates the nominal wage to counteract the deflation of home tradable prices and the appreciation of nominal exchange rates. The foreign monetary stance extends slightly to offset the deflation caused by the decreased price of imports. Foreign exporters lower their retail price due to the appreciation. ERPT is incomplete as a 0.5% appreciation leads to a 0.15% decrease in price. Since the price decrease is triggered by the exchange rate change rather than the change in marginal cost, the effect of the increasing demand is dominated by the rise in the local distribution cost. Therefore, a small decrease in home retail capacity for foreign exporters is observed. As to the foreign tradable goods selling in the foreign market, only a very small increase in its price is observed due to a slight rise of its marginal cost[25]. Nevertheless, a lower demand reduces the incentive of holding retail capacity and K_F decreases over time.

To understand how trade elasticity evolves after a shock, it is helpful to decompose the wholesale price elasticity of substitution between home and foreign goods into three parts:

$$
\begin{aligned}
-\frac{\partial \log\left(\frac{C_F}{C_H}\right)}{\partial \log\left(\frac{\overline{P}_F}{\overline{P}_H}\right)} &= -\frac{\partial \log\left(\frac{S_F\left(\frac{P_F}{P_T}\right)^{-\rho} C_T}{S_H\left(\frac{P_H}{P_T}\right)^{-\rho} C_T}\right)}{\partial \log\left(\frac{\overline{P}_F}{\overline{P}_H}\right)} = -\frac{1}{\theta_K}\frac{\partial \log\left(\frac{K_F}{K_H}\right)}{\partial \log\left(\frac{\overline{P}_F}{\overline{P}_H}\right)} + \rho\frac{\partial \log\left(\frac{P_F}{P_H}\right)}{\partial \log\left(\frac{\overline{P}_F}{\overline{P}_H}\right)} \\
&= -\frac{1}{\theta_K}\frac{\partial \log\left(\frac{K_F}{K_H}\right)}{\partial \log\left(\frac{\overline{P}_F}{\overline{P}_H}\right)} + \rho\frac{\partial \log\left[\frac{\left(\chi(k_F)\frac{P_N}{\overline{P}_F}+1\right)\overline{P}_F}{\left(\chi(k_H)\frac{P_N}{\overline{P}_H}+1\right)\overline{P}_H}\right]}{\partial \log\left(\frac{\overline{P}_F}{\overline{P}_H}\right)} \\
&= -\frac{1}{\theta_K}\frac{\partial \log\left(\frac{K_F}{K_H}\right)}{\partial \log\left(\frac{\overline{P}_F}{\overline{P}_H}\right)} + \rho\frac{\partial \log\left(\frac{\chi(k_F)\frac{P_N}{\overline{P}_F}+1}{\chi(k_H)\frac{P_N}{\overline{P}_H}+1}\right)}{\partial \log\left(\frac{\overline{P}_F}{\overline{P}_H}\right)} + \rho
\end{aligned}
\qquad (1.6)
$$

In addition to the conventional price effect of demand ρ, the change in relative demand is affected by two additional channels reflecting the effect of changing retail capacity. The first term in equation (1.6) captures the "retail market share channel" of demand where the change in the relative price affects the optimal retail capacity and alters relative retail network

[24]Due to the strict inflation targeting, the change of the real exchange rate is the same as that of the nominal exchange rate.

[25]Due to the strict inflation targeting, the foreign monetary stance extends slightly to offset the deflation introduced by a decreased price of imports.

shares. The second term reflects the "distribution margin channel" where the wholesale price does not move one to one with the retail price due to the existence of the distribution margin. It is easy to verify that this part is negative with a given price of nontradables and constant retail capacity.

Similarly, the elasticity of substitution measured at the consumer price can be derived as follows.

$$-\frac{\partial \log\left(\frac{C_F}{C_H}\right)}{\partial \log\left(\frac{P_F}{P_H}\right)} = -\frac{1}{\theta_K}\frac{\partial \log\left(\frac{K_F}{K_H}\right)}{\partial \log\left(\frac{P_F}{P_H}\right)} + \rho \tag{1.7}$$

I now compare the model using the conventional CES aggregator, denoted as CES (black), with the model using the retail network CES aggregator, denoted as RNCES (blue). For the CES model, the trade elasticity with respect to the producer (wholesale) price is low in the short run and gradually back to the calibrated value in the long-run. The relative distribution margin decreases immediately after the shock but gradually returns to zero in the steady state. However, the trade elasticity with respect to consumer prices stays constant since this elasticity is not affected by the distribution margin channel.

The RNCES model amplifies the difference between shot-run and long-run trade elasticities. It is important to note that this amplification effect does not necessarily work through enlarging the magnitude of responses. It is the relative magnitude of relative changes that matters. The lower panel of graphs in figure 1.3 shows that the magnitude of responses measured by their percentage deviation from the steady state is actually smaller for the RNCES model. For the CES model, the relative consumer price of imports over home products increases gradually by a large amount. However, the increase in the relative consumption is even bigger. The increase in the relative consumption is twice as much as the increase in the relative consumer price, resulting in a constant trade elasticity at consumer prices of 2. For the RNCES model, the change in the relative price is much smaller than that for the CES model but the difference in changes in the relative consumption between RNCES and CES models is relatively small, resulting in a steady increase in the trade elasticity at the consumer and producer prices over time.

1.5 The Full Model

This section considers a more realistic market structure and allows for four types of firms interacting with each other, namely retailing manufacturers, non-retailing manufacturers, local retailers and nontradable goods producers. For each type of tradable goods, there is a proportion of non-retailing manufacturers who sell to local retailers and do not manage their own retail network. Local retailers sell a range of home and foreign tradables goods, set the retail price for each product and adjust their retail networks according to all products they sell.

The optimisation problem of foreign local retailer r is as follows. The foreign local retailer r buys a range of imports and home products and sets the retail price $P_t^*(h_r,r)$, $P_t^*(f_r,r)$ to maximise its profits subject to the demand of product $D_t^*(h_r,r)$, $D_t^*(f_r,r)$ and chooses this optimal investment $i_t^*(r)$ for all products it sells subject to the law of motion of retail capacity. For each individual product h_r, f_r the marginal cost is $P_{im,t}^*(h_r,r)$, $P_{local,t}^*(f_r,r)$. The total profit is constructed by aggregating profits and subtracting the total cost of investments made $i_t^*(r)P_{N,t}^*$ to extend retail capacity.

$$\max_{\{P_\iota^*(h_r,r),P_\iota^*(f_r,r),i_\iota^*(r)\}_{\iota=t}^{\infty}} E_t \sum_{\iota=t}^{\infty} Q_{t,t+\iota} \left\{ \begin{array}{c} \int_x^1 \left[P_\iota^*(h_r,r) - P_{im,\iota}^*(h_r,r)\right] D_\iota^*(h_r,r)dh_r + \\ \int_x^1 \left[P_\iota^*(f_r,r) - P_{local,\iota}^*(f_r,r)\right] D_\iota^*(f_r,r)df_r - i_\iota^*(r)P_{N,\iota}^* \end{array} \right\}$$

subject to

$$D_t^*(h_r,r) = \left(\frac{k_t^*(r)}{K_{H,t}^*}\right)^{\frac{1}{\theta_K}} \left(\frac{P_t^*(h_r,r)}{P_{H,t}^*}\right)^{-\theta_H} C_{H,t}^*$$

$$D_t^*(f_r,r) = \left(\frac{k_t^*(r)}{K_{F,t}^*}\right)^{\frac{1}{\theta_K}} \left(\frac{P_t^*(f_r,r)}{P_{F,t}^*}\right)^{-\theta_H} C_{F,t}^*$$

$$k_{t+1}^*(r) = i_t^*(r) + (1-\delta)k_t^*(r) - \chi_2(k_t^*(r)) \int_x^1 D_t^*(h_r,r)dh_r - \chi(k_t^*(r)) \int_x^1 D_t^*(f_r,r)df_r$$

The law of motion of retail capacity is similar to (1.3) where $\chi_2(k_t^*(r)) \int_x^1 D_t^*(h_r,r)dh_r$ and $\chi(k_t^*(r)) \int_x^1 D_t^*(f_r,r)df_r$ are the distribution costs for the sales of all imports and local products respectively.[26] In this case, home product h_r competes with other home products based on retail networks of its embedded local retailers and has a retail market share of $\frac{k_t^*(r)}{K_{H,t}^*}$. The cost of selling to local retailers is the drop in sales because the product is sold at a higher price due to the additional markup charged by local retailers.

The optimal retail prices are given by

$$P_t^*(h_r,r) = \frac{\theta_H}{\theta_H - 1}\left[P_{im,t}^*(h_r,r) + \chi_2(k_t^*(r))P_{N,t}^*\right]$$

$$P_t^*(f_r,r) = \frac{\theta_H}{\theta_H - 1}\left[P_{local,t}^*(f_r,r) + \chi(k_t^*(r))P_{N,t}^*\right]$$

The optimal price reflects two components that affect ERPT, namely the double marginalisation and the distribution margin.

[26] I allow for different distribution cost functions to promote local and foreign products, $\chi(.)$ and $\chi_2(.)$.

The optimal retail capacity for foreign local retailers hinges on the purchase price and demand of all home and foreign products.

$$P^*_{N,t} = E_t Q_{t,t+1} \left\{ \begin{array}{c} \int_x^1 \left[\frac{P^*_t(h_r,r) - P^*_{im,t}(h_r,r) - \chi_2(k^*_t(r)) P^*_{N,t}}{\theta_K k^*_{t+1}(r)} - \chi_2'\left(k^*_{t+1}(r)\right) P^*_{N,t+1} \right] D^*_{t+1}(h_r) dh_r \\ + \int_x^1 \left[\frac{P^*_t(f_r,r) - P^*_{local,t}(f_r,r) - \chi(k^*_t(r)) P^*_{N,t}}{\theta_K k^*_{t+1}(r)} - \chi'\left(k^*_{t+1}(r)\right) P^*_{N,t+1} \right] D^*_{t+1}(f_r) df_r \\ +(1-\delta) P^*_{N,t+1} \end{array} \right\}$$

For a home exporter h_r its demand is the sum of demand of all foreign local retailers:

$$D^*_t(h_r) = \int_0^1 D^*_t(h_r,r) dr = \int_0^1 \left(\frac{k^*_t(r)}{K^*_{H,t}} \right)^{\frac{1}{\theta_K}} \left(\frac{P^*_t(h_r,r)}{P^*_{H,t}} \right)^{-\theta_H} C^*_{H,t} dr$$

Substituting $P^*_t(h_r,r)$ with the price set by local retailers and using the assumption that local retailers are homogeneous, the demand can be written as

$$D^*_t(h_r) = \left(\frac{K^*_{R,t}}{K^*_{H,t}} \right)^{\frac{1}{\theta_K}} \left(\frac{\frac{\theta_H}{\theta_H - 1} \left[P^*_{im,t}(h_r) + \chi_2(K^*_{R,t}) P^*_{N,t} \right]}{P^*_{H,t}} \right)^{-\theta_H} C^*_{H,t}$$

Given the demand from local retailers, the non-retailing exporter's problem is therefore

$$\max_{P^*_{im,t}(h_r)} \quad \left[\varepsilon_t P^*_{im,t}(h_r) - MC_t(h_r) \right] \left(\frac{K^*_{R,t}}{K^*_{H,t}} \right)^{\frac{1}{\theta_K}} \left(\frac{\frac{\theta_H}{\theta_H - 1} \left[P^*_{im,t}(h_r) + \chi_2(K^*_{R,t}) P^*_{N,t} \right]}{P^*_{H,t}} \right)^{-\theta_H} C^*_{H,t}$$

Its optimal price is given by

$$P^*_{im,t}(h_r) = \frac{\theta_H}{\theta_H - 1} \left[\frac{1}{\theta_H} \chi_2(K^*_{R,t}) P^*_{N,t} + \frac{MC_t(h_r)}{\varepsilon_t} \right]$$

Similarly, the optimal price for local products can be derived as

$$P^*_{local,t}(f_r) = \frac{\theta_H}{\theta_H - 1} \left[\frac{1}{\theta_H} \chi(K^*_{R,t}) P^*_{N,t} + MC_t(f_r) \right]$$

The retailer's price expressed in terms of producer's marginal cost and distribution margin is:

$$P^*_t(h_r,r) = \left(\frac{\theta_H}{\theta_H - 1} \right)^2 \left[\chi_2(k^*_t(r)) P^*_{N,t} + \frac{MC_t(h_r)}{\varepsilon_t} \right]$$

where $\left(\frac{\theta_H}{\theta_H - 1} \right)^2$ reflects double marginalisation. The ERPT at retail prices is given by

$$\frac{\frac{MC_t(h_r)}{\varepsilon_t}}{\chi_2(k_t^*(r))P_{N,t}^* + \frac{MC_t(h_r)}{\varepsilon_t}}$$

The analytical ERPT of the retail price is similar to the expression of equation (1.4) implying that double marginalisation does not affect ERPT at the consumer price. ERPT will be the same if retailing and non-retailing manufacturers have the same retail capacity. However, simulation shows that local retailers accumulate the largest amount of retail capacity, followed by home retailing manufacturers and foreign exporters. By selling to the local retailer, non-retailing exporters are able to utilise the large retail network accumulated by local retailers and lower their distribution margin, resulting in a high ERPT.

The home bias in tradables arises naturally under two circumstances: (1) There exists an iceberg trade cost or trade barriers which increase the optimal prices charged by exporters. The increase in price lowers the demand and thus lowers the gain from investing in the retail capacity. As a result, the foreign retailing exporters optimally reduce their retail capacity, which lowers the market share of foreign goods and leads to home bias. (2) *Ceteris paribus*, it is more costly to promote and sell the foreign good, i.e., distribution cost function for foreign products $\chi_2(k)$ first order stochastically dominates that for home products $\chi(k)$. The higher cost for each unit of product sold leads to a lower level of the retail capacity for foreign retailing exporters. In this case, the lower market share reflects the difficulty to promote and distribute imported goods in the local market if domestic tradables and foreign tradable goods have the same quality. With an iceberg trade cost of 3% and 10% additional cost of distributing foreign products, the benchmark model gives an import share of 8%.

The rest of the model and their associated equilibrium conditions are presented in section 1.A in appendix. All equilibrium conditions for the extended model are summarized in section 1.B in appendix.

1.5.1 Calibration and key statistics

On top of the calibration of the basic model, seven additional parameters need to be calibrated, namely τ, φ_H, φ_F, ϑ, ζ_1, ζ_2 and the proportion of retailing manufacturers Ξ. Following Ghironi and Melitz (2005), the iceberg trade cost τ is set to 3%. The fixed part of the distribution cost function ϑ is calibrated such that the distribution margin is in the range of 40% to 60%. With 10% additional distribution cost for distributing foreign products ($\varphi_F = 1.1\varphi_H$), the model could generate a high degree of home bias and give an import share of around 8%. The proportion of retailing manufacturers Ξ is a parameter which may differ across markets. I have not found enough empirical evidence to calibrate it. In the benchmark model, I select the Ξ which matches the best with the calculated moments from data.[27] ζ_1 and ζ_2 are related

[27]The basic model analyses the response of the economy assuming $\Xi = 1$.

to endogenous discount factors.[28] ζ_1 controls the variance of the discount factor and ζ_2 is chosen such that the steady state real interest rate is around 1 percent per quarter. In figure 1.C.1, I plot simulated series of β_t for different values of ζ_1. Technology shocks are assumed to follow a trend-stationary AR(1) process, $\mathbf{Z}_t' = \psi \mathbf{Z}_{t-1}' + \boldsymbol{u}_t$,where $\mathbf{Z}_t = [Z_{T,t}, Z_{N,t}]$ is a vector of the productivity of tradable and nontradable goods respectively. I use the same auto-correlation and variance-covariance matrices as in Corsetti, Dedola, and Leduc (2008a).

Table 1.C.3 in the appendix estimates ERPT for all types of firms using data generated by the model. Estimation regressions take the same form as the empirical estimation regression (1.3.1). Together with the average distribution margin reported in the key statistics table, it can be seen that a higher distribution margin reduces the pass through. This result is consistent with the analytical expression where the level of pass through is negatively correlated with the distribution margin. Furthermore, the difference between the short-run and the long-run ERPT increases in the level of the distribution margin.

The model's ability to match data patterns of international business cycles and international relative prices is summarised by tables 1.C.1 and 1.C.2 in the appendix. Empirical moments are calculated based on quarterly data from 1980:1 - 2013:2. The model performs better in terms of the correlation of international prices. In addition, the benchmark model can successfully generate a smaller international consumption correlation compared to the output correlation. With reasonably small frictions on distributing foreign products and a standard iceberg trade cost, the model can generate a sizeable home bias per effect of the retail resource competition. Introducing frictions on capacity adjustment to the benchmark model lowers the variance of investment and further improves the fitness.

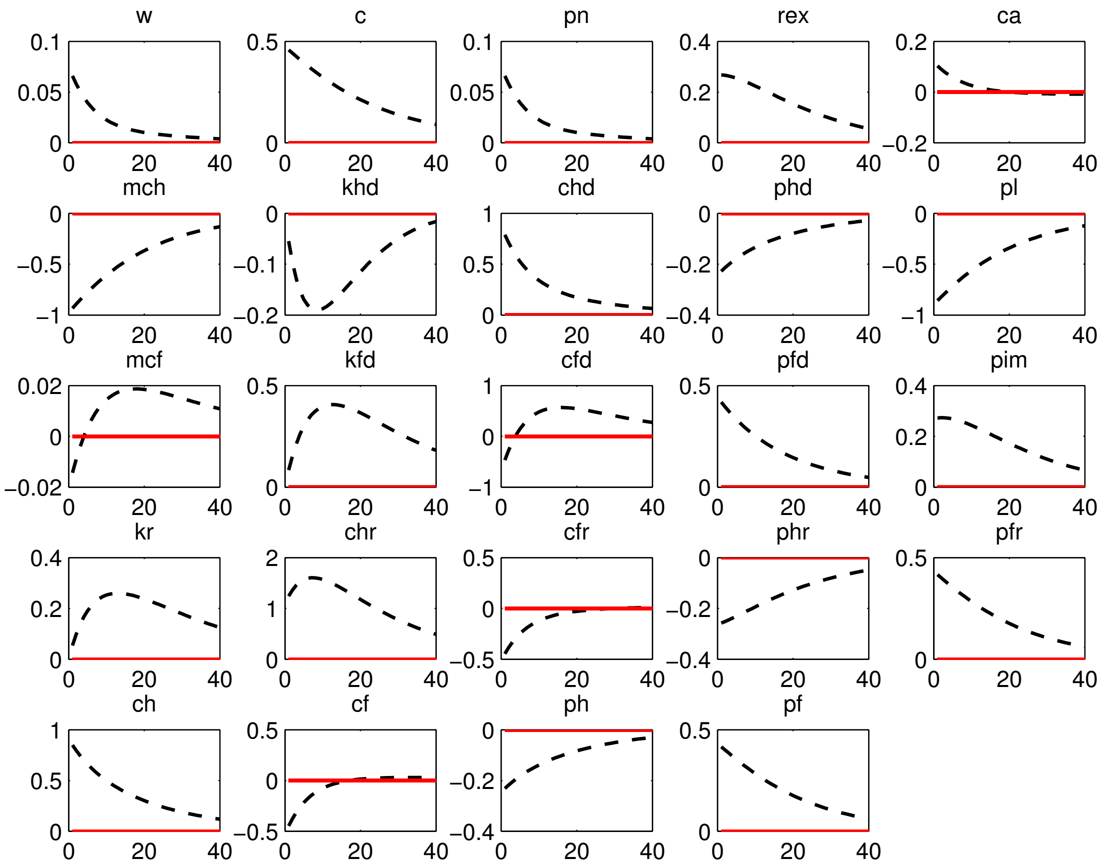
1.5.2 The analysis of impulse responses

The response to a positive productivity shock on tradable goods

Figure 1.4 presents impulse responses to a persistent positive productivity shock in the benchmark model with frictions on capacity adjustment. The productivity shock is assumed to follow an AR(1) process with a persistent parameter equal to 0.95. Upon the shock, home producers of tradable goods face a lower marginal cost and choose to lower their prices. At the equilibrium, more products are produced and the demand for labour increases, which in turn increases the equilibrium wage. As a result, the representative household consumes less leisure, supplies more labour and earns more. Due to the wealth effect, the representative household consumes more domestic products as well as imports. The real exchange rate depreciates and home products are sold more in both domestic market and foreign markets. As tradable goods and nontradable goods are complements, more

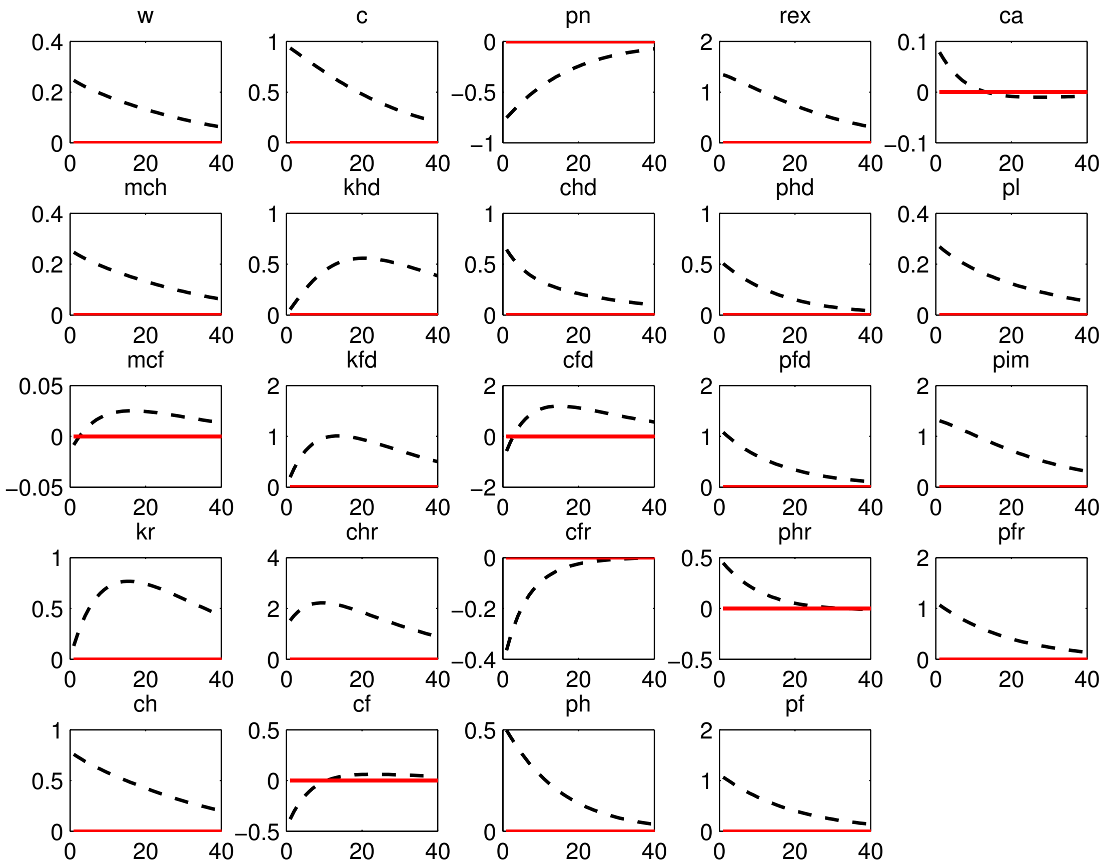
[28]Following Schmitt-Grohé and Uribe (2003) and Corsetti, Dedola, and Leduc (2008a), the extended model introduces the endogenous discount factor for the representative household taking the following form: $\beta_t = \ln\left\{\zeta_1\left[1+\zeta_2\left(C_t + \frac{M_{t+1}}{P_t} + \alpha\left(1-L_t\right)\right)\right]\right\}$

Fig. 1.4 In response to a positive persistent productivity shock on home tradables



Note: The black dashed line denotes the impulse response of the benchmark model with frictions on capacity adjustment. Impulse responses are expressed in percentage deviations from the steady state. The horizontal axis denotes quarters after the shock. Please check table 1.B.2 for references of variable names.

Fig. 1.5 In response to a positive persistent productivity shock on home nontradable goods



Note: The black dashed line denotes the impulse response of the benchmark model with frictions on capacity adjustment. Impulse responses are expressed in percentage deviations from the steady state. The horizontal axis denotes quarters after the shock. Please check table 1.B.2 for references of variable names.

nontradable goods are needed as sales of tradable goods rise. The price of nontradable goods goes up. The expected increase in sales for home products drops over time as the technology shock dies out. Similarly, the price of nontradable good falls gradually after a large impulse increase. Knowing that the price of nontradables is high today but low tomorrow, home retailing manufacturers choose to disinvest today and increase their investment when the price of investment is at a relatively low level. This gives an example where a lower price and a higher expected demand are not accompanied by an increase in the retail capacity due to horizontal competitions of retail resources with local retailers.

Due to the spillover effect, foreign exporters face an increase in marginal cost. Upon the shock, foreign retailing manufacturers increase their price and face a drop in demand of around 5%. As a result, the consumption of home products increases while the consumption of foreign products goes down. The relative consumption drops. The foreign exporters will have a large drop in quantities sold in home markets if they do not adjust their distribution capacity. Although the price of investment is relative high now, the benefit of the gain in demand outweighs the cost of investment, and foreign retailing manufacturers invest gradually to increase their retail market share. In this case, the relative retail capacity goes up gradually. The positive effect of a higher retail market share on demand dominates the negative effect of a high price. The demand for foreign products increases.

As specified in the previous section, the investment decision of local retailers depends on the quantity and price of both home and foreign products. Non-retailing manufacturers do not have controls over their retail networks and their prices decrease proportionally to the marginal cost shock. Similarly, import prices of foreign non-retailing manufacturers increase slightly per effect of the exchange rate depreciation and the small increase in their marginal cost. The relative price of retailing manufacturers decreases. Retailers face an increase in demand for home produced products and a decrease in demand for imported goods. At equilibrium, they end up choosing to extend their retail networks to amplify the gain from home products and partially make up their loses from imports.

Shock on nontradable goods

A positive and persistent shock on nontradable goods reduces the marginal cost of producers of nontradable products. The price of nontradables thus decreases and sales of nontradables increase. The demand of labour increases, giving rise to a higher wage. At the equilibrium, the representative household supplies more labor and consumes more, increasing the demand of all products.

Local retailers, home and foreign retailing manufacturers choose to expand their retail capacity subject to two effects, namely the demand effect and price effect. On the one hand, expected demand increases, which gives rise to larger total benefit of investment in retail capacity and thus induces a higher retail capacity holding. On the other hand, the price of investment is low at the moment and is expected to increase. It is optimal to invest now and disinvest when the cost is high. Due to frictions on adjusting retail capacity, the capacity cannot adjust directly to the desired level. Simulation shows that the price effect outweighs the demand effect with the retail capacity reaching the highest after 15 quarters.

The marginal cost of home tradable goods increases and home retailing manufacturers increase their price. However, the price increase of home retailing manufacturers is less than the drop in their marginal cost as the distribution cost drops. Prices of imports increase more than home prices due to the depreciation of real exchange rates. Note that the pass through for import prices is very high. Given a small increase in the marginal cost, the import price almost increases one-to-one to the real exchange rate. At the consumer level, the price of these products increases less than the price of imported products does. The relative price of local retailer selling products rises. As a consequence, sales of foreign retailing manufacturers drop initially. To prevent a further decline in their demand, it is optimal for them to invest more in the retail capacity compared to home retailing manufacturers. Their sales gradually go up as their retail networks expand. The choice of expanding the retail capacity is also based on fact that the expected demand in the future is relatively high as the exchange rate of home country depreciates. The relative demand is initially lower than the steady state value but gradually goes up as the foreign retail capacity expands until the depreciation stops.

1.5.3 Trade elasticities

Table 1.1 presents OLS estimation results of elasticity of substitutions based on simulated series. In the benchmark model without frictions on adjusting investment, any shock that changes the desired retail capacity will be adjusted instantly. In column (1) , regressions controlling for relative retail capacity (part a) reflect the short-run trade elasticity. This elasticity is captured by the coefficient on the change of relative price, e.g. $|-1.9782|$. Regression results of part (b) imply the long-run trade elasticity.

The first set of regressions estimate the trade elasticity between home and foreign tradables goods. These goods include products of both retailing and non-retailing manufacturers. The estimated short run trade elasticity is around the calibrated value of 2. The coefficient on the relative change in retail capacity is positive, consistent with the analytical decomposition in equation (1.6). The second set of regressions estimate the elasticity of substitution between home and foreign retailing manufacturers. This elasticity in the short-run is slightly lower than the calibrated value and the implied long-run elasticity is as high as 4.22. The last set of regressions measure the elasticity of substitution between retailers selling home products and imports. Estimations of part (c) and (d) in the bottom panel are based on relative retail prices and relative producer prices respectively.

Table 1.1 Estimated elasticity of substitution between home and foreign products

Dependent Variable		Explanatory Variables	(1) Benchmark	(2) Benchmark + Capacity Adjustment Friction
$Y : \log \frac{C_{F,t}}{C_{H,t}}$	(a):	$\log \frac{P_{F,t}}{P_{H,t}}$	-1.9782	-2.0367
			(0.0110)	(0.0010)
		$\log \frac{K_{F,t}}{K_{H,t}}$	0.6857	0.7141
			(0.0050)	(0.0006)
		R^2	0.9810	0.9981
	(b):	$\log \frac{P_{F,t}}{P_{H,t}}$	-3.2442	-1.9029
			(0.0076)	(0.0110)
		R^2	0.9476	0.7504
$Y : \log \frac{C_{FD,t}}{C_{HD,t}}$	(a):	$\log \frac{P_{FD,t}}{P_{HD,t}}$	-1.7962	-0.8125
			(0.0161)	(0.0200)
		$\log \frac{K_{FD,t}}{K_{HD,t}}$	0.8549	0.4982
			(0.0044)	(0.0079)
		R^2	0.9539	0.3323
	(b):	$\log \frac{P_{FD,t}}{P_{HD,t}}$	-4.2215	-0.6137
			(0.0226)	(0.0267)
		R^2	0.7777	0.0647
$Y : \log \frac{C_{RF,t}}{C_{RH,t}}$	(c):	$\log \frac{P_{RF,t}}{P_{RH,t}}$	-3.5466	-3.7320
			(0.0462)	(0.0148)
		R^2	0.3710	0.8634
	(d):	$\log \frac{P_{IM,t}}{P_{L,t}}$	-2.1369	-2.2249
			(0.0277)	(0.0146)
		R^2	0.3749	0.8620

Note: Standard errors are reported in parentheses.

The column (2) presents results of the model with frictions on retail capacity adjustments. The short-run trade elasticity is estimated to be 2.03, similar to the benchmark model. Since the retail capacity cannot be adjusted immediately, estimates of part (b) are similar to those of part (a). Note that adding frictions on retail capacity adjustment changes the dynamics of responses of home and foreign retailing manufacturers. Decisions on prices and optimal capacity are more complicated as these firms react based on their expectations of future nontradable price and their shares of retail networks. The estimated trade elasticity reacts in a smaller magnitude compared to the benchmark model due to a stronger competition effect from local retailers. The associated R-squared is much lower compared to the benchmark model, indicating that there exists non-pricing explanatory factors and that the current regression specification fails to capture the adjustment process of relative consumption.

In the bottom panel, the estimated coefficients are very similar in two model specifications. This is due to the separation between pricing and retailing decisions. Non-retailing manufacturers only make price decisions at each period, taking the local retailer's decision on the optimal retail capacity as given. In addition, when the local retailer adjusts its retail capacity, home and foreign non-manufacturers are affected equally as they share the same retailing network.

1.5.4 Price volatility and distribution margin

In this subsection, I discuss the model implications on the connections among distribution margin, price volatility and ERPT. Specifically, I derive the short-run analytical expression of the correlation between price volatility and ERPT in terms of the distribution margin, and explore the long-run relationship using simulated data from the model.

Recall that the optimal price of home retailing manufacturers at the foreign country can be expressed as:

$$P_t^*\left(h_d\right) = \frac{\theta_H}{\theta_H - 1}\left[\frac{MC_{H,t}}{\varepsilon_t} - \chi_2(k_t^*\left(h_d\right))P_{N,t}^*\right]$$

This expression can be approximated by

$$p_t^*\left(h_d\right) \approx -\frac{\frac{MC_t(h_d)}{\varepsilon_t}}{\frac{MC_t(h_d)}{\varepsilon_t} + \chi_2(k_t^*\left(h_d\right))P_{N,t}^*}\left(z_t\left(h_d\right) + e_t\right) + \frac{\chi_2(k_t^*\left(h_d\right))P_{N,t}^*}{\frac{MC_t(h_d)}{\varepsilon_t} + \chi_2(k_t^*\left(h_d\right))P_{N,t}^*}\left(\omega_t\left(h_d\right)\right)$$

where $p_t^*\left(h_d\right)$ is approximated at the first order around time t, and $p_t^*\left(h_d\right), z_t\left(h_d\right), e_t$, $\omega_t\left(h_d\right)$ represent the first difference of logged variable $P_t^*\left(h_d\right)$, $Z_t\left(h_d\right), \varepsilon_t$ and distribution cost $\chi_2\left(k_t^*\left(h_d\right)\right) D_t^*\left(h_d\right)$.

Define $\eta_t(h_d) \equiv \frac{\chi_2(k_t^*(h_d))P_{N,t}^*}{\frac{MC_t(h_d)}{\varepsilon_t}+\chi_2(k_t^*(h_d))P_{N,t}^*}$ which is proportional to the distribution margin at time t and the analytical pass through equals $1-\eta_t(h_d)$. The short-run variance around t can be written as

$$\begin{aligned} var\left(p_t^*(h_d)\right) = \; & [1-\eta_t(h_d)]^2 var\left(z_t(h_d)+e_t\right)+\eta_t(h_d)^2 var\left(\omega_t(h_d)\right) \\ & -2\eta_t(h_d)[1-\eta_t(h_d)]\,cov\left(z_t(h_d)+e_t,\omega_t(h_d)\right) \end{aligned}$$

Note that both short-run pass though and price volatility are functions of distribution margin. Although the retail capacity fails to adjust in the short run, the price of nontradables may not be constant. With idiosyncratic shocks that do not alter the price of nontradables, the second and third terms on the right hand side are nearly zero and the exchange rate pass through is perfectly correlated with price volatility. In order for short-run pass through to be positively correlated with the short-run price volatility, the following expression must be true:

$$\frac{\partial var\left(p_t^*(h_d)\right)}{\partial[1-\eta_t(h_d)]} = \frac{\begin{array}{c} 2[1-\eta_t(h_d)]\,var\left(z_t(h_d)+e_t\right)-2\eta_t(h_d)\,var\left(\omega_t(h_d)\right) \\ -2[1-2+2\eta_t(h_d)]\,cov\left(z_t(h_d)+e_t,\omega_t(h_d)\right) \end{array}}{} > 0$$

Since idiosyncratic shock is unlikely to be correlated with change in aggregate prices and exchange rates, we have

$$\begin{aligned} var\left(z_t(h_d)\right) \;>\; & \frac{1}{1-\eta_t(h_d)}\left[var\left(\omega_t(h_d)\right)+cov\left(z_t(h_d)+e_t,\omega_t(h_d)\right)\right] \\ & -var\left(z_t(h_d)+e_t+\omega_t(h_d)\right) \end{aligned}$$

In the short run where the retail capacity fails to adjust, the exchange rate pass through of a firm is positively correlated with the volatility of its import price if (a) the size of the idiosyncratic shocks is very large and (b) its distribution margin is low. According to (b), an increase in distribution margin reduces the correlation between ERPT and price volatility, i.e.,

$$\frac{\partial^2 var\left(p_t^*(h_d)\right)}{\partial[1-\eta_t(h_d)]\,\partial\eta_t(h_d)} = -2\left[var\left(z_t(h_d)+e_t\right)+var\left(\omega_t(h_d)\right)+2cov\left(z_t(h_d)+e_t,\omega_t(h_d)\right)\right] < 0$$

According to Berger and Vavra (2013), firm idiosyncratic shocks account for more than 90% of the total variance of prices, which suggests a high correlation between ERPT and exchange volatility. If the relationship holds in short run, the long-run volatility of price must be decreasing in the distribution margin.

To understand the long-run properties of price volatility, I run several regressions of simulated response of prices to idiosyncratic shocks as well as aggregate shocks. I examine additional idiosyncratic shocks of different sizes to individual firms and calculate firms' responses over 10000 periods based on simulated aggregate variables. Variances are calculated based on logged and first differenced variables. pfr, pim and pfd denote home retail prices of foreign non-retailing manufacturers, home import prices of foreign non-retailing manufacturers and home retail prices of foreign retailing manufacturers respectively. Results are summarised in table 1.2. The first regression for each price is similar to the one run by Berger and Vavra (2013). In the first set of regressions (labelled with (1)), the corresponding coefficients on the variance of nominal exchange rate are always significant. However, the low R-squared of these regressions suggests that the change in exchange rate only explains a very small part of the variance of the import prices.

The second set of regressions (labelled with (2)) add the variance of marginal cost of the firm and its distribution margin. Coefficients on marginal costs are highly significant for all three prices. The contribution of the variance of distribution margin is low for pfr and pim. Note that these types of firms sell their product through local retailers and cannot adjust their retail capacity under the idiosyncratic shocks. These two regressions illustrate that the change in nontradable prices does not affect the variance of the price significantly for non-retailing manufacturers. On the contrary, the variance of distribution margin explains 40% for retailing manufacturers.

The last set of regressions (labelled with (3)) separate the idiosyncratic shocks from aggregate shocks and separate the change in retail capacity from the change in nontradable prices. Results show that idiosyncratic shocks are important in explaining variance of prices. The effect of aggregate shocks is significant for non-retailing manufacturers but not for retailing manufacturers. In addition, the coefficient of variance of aggregate productivity shock on retail prices of non-retailing manufacturers is lower compared to the second set of regressions, suggesting that local retailers partially adjust their retail networks.

Table 1.2 Decomposition of price volatility

	$var(pfr(i)_t)$			$var(pim(i)_t)$			$var(pfd(i)_t)$		
Coefficients	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
$var(\varepsilon_t)$	0.9351	0.9149	0.4306	2.0248	1.8851	1.3645	0.4050	0.8677	0.0787
	(0.1968)	(0.0490)	(0.0075)	(0.5625)	(0.1396)	(0.0140)	(0.2137)	(0.0466)	(0.0269)
$var(mcf(i)_t)$		0.3134			0.9130			0.2505	
		(0.0031)			(0.0057)			(0.0099)	
$var(dc(i)_t)$		0.0002			0.0022			2.8549	
		(0.0030)			(0.0057)			(0.2819)	
$var(e_t)^a$			0.6548			1.8490			0.0650
			(0.0059)			(0.0169)			(0.0874)
$var(MCF_t)$			0.0503			1.1590			-0.0462
			(0.0318)			(0.0908)			(0.0301)
$var(K_{R,t}\ or\ k_{fd,t})$			-0.3474			-3.2638			1.7211
			(0.7743)			(2.2071)			(0.2296)
$var(P_{N,t})$			-1.8472			-4.9040			-1.8883
			(0.4215)			(1.2015)			(0.3993)
R^2	0.0434	0.9630	0.9565	0.0254	0.9625	0.9810	0.0072	0.9710	0.9548
Shock Decomposition for Regression (2) $var(Prices)$		3.12e-04			9.41e-04			3.10e-04	
$var(\varepsilon_t)$		2.75e-04			2.75e-04			2.75e-04	
		(37.84%)			(39.80%)			(6.98%)	
$var(mcf(i)_t)$		6.21e-04			6.21e-04			6.21e-04	
		(62.28%)			(60.21%)			(50.97%)	
$var(dc(i)_t)$		1.92e-05			1.92e-05			4.58e-05	
		(0.00%)			(0.00%)			(42.22%)	

Note: Regression statistics are based on simulated data generated by model "Benchmark + Capacity Adjustment Friction". I examine additional idiosyncratic shocks to individual firms and calculate their responses over 10000 periods based on simulated aggregate variables. Variances are calculated based on 500 bootstrapped firms.

[a] e_t is the idiosyncratic productivity shock to individual firms.

1.6 Conclusion

Understanding movements of international prices lies in the heart of open macroeconomic studies. The degree of ERPT, the trade elasticity and the volatility of import price are three key measures that govern the behaviour of international prices. This paper explores the role of the dynamics of local distribution margin in explaining the connections among these three measures.

I investigate the empirical property of distribution margin and find the following patterns. First, distribution margin differs greatly across industries within a country. Second, while the difference in distribution margin is relatively large for the same industry across countries, this difference is substantially smaller at the aggregate level. Third, the variance of the retail price margin shows a high degree of heterogeneity. Fourth, I find a positive relationship between estimated distribution margin and ERPT for 10 European countries as predicted by theoretical models.

I extend Corsetti and Dedola (2005) and model distribution as an investment decision in retail capacity. The slow adjustment of retail capacity restricts the change in demand in the short-run and gives a natural explanation to the trade elasticity puzzle. I show that the trade elasticity can be decomposed into 3 channels and estimate the quantitative importance of each channel in explaining the short-run and long-run discrepancies in the trade elasticity.

The model contributes to the literature on market structures with strategic vertical and horizontal interactions. In the extended model, retailing manufacturers compete for retail resources with local retailers. The optimal decisions of retailing manufacturers depend on the responses of local retailers and non-retailing manufacturers. Since the price elasticity of demand is a function of the distribution margin, the aggregate level of ERPT and trade elasticity are sensitive to the proportion of each type of firms.

As an extension of the analysis, I investigate the role of retail capacity in explaining the empirical positive relationship between price volatility and ERPT. With the proposed model, I show analytically and quantitatively that an increase in the long-run average distribution margin reduces the correlation between price volatility and ERPT. In the short run where the retail capacity fails to adjust, ERPT of a firm is positively correlated with the volatility of its import price if (a) the size of idiosyncratic shocks is large and (b) its distribution margin is low.

Appendix

Appendix 1.A Derivations

1.A.1 Solve for the consumer's problem

The representative household's problem can be rewritten as:

$$U_t = \max E_t \sum_{\iota=t}^{\infty} \prod_{\omega=t}^{\iota} \beta_\omega \left\{ \begin{array}{c} \frac{1}{1-\sigma} \left[\begin{array}{c} \frac{M_{\iota+1}}{P_\iota} + (1+i_\iota)\frac{B_{H,\iota}}{P_\iota} + (1+i_\iota^*)\frac{\varepsilon_\iota B_{F,\iota}}{P_\iota} + \frac{\int_0^1 \pi_\iota(c,j)dc}{P_\iota} \\ + \frac{\int_0^1 \pi_\iota(h,j)dh}{P_\iota} + \frac{W_\iota L_\iota}{P_\iota} - \frac{T_\iota}{P_\iota} - \frac{M_\iota}{P_\iota} - \frac{B_{H,\iota+1}}{P_\iota} - \frac{\varepsilon_t B_{F,\iota+1}}{P_\iota} \end{array} \right]^{1-\sigma} \\ + \frac{1}{1-\sigma}\left(\frac{M_{\iota+1}}{P_\iota}\right)^{1-\sigma} + \alpha \frac{(1-L_\iota)^{1-\nu}}{1-\nu} \end{array} \right\}$$

First order conditions:

The Euler equation:

$B_{H,t+1}$,$B_{F,t+1}$:

$$\frac{C_t^{-\sigma}}{P_t} = \beta_t \left(1 + i_{t+1}\right) E_t \left(\frac{C_{t+1}^{-\sigma}}{P_{t+1}}\right)$$

$$\frac{\varepsilon_t C_t^{-\sigma}}{P_t} = \beta_t \left(1 + i_{t+1}^*\right) E_t \left(\frac{\varepsilon_{t+1} C_{t+1}^{-\sigma}}{P_{t+1}}\right)$$

Foreign counterparts:

$B_{H,t+1}^*$,$B_{F,t+1}^*$:

$$\frac{\left(C_t^*\right)^{-\sigma}}{\varepsilon_t P_t^*} = \beta_t \left(1 + i_{t+1}\right) E_t \left[\frac{\left(C_{t+1}^*\right)^{-\sigma}}{\varepsilon_{t+1} P_{t+1}^*}\right]$$

$$\frac{\left(C_t^*\right)^{-\sigma}}{P_t^*} = \beta_t \left(1 + i_{t+1}^*\right) E_t \left[\frac{\left(C_{t+1}^*\right)^{-\sigma}}{P_{t+1}^*}\right]$$

Optimal money holding M_{t+1} :

$$M_{t+1} = \left(\chi_t \frac{1 + i_{t+1}}{i_{t+1}}\right)^{\frac{1}{\sigma}} P_t C_t$$

Labour supply L_t :

$$\frac{W_t C_t^{-\sigma}}{P_t} = \alpha \left(1 - L_t\right)^{-\nu}$$

Combine the first order conditions with respect to $B_{H,t+1}$ and $B^*_{H,t+1}$ or $B^*_{F,t+1}$ and $B_{F,t+1}$ to get the international risk sharing condition:

$$\frac{\varepsilon_t P_t^* \left(C_t^*\right)^\sigma}{P_t \left(C_t\right)^\sigma} = \frac{\beta_t E_t \left[\frac{1}{P_{t+1}(C_{t+1})^\sigma}\right]}{\beta_t^* E_t \left[\frac{1}{\varepsilon_{t+1} P_{t+1}^* \left(C_{t+1}^*\right)^\sigma}\right]} = \frac{\beta_t E_t \left[\frac{\varepsilon_{t+1}}{P_{t+1}(C_{t+1})^\sigma}\right]}{\beta_t^* E_t \left[\frac{1}{P_{t+1}^* \left(C_{t+1}^*\right)^\sigma}\right]}$$

1.A.2 Solve the producer's problem with quadratic investment adjustment costs

$$\max_{\left\{P_\iota^*(h_d), k_{\iota+1}^*(h_d)\right\}_{\iota=t}^{\infty}} \quad E_t \sum_{\iota=t}^{\infty} Q_{t,t+\iota} \left\{ \begin{array}{c} \left[\varepsilon_t P_t^*(h_d) - (1+\tau) MC_{H,t}\right] D_t^*(h_d) \\ - \left[i_\iota^*(h_d) + \frac{\gamma\left(i_\iota^*(h_d) - \bar{i}\right)^2}{2k_\iota^*(h_d)}\right] \varepsilon_\iota P_{N,\iota}^* \end{array} \right\}$$

subject to

$$\begin{aligned} k_{t+1}^*(h_d) &= i_t^*(h_d) + (1-\delta) k_t^*(h_d) - \chi_2\left(k_t^*(h_d)\right) D_t^*(h_d) \quad \left(Q_{t,t+\iota} \varepsilon_t P_{N,\iota}^* \lambda_{1,t}^*(h_d)\right) \\ D_t^*(h_d) &= \left(\frac{k_t^*(h_d)}{K_{H,t}^*}\right)^{\frac{1}{\theta_K}} \left(\frac{P_t^*(h_d)}{P_{H,t}^*}\right)^{-\theta_H} C_{H,t}^* \quad \left(Q_{t,t+\iota} \varepsilon_t P_{N,\iota}^* \lambda_{2,t}^*(h_d)\right) \\ MC_{H,t} &= \frac{W_t}{Z_{H,t}}, \end{aligned}$$

where the Lagrangian multiplier for the corresponding constraint is represented in the bracket. The optimal pricing function and its corresponding retail capacity are derived by taking first order conditions with respect to the following variables:

$P_t^*(h_d)$:

$$\varepsilon_t P_t^*(h_d) = \theta_{h_d} \varepsilon_t P_{N,t}^* \lambda_{2,t}^*(h_d)$$

$D_t^*(h_d)$:

$$\left[\varepsilon_t P_t^*(h_d) - (1+\tau) MC_{H,t}\right] = -P_{N,t}^* \varepsilon_t \lambda_{1,t}^*(h_d) \chi_2\left(k_t^*(h_d)\right) - \lambda_{2,t}^*(h_d) \varepsilon_t P_{N,t}^*$$

$i_t^*(h_d)$:

$$1 + \frac{\gamma\left(i_t^*(h_d) - \bar{i}\right)}{k_t(h_d)} = -\lambda_{1,t}^*(h_d)$$

where $-\lambda_{1,t}^*(h_d)$ measures the marginal cost of adjusting capacity.

$k_{t+1}^*(h_d)$:

$$-\varepsilon_t P_{N,t}^* \lambda_{1,t}^*(h_d) = E_t Q_{t,t+1} \left\{ \begin{array}{c} \varepsilon_{t+1} P_{N,t+1}^* \frac{\gamma\left(i_{t+1}^*(h_d) - \bar{i}\right)^2}{2k_{t+1}^*(h_d)^2} - \frac{\varepsilon_{t+1} P_{t+1}^*(h_d)}{\theta_H} \frac{D_{t+1}^*(h_d)}{\theta_K k_{t+1}^*(h_d)} \\ -P_{N,t+1}^* \varepsilon_{t+1} \lambda_{1,t+1}^*(h_d) \left[(1-\delta) - \chi_2'\left(k_{t+1}^*(h_d)\right) D_{t+1}^*(h_d)\right] \end{array} \right\}$$

Combine the first order conditions with respect to $P_t^*(h_d)$ and $D_t^*(h_d)$ to get the optimal price:

$$P_t^*(h_d) = \frac{\theta_H}{\theta_H - 1}\left[\frac{MC_{H,t}}{\varepsilon_t} - \lambda_{1,t}^*(h_d)\chi_2(k_t^*(h_d))P_{N,t}^*\right]$$

1.A.3 Solve the local retailer's problem with quadratic investment adjustment costs

$$\max \quad E_t \sum_{\iota=t}^{\infty} Q_{t,t+\iota} \left\{ \begin{array}{c} \int_x^1 \left[P_\iota^*(h_r,r) - P_{im,\iota}^*(h_r,r)\right] D_\iota^*(h_r,r)dh_r + \\ \int_x^1 \left[P_\iota^*(f_r,r) - P_{local,\iota}^*(f_r,r)\right] D_\iota^*(f_r,r)df_r \\ -\left[i_\iota^*(r) + \frac{\gamma\left(i_\iota^*(r)-\bar{i}\right)^2}{2k_\iota(r)}\right] P_{N,\iota}^* \end{array} \right\}$$

subject to

$$\begin{aligned}
k_{t+1}^*(r) &= i_t^*(r) + (1-\delta)k_t^*(r) \\
&\quad - \chi_2(k_t^*(r)) \int_x^1 D_\iota^*(h_r,r)dh_r - \chi(k_t^*(r)) \int_x^1 D_\iota^*(f_r,r)df_r \quad \left(Q_{t,t+\iota}P_{N,\iota}^*\lambda_{1,t}^*(r)\right) \\
D_t^*(h_r,r) &= \left(\frac{k_t^*(r)}{K_{H,t}^*}\right)^{\frac{1}{\theta_K}} \left(\frac{P_\iota^*(h_r,r)}{P_{H,t}^*}\right)^{-\theta_H} C_{H,t}^* \quad \left(Q_{t,t+\iota}P_{N,\iota}^*\lambda_{2,t}^*(h_r,r)\right) \\
D_t^*(f_r,r) &= \left(\frac{k_t^*(r)}{K_{F,t}^*}\right)^{\frac{1}{\theta_K}} \left(\frac{P_\iota^*(f_r,r)}{P_{F,t}^*}\right)^{-\theta_H} C_{F,t}^* \quad \left(Q_{t,t+\iota}P_{N,\iota}^*\lambda_{2,t}^*(f_r,r)\right),
\end{aligned}$$

where the Lagrangian multiplier for the corresponding constraint is represented in the bracket. The optimal pricing function and its corresponding retail capacity are derived by taking first order conditions with respect to the following variables:

$P_t^*(h_r,r)$:

$$P_t^*(h_r,r) = \theta_H P_{N,\iota}^*\lambda_{2,t}^*(h_r,r)$$

$P_t^*(f_r,r)$:

$$P_t^*(f_r,r) = \theta_H \varepsilon_t P_{N,t}^*\lambda_{2,t}^*(f_r,r)$$

$D_t^*(h_r,r)$:

$$\left[P_t^*(h_r,r) - P_{im,t}^*(h_r,r)\right] = -P_{N,t}^*\lambda_{1,t}^*(r)\chi_2(k_t^*(h_r,r)) - \lambda_{2,t}^*(h_r,r)P_{N,t}^*$$

$D_t^*(f_r,r)$:

$$\left[P_t^*(f_r,r) - P_{local,t}^*(f_r,r)\right] = -P_{N,t}^*\lambda_{1,t}^*(r)\chi(k_t^*(f_r,r)) - \lambda_{2,t}^*(f_r,r)P_{N,t}^*$$

$i_t^*(r)$:

$$1+\frac{\gamma\left(i_t^*(r)-\bar{i}\right)}{k_t(r)}=-\lambda_{1,t}^*(r)$$

$k_{t+1}^*(r)$:

$$-P_{N,t}^*\lambda_{1,t}^*(r)=E_tQ_{t,t+1}\left\{\begin{array}{c} P_{N,t+1}^*\frac{\gamma\left(i_{t+1}^*(r)-\bar{i}\right)^2}{2k_{t+1}^*(r)^2}-\int_x^1\frac{P_{t+1}^*(h_r,r)}{\theta_H}\frac{D_{t+1}^*(h_r,r)}{\theta_K k_{t+1}^*(r)}dh_r \\ -\int_x^1\frac{P_{t+1}^*(f_r,r)}{\theta_H}\frac{D_{t+1}^*(f_r,r)}{\theta_K k_{t+1}^*(r)}df_r-P_{N,t+1}^*\lambda_{1,t+1}^*(r)(1-\delta) \\ P_{N,t+1}^*\lambda_{1,t+1}^*(r)\left[\begin{array}{c}\chi_2'\left(k_{t+1}^*(r)\right)\int_x^1 D_{t+1}^*(h_r,r)dh_r \\ +\chi'\left(k_{t+1}^*(r)\right)\int_x^1 D_{t+1}^*(f_r,r)df_r\end{array}\right]\end{array}\right\}$$

Combine the first order conditions with respect to $P_t^*(h)$ and $D_t^*(h)$ to get the optimal price:

$$P_t^*(h_r,r)=\frac{\theta_H}{\theta_H-1}\left[P_{im,t}(h_r,r)-\lambda_{1,t}^*(r)\chi_2(k_t^*(h_r,r))P_{N,t}^*\right]$$

1.A.4 Aggregations and resource constraints

Aggregate consumptions, retail capacity and prices are aggregated in the same way as in the basic model. The only difference is that we need to take into account the proportion of two groups of firms.

$$\begin{aligned}
C_{H,t} &\equiv \left[\int_0^x\left(\frac{k_t(h_d)}{K_{H,t}}\right)^{\frac{1}{\theta_H\theta_K}}C_t(h_d)^{\frac{\theta_H-1}{\theta_H}}dh_d+\int_x^1\left(\frac{k_t(h_r)}{K_{H,t}}\right)^{\frac{1}{\theta_H\theta_K}}C_t(h_r)^{\frac{\theta_H-1}{\theta_H}}dh_r\right]^{\frac{\theta_H}{\theta_H-1}} \\
C_{F,t} &\equiv \left[\int_0^x\left(\frac{k_t(f_d)}{K_{F,t}}\right)^{\frac{1}{\theta_H\theta_K}}C_t(f_d)^{\frac{\theta_H-1}{\theta_H}}df_d+\int_x^1\left(\frac{k_t(f_r)}{K_{F,t}}\right)^{\frac{1}{\theta_H\theta_K}}C_t(f_r)^{\frac{\theta_H-1}{\theta_H}}df_r\right]^{\frac{\theta_H}{\theta_H-1}} \\
K_{H,t} &\equiv \left[\int_0^x k_t(h_d)^{\frac{1}{\theta_K}}dh_d+\int_x^1 k_t(h_r)^{\frac{1}{\theta_K}}dh_r\right]^{\theta_K}, \\
K_{F,t} &\equiv \left[\int_0^x k_t(f_d)^{\frac{1}{\theta_K}}df_d+\int_x^1 k_t(f_r)^{\frac{1}{\theta_K}}df_r\right]^{\theta_K} \\
P_{H,t} &= \left[\int_0^x\left(\frac{k_t(h_d)}{K_{H,t}}\right)^{\frac{1}{\theta_K}}p_t(h_d)^{1-\theta_H}dh_d+\int_x^1\left(\frac{k_t(h_r)}{K_{H,t}}\right)^{\frac{1}{\theta_K}}p_t(h_r)^{1-\theta_H}dh_r\right]^{\frac{1}{1-\theta_H}}, \\
P_{F,t} &= \left[\int_0^x\left(\frac{k_t(f_d)}{K_{F,t}}\right)^{\frac{1}{\theta_K}}p_t(f_d)^{1-\theta_H}df_d+\int_x^1\left(\frac{k_t(f_r)}{K_{F,t}}\right)^{\frac{1}{\theta_K}}p_t(f_r)^{1-\theta_H}df_r\right]^{\frac{1}{1-\theta_H}}
\end{aligned}$$

At the aggregate level the investment of the retail capacity for home tradables $I_{DH,t}$, home retailers $I_{R,t}$ and foreign tradables $I_{DF,t}$ are given by:

$$\begin{aligned}
I_{DH,t} &= \int_0^x \left[k_{t+1}(h_d) + \chi(k_t(h_d))C_t(h_d) - (1-\delta)k_t(h_d)\right] dh_d \\
I_{DF,t} &= \int_0^x \left[k_{t+1}(f_d) + \chi_2(k_t(f_d))C_t(f_d) - (1-\delta)k_t(f_d)\right] df_d \\
I_{R,t} &= K_{R,t+1} + \chi_2(K_{R,t})C_{FR,t} + \chi(K_{R,t})C_{HR,t} - (1-\delta)K_{R,t}
\end{aligned}$$

The demand for nontradable goods is consumer demand for the nontradables plus the sum of the investment of retail capacity for home tradables and foreign tradables.

$$D_t(n) = \left(\frac{p_t(n)}{P_{N,t}}\right)^{-\theta_N} \left[C_{N,t} + I_{DF,t} + I_{DH,t} + I_{R,t}\right]$$

The optimal price is given by

$$p_t(n) = P_{N,t} = \frac{\theta_N}{\theta_N - 1} MC_{N,t}$$

From the aggregate resource constraint, the aggregate labour used is equal to the per unit labour cost times the total quantity produced.

$$L_t = \frac{Y_{N,t}}{Z_{N,t}} + \frac{Y_{H,t}}{Z_{H,t}} = \frac{I_{DH,t} + I_{DF,t} + I_{R,t} + C_{N,t}}{Z_{N,t}} + \frac{C_{H,t} + C_{H,t}^*}{Z_{H,t}}$$

1.A.5 The current account

The stochastic discount factor $Q_{t,t+1}$ is defined such that the expectation of this discount factor is just the inverse of the nominal interest rate[29].

$$Q_{t,t+1} \equiv \beta_t \frac{P_t C_t^\sigma}{P_{t+1} C_{t+1}^\sigma}, EQ_{t,t+1} = \frac{1}{1+i_{t+1}}, E\left[Q_{t,t+1}\frac{\varepsilon_{t+1}}{\varepsilon_t}\right] = \frac{1}{1+i_{t+1}^*}$$

The following three equations describe the equilibrium conditions of the international bond holding and balance of the current account.

$$B_{H,t} = -B_{H,t}^*, \quad B_{F,t} = -B_{F,t}^*$$

$$\overline{A_t} \equiv A_{t+1} - M_t = B_{H,t} + \varepsilon_t B_{F,t} = -B_{H,t}^* - \varepsilon_t B_{F,t}^* = -\varepsilon_t \left(\frac{B_{H,t}^*}{\varepsilon_t} + B_{F,t}^*\right) = -\varepsilon_t \overline{A_t^*}$$

[29] See the first order condition of the nominal bond holding for the consumer's optimisation problem in the appendix.

When calculating the price of imports and exports, I subtract the distribution cost from the consumer price for the price of direct exporters.

$$\begin{aligned} NX_t &= \varepsilon_t \left(P^*_{HD,t} - \chi_2 \left(K^*_{HD,t+1} \right) P^*_{N,t} \right) C^*_{HD,t} + \varepsilon_t P^*_{IM,t} C^*_{HR,t} \\ &\quad - \left(P_{FD,t} - \chi_2 \left(K_{FD,t+1} \right) P_{N,t} \right) C_{FD,t} - P_{IM,t} C_{FR,t} \end{aligned}$$

The balance of trade is simply adding the net export to the capacity account.

$$E_t \left\{ Q_{t,t+1} \overline{A}_{t+1} \right\} = \overline{A}_t + NX_t$$

Appendix 1.B Summary of Equilibrium Conditions

The world equilibrium is characterised as follows. Given the strict inflation targeting monetary policy, the stochastic process of the productivity shock, and the initial conditions of bond holding and money holding, the equilibrium is given by the above set of endogenous variables which satisfy (a) the consumer's optimisation constraints such that home and foreign representative households maximise their utility; (b) producers' optimal price choice and retailers' optimal capacity choice such that producers and retailers in the home and foreign country maximise their profit; (c) market clear constraints and (d) aggregate resource constraints. The summary of equilibrium conditions of the home country is given as follows, the corresponding foreign equilibrium conditions can be easily derived:

Table 1.B.1 Summary of equilibrium conditions

Endogenous discount factor of consumer preference:	$\beta_t \equiv \log \left\{ \zeta_2 \left[1 + \zeta_1 \left(C_t + \alpha \left(1 - L_t \right) \right) \right] \right\}$
Definition of the stochastic discount factor:	$Q_{t,t+1} \equiv \beta_t \frac{P_t C_t^{\sigma}}{P_{t+1} C_{t+1}^{\sigma}}$
Accounting identity for retail capacity investment:	$I_{HD,t} = K_{HD,t+1} + \chi(K_{HD,t}) C_{HD,t} - (1-\delta) K_{HD,t}$ $I_{FD,t} = K_{FD,t+1} + \chi_2(K_{FD,t}) C_{FD,t} - (1-\delta) K_{FD,t}$ $I_{R,t} = K_{R,t+1} + \chi(K_{R,t}) C_{HR,t} + \chi_2(K_{R,t}) C_{FR,t} - (1-\delta) K_{R,t}$
Aggregate resource constraint:	$L_t = \frac{Y_{N,t}}{Z_{N,t}} + \frac{Y_{H,t}}{Z_{H,t}} = \frac{I_{HD,t} + I_{FD,t} + I_{R,t} + C_{N,t}}{Z_{N,t}} + \frac{C_{H,t} + C^*_{H,t}(1+\tau)}{Z_{H,t}}$

Continued on next page...

Table 1.B.1: Summary of equilibrium conditions – continued

Labour equilibrium:	$\frac{W_t C_t^{-\sigma}}{P_t} = \alpha\left(1-L_t\right)^{-\nu}$
Retail capacity aggregation:	$K_{H,t} \equiv \left[x K_{HD,t}^{\frac{1}{\theta_K}} + (1-x) K_{R,t}^{\frac{1}{\theta_K}}\right]^{\theta_K}$, $K_{F,t} \equiv \left[x K_{FD,t}^{\frac{1}{\theta_K}} + (1-x) K_{R,t}^{\frac{1}{\theta_K}}\right]^{\theta_K}$
Retail market share:	$S_{H,t} = \frac{(K_{H,t})^{\frac{1}{\theta_K}}}{(K_{H,t})^{\frac{1}{\theta_K}} + (K_{F,t})^{\frac{1}{\theta_K}}}, S_{F,t} = \frac{(K_{F,t})^{\frac{1}{\theta_K}}}{(K_{H,t})^{\frac{1}{\theta_K}} + (K_{F,t})^{\frac{1}{\theta_K}}}$ $S_{HD,t} = \left(\frac{K_{HD,t}}{K_{H,t}}\right)^{\frac{1}{\theta_K}}, S_{FD,t} = \left(\frac{K_{FD,t}}{K_{F,t}}\right)^{\frac{1}{\theta_K}}$
Consumer's demand:	$C_{N,t} = \left(\frac{P_{N,t}}{P_t}\right)^{-\phi} C_t, C_{T,t} = \left(\frac{P_{T,t}}{P_t}\right)^{-\phi} C_t$ $C_{H,t} = S_{H,t}\left(\frac{P_{H,t}}{P_{T,t}}\right)^{-\rho} C_{T,t}, C_{F,t} = S_{F,t}\left(\frac{P_{F,t}}{P_{T,t}}\right)^{-\rho} C_{T,t}$ $C_{HD,t} = x S_{HD,t}\left(\frac{P_{HD,t}}{P_{H,t}}\right)^{-\rho} C_{H,t}, C_{FD,t} = x S_{FD,t}\left(\frac{P_{FD,t}}{P_{F,t}}\right)^{-\rho} C_{F,t}$ $C_{HR,t} = (1-x)\left(1-S_{HD,t}\right)\left(\frac{P_{HR,t}}{P_{H,t}}\right)^{-\theta_H} C_{H,t}$, $C_{FR,t} = (1-x)\left(1-S_{FD,t}\right)\left(\frac{P_{FR,t}}{P_{F,t}}\right)^{-\theta_H} C_{F,t}$
Price aggregation:	$P_{T,t} = \left[S_{H,t} P_{H,t}^{1-\rho} + S_{F,t} P_{F,t}^{1-\rho}\right]^{\frac{1}{1-\rho}}, P_t = \left[P_{T,t}^{1-\phi} + P_{N,t}^{1-\phi}\right]^{\frac{1}{1-\phi}}$ $P_{H,t} = \left[x S_{HD,t} P_{HD,t}^{1-\theta_H} + (1-x)\left(1-S_{HD,t}\right) P_{HR,t}^{1-\theta_H}\right]^{\frac{1}{1-\theta_H}}$ $P_{F,t} = \left[x S_{FD,t} P_{FD,t}^{1-\theta_H} + (1-x)\left(1-S_{FD,t}\right) P_{FR,t}^{1-\theta_H}\right]^{\frac{1}{1-\theta_H}}$
Current account:	$E_t\left\{Q_{t,t+1}\overline{A}_{t+1}\right\} = \overline{A}_t + CA_t$ $CA_t = \varepsilon_t\left(P_{HD,t}^* - \chi_2\left(K_{HD,t+1}^*\right)\right) C_{HD,t}^* + \varepsilon_t P_{IM,t}^* C_{HR,t}^* - \left(P_{FD,t} - \chi_2\left(K_{FD,t+1}\right)\right) C_{FD,t} - P_{IM,t} C_{FR,t}$

Continued on next page...

Table 1.B.1: Summary of equilibrium conditions – continued

International risk sharing condition:	$\frac{\varepsilon_t P_t^*(C_t^*)^{\sigma}}{P_t(C_t)^{\sigma}} = \frac{\beta_t E_t\left[\frac{1}{P_{t+1}(C_{t+1})^{\sigma}}\right]}{\beta_t^* E_t\left[\frac{1}{\varepsilon_{t+1}P_{t+1}^*\left(C_{t+1}^*\right)^{\sigma}}\right]} = \frac{\beta_t E_t\left[\frac{\varepsilon_{t+1}}{P_{t+1}(C_{t+1})^{\sigma}}\right]}{\beta_t^* E_t\left[\frac{1}{P_{t+1}^*\left(C_{t+1}^*\right)^{\sigma}}\right]}$
Optimal retail capacity:	$-P_{N,t}\lambda_{1,t}^R = E_t Q_{t,t+1}\left\{\begin{array}{c} P_{N,t+1}\frac{\gamma\left(I_{R,t+1}-\bar{i}\right)^2}{2K_{R,t+1}{}^2} - \frac{P_{HR,t+1}}{\theta_H}\frac{D_{HR,t+1}}{\theta_K K_{R,t+1}} \\ -\frac{P_{FR,t+1}}{\theta_H}\frac{D_{FR,t+1}}{\theta_K K_{R,t+1}} - P_{N,t+1}\lambda_{1,t+1}^R(1-\delta) \\ P_{N,t+1}\lambda_{1,t+1}^R\left[\begin{array}{c}\chi'\left(K_{R,t+1}\right)D_{HR,t+1} \\ +\chi_2'\left(K_{R,t+1}\right)D_{FR,t+1}\end{array}\right]\end{array}\right\}$ $-\frac{P_{N,t}\lambda_{1,t}^{FD}}{\varepsilon_t} = E_t Q_{t,t+1}\left\{\begin{array}{c} \frac{P_{N,t+1}}{\varepsilon_{t+1}}\frac{\gamma\left(I_{FD,t+1}-\bar{I}_{FD}\right)^2}{2(K_{FD.t+1})^2} - \frac{P_{FD,t+1}}{\varepsilon_{t+1}\theta_H}\frac{C_{FD,t+1}}{\theta_K K_{FD,t+1}} \\ -\frac{P_{N,t+1}\lambda_{1,t+1}^{FD}}{\varepsilon_{t+1}}\left[(1-\delta)-\chi_2'\left(K_{FD,t+1}\right)C_{FD,t+1}\right]\end{array}\right\}$ $-P_{N,t}\lambda_{1,t}^{HD} = E_t Q_{t,t+1}\left\{\begin{array}{c} P_{N,t+1}\frac{\gamma\left(I_{HD,t+1}-\bar{I}_{HD}\right)^2}{2(K_{HD.t+1})^2} - \frac{P_{HD,t+1}}{\theta_H}\frac{C_{HD,t+1}}{\theta_K K_{HD,t+1}} \\ -P_{N,t+1}\lambda_{1,t+1}^{HD}\left[(1-\delta)-\chi'\left(K_{HD,t+1}\right)C_{HD,t+1}\right]\end{array}\right\}$ $1+\frac{\gamma\left(I_{FD,t}-\bar{I}_{FD}\right)}{K_{FD,t}} = -\lambda_{1,t}^{FD}$ $1+\frac{\gamma\left(I_{HD,t}-\bar{I}_{HD}\right)}{K_{HD,t}} = -\lambda_{1,t}^{HD}$ $1+\frac{\gamma\left(I_{R,t}-\bar{I}_{R}\right)}{K_{R,t}} = -\lambda_{1,t}^{R}$
Optimal price setting:	$P_{FD,t} = \frac{\theta_H}{\theta_H-1}\left[\varepsilon_t MC_{F,t}(1+\tau) - \lambda_{1,t}^{FD}\chi_2(K_{FD,t})P_{N,t}\right]$ $P_{HD,t} = \frac{\theta_H}{\theta_H-1}\left[MC_{H,t} - \lambda_{1,t}^{HD}\chi(K_{HD,t})P_{N,t}\right]$ $P_{HR,t} = \frac{\theta_H}{\theta_H-1}\left[P_{L,t} - \lambda_{1,t}^{R}\chi(K_{R,t})P_{N,t}\right]$ $P_{FR,t} = \frac{\theta_H}{\theta_H-1}\left[P_{IM,t} - \lambda_{1,t}^{R}\chi_2(K_{R,t})P_{N,t}\right]$ $P_{L,t} = \frac{\theta_H}{\theta_H-1}\left[MC_{H,t} - \frac{1}{\theta_H}\lambda_{1,t}^{R}\chi(K_{R,t})P_{N,t}\right]$

Continued on next page...

Table 1.B.1: Summary of equilibrium conditions – continued

	$P_{IM,t} = \frac{\theta_H}{\theta_H - 1}\left[\varepsilon_t MC_{F,t}\left(1+\tau\right) - \frac{1}{\theta_H}\lambda^R_{1,t}\chi 2(K_{R,t})P_{N,t}\right]$
	$P_{N,t} = \frac{\theta_N}{\theta_N - 1} MC_{N,t}$
Marginal costs:	$MC_{N,t} = \frac{W_t}{Z_{N,t}}, MC_{H,t} = \frac{W_t}{Z_{H,t}}$

Table 1.B.2 List of aggregate variables

Variables	Description	Simulation Form
ε_t	Nominal exchange rate	`q`
P_t	CPI	`p`
$P_{T,t}$	Price index of tradable goods	`pt`
$P_{F,t}$	Consumer price index of imports	`pf`
$P_{H,t}$	Consumer price index of home produced goods	`ph`
$P_{N,t}$	Price index of nontradable goods	`pn`
$P_{HD,t}$	Consumer price of goods produced by home retailing manufacturers (Subscript D means distributing by itself)	`phd`
$P_{HR,t}$	Consumer price of goods produced by home non-retailing manufacturers (Subscript R means selling to local retailer)	`phr`
$P_{FD,t}$	Consumer price of goods imported from foreign retailing manufacturers (Subscript D means distributing by itself)	`pfd`
$P_{FR,t}$	Consumer price of goods imported from foreign non-retailing manufacturers (Subscript R means selling to local retailer)	`pfr`
$P_{L,t}$	Wholesale price of goods produced by home retailing manufacturers	`pl`
$P_{IM,t}$	Wholesale price of goods produced by home non-retailing manufacturers	`pim`
C_t	Aggregate consumption	`c`
$C_{T,t}$	Consumptions of tradable goods	`ct`
$C_{H,t}$	Consumptions of home produced goods	`ch`
$C_{F,t}$	Consumptions of imports	`cf`
$C_{N,t}$	Consumptions of nontradable goods	`cn`
$C_{HD,t}$	Consumptions of goods produced by home retailing manufacturers	`chd`
$C_{HR,t}$	Consumptions of goods produced by home non-retailing manufacturers	`chr`
$C_{FD,t}$	Consumptions of goods imported from foreign retailing manufacturers	`cfd`
$C_{FR,t}$	Consumptions of goods imported from foreign non-retailing manufacturers	`cfr`
$MC_{N,t}$	Marginal cost of nontradable good producers	`mcn`
$MC_{H,t}$	Marginal cost of domestic tradable good producers	`mch`
$K_{HD,t}$	retail capacity of home retailing manufacturers	`kdh`
$K_{FD,t}$	retail capacity of foreign retailing manufacturers	`kdf`
$K_{R,t}$	retail capacity of retailers	`kr`
$I_{HD,t}$	Investment made by home retailing manufacturers	`idh`
$I_{FD,t}$	Investment made by foreign retailing manufacturers	`idf`
$I_{R,t}$	Investment made by retailers	`ir`
L_t	Equilibrium employment	`l`
W_t	Nominal wage	`w`
$Q_{t,t+1}$	Stochastic discount factor	`dis`
$\overline{A}_t$	Net international bond holding	`a`
REX_t	Real exchange rate $\equiv \frac{\varepsilon_t P_t^*}{P_t}$	`rex`
TOT_t	Terms of trade $\equiv \frac{\overline{P}_{F,t}}{\varepsilon_t \overline{P}^*_{H,t}}$	`tot`

Note: This list presents the simulation form of home variables. The simulation forms of relative prices are measure by the foreign variable over the home variable. Simulation forms of foreign variables are denoted with a suffix “s”.

Table 1.B.3 Calibration

Parameters	Description	Benchmark	Calibration for the basic model
β	Discount factor	0.99	0.99
δ	Depreciation rate on retail capacity	0.025	0.025
ν	Disutility of labour	2	2
σ	Risk aversion	2	2
ρ	Elasticity of substitution between home and foreign tradable goods	2	2
ϕ	Elasticity of substitution between tradable and nontradable goods	0.74	0.74
θ_H	Elasticity of substitution within home tradable goods	15.3	15.3
θ_N	Elasticity of substitution within nontradable goods	7.7	7.7
θ_K	Elasticity of retail aggregation factor	1.4	1.4
φ_H	Effectiveness of retail capacity on distributing home products	0.3	0.3
φ_F	Effectiveness of retail capacity on distributing foreign products	0.33	0.3
ϑ	Fixed distribution cost per unit	0.5	0
Ξ	Proportion of retailing manufacturers	0.2	–
τ	Iceberg trade cost	0.03	–
γ	Frictions on retail capacity adjustment	0.00	10
ζ_1	Parameter 1 of the endogenous discount factor	-0.005	–
ζ_2	Parameter 2 of the endogenous discount factor	3	–

Note: "Benchmark + Retail Capacity Adjustment Friction" in the paper defers to the benchmark calibration with frictions on retail capacity adjustment $\gamma = 10$.

Appendix 1.C Statistics and Graphs

Table 1.C.1 Key statistics

Statistics	U.S. data	Benchmark	Benchmark + Retail Capacity Adjustment Friction	BKK
Business Cycle Correlations				
rGDP, Consumption	0.91	0.97	1.00	0.79
rGDP, Employment	0.85	0.89	0.99	0.94
rGDP, Investment	0.93	0.92	0.96	0.27
rGDP, Net exports	-0.69	-0.26	-0.21	-0.02
TOT, Net exports	-0.17	0.07	-0.06	-0.84
$C/C^*, REX$	-0.08	-0.01	0.00	0.98
Volatility (standard deviation) relative to rGDP				
Consumption	0.85	0.97	1.07	0.79
Employment	0.75	0.45	0.59	0.47
Investment	3.10	–	–	10.94
$(a)\ I_{D,H}$	–	2.79	0.49	–
$(b)\ I_{D,F}$	–	4.92	0.92	–
$(c)\ I_R$	–	3.20	0.76	–
Net exports	0.29	0.11	0.86	2.90
International Correlations				
$Z_{H,t}, Z_{F,t}$		0.42	0.47	0.30
$Z_{N,t}, Z^*_{N,t}$		-0.08	-0.05	–
Real GDP	0.68	0.34	0.58	-0.18
Consumption	0.46	0.29	0.65	0.88
labour	0.42	0.13	0.64	0.47
P_{import}, P_{export}	0.89	0.36	0.00	-1.00
P_{import}, REX	0.57	0.38	-0.34	-1.00
P_{export}, REX	0.58	0.33	0.85	1.00
TOT, REX	0.47	0.48	0.99	1.00
Avg. Distribtuion Margin				
Home retailing manufacturers		0.36	0.36	
Foreign retailing manufacturers at Home		0.47	0.47	
Home local retailer selling home product		0.38	0.38	
Home local retailer selling foreign product		0.38	0.38	

Note: Statistics are calculated based on logged & HP-filtered quarterly time series. The "U.S. data" column presents statistics calculated during the period 1995:1-2012:12. Data sources can be found in table 1.C.4.

Table 1.C.2 Cyclical properties

Data	*Autocorrelations*									
		(-1)	(-2)	(-3)	(-4)					
	REX	0.80	0.55	0.34	0.14					
	rGDP	0.78	0.60	0.40	0.19					
	Cross Correlations with Output									
		(-4)	(-3)	(-2)	(-1)	(0)	(+1)	(+2)	(+3)	(+4)
	Consumption	0.08	0.24	0.42	0.57	0.68	0.63	0.51	0.36	0.18
	Hours Worked	-0.20	0.01	0.29	0.55	0.73	0.75	0.67	0.56	0.45
	Investment	0.45	0.64	0.79	0.91	0.96	0.88	0.73	0.55	0.37
Benchmark Model	*Autocorrelations*									
		(-1)	(-2)	(-3)	(-4)					
	REX	0.71	0.46	0.26	0.11					
	rGDP	0.68	0.42	0.22	0.06					
	Cross Correlations with Output									
		(-4)	(-3)	(-2)	(-1)	(0)	(+1)	(+2)	(+3)	(+4)
	Consumption	0.13	0.28	0.47	0.70	0.96	0.65	0.39	0.19	0.03
	Hours Worked	0.22	0.35	0.50	0.66	0.83	0.55	0.31	0.12	-0.02
	Investment	-0.07	0.06	0.25	0.52	0.94	0.67	0.44	0.25	0.11

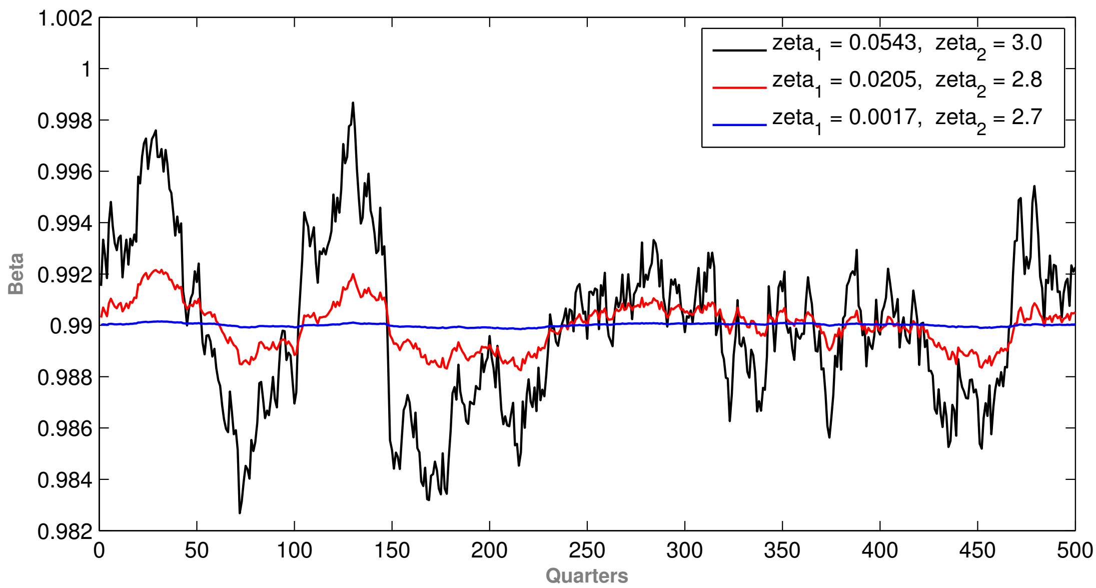
Note: Statistics are calculated based on logged and HP-filtered quarterly seasonal adjusted private final consumption expenditure at constant prices and gross domestic products at constant prices from periods 1995:1-2013:2. Data source: OECD Main Economic Indicators.

Table 1.C.3 Estimated ERPT from simulated data

Prices	Benchmark	
	Short Run	Long Run
P^*_{HD}	0.3794	0.4746
	(0.0062)	
P^*_{HR}	0.4594	0.5178
	(0.0047)	
P^*_{IM}	0.9406	0.9471
	(0.0005)	
P^*_{H}	0.4532	0.5153
	(0.0048)	

Note: Standard errors are reported in parentheses.

Fig. 1.C.1 Plot of the endogenous discount factor



1.C.1 Empirical distribution margin, ERPT and international business cycle estimates

Table 1.C.4 Data sources

Series	Frequency	Periods	Source
Export, Import Prices of the United States	Quarterly	1990:1-2013:2	Export/import price index all commodities, the U.S. Bureau of Labour Statistics
Real Effective Exchange Rate; Bilateral Nominal Exchange Rate	Quarterly; Monthly	1990:1-2013:2	International Financial Statistics, IMF
Terms of Trade	Quarterly	1990:1-2013:2	Datastream
Business Cycle Series: Consumption, Investment, Total Hours Worked, Net Exports	Quarterly	1995:1-2013:2	OECD Main Economic Indicators; OECD Economic Outlook
Distribution Margins	Annual	1995-2010	Supply Table at current prices of the National Accounts, Eurostat Database
Sectoral Retail Price Margins	Monthly	2009:3-2013:8	The U.S. Bureau of Labour Statistics
Bilateral Import/Export Unit Value Indices	Monthly	1995:1-2001:12	COMEXT (Eurostat) Database

Table 1.C.5 Variances of retail price margins in the United States

Industries	Variance (%)
Food and alcohol retailing	0.0353
Health and beauty care retailing, including optical goods	0.1098
Apparel, jewellery, footwear, and accessories retailing	0.1400
Computer hardware, software, and supplies retailing	1.0974
TV, video, and photographic equipment and supplies retailing	1.0658
Automobiles and automobile parts retailing	0.0539
Manufactured (mobile) homes retailing	0.0813
RVs, trailers, and campers retailing	0.0592
Sporting goods, including boats, retailing	0.0422
Lawn, garden, and farm equipment and supplies retailing	0.0525
Furniture retailing	0.1889
Flooring and floor coverings retailing	0.1820
Hardware and building materials and supplies retailing	0.2181
Major household appliance retailing	1.6423
Fuels and lubricants retailing	1.4924
Cleaning supplies and paper products retailing	0.1188
Book retailing	0.0451
Other merchandise retailing (partial)	0.0307
All products	0.0199

Note: Statistics are calculated based on Retail Producer Price Indexes of the U.S. Bureau of Labour Statistics from 2009:3 to 2013:8.

Table A: The Estimated Distribution Margin in the Uinted Kingdom from 1997 to 2009

Industries \ Year	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Max	Min	Average
Basic metals	0.1655	0.1757	0.2003	0.1608	0.1675	0.1613	0.1590	0.1791	0.1667	0.1549	0.1553	0.1715	0.1475	0.2003	0.1475	0.1665
Basic pharmaceutical products and pharmaceutical preparations	0.2420	0.2663	0.2759	0.2688	0.2781	0.2728	0.2783	0.2732	0.2694	0.2677	0.2586	0.2514	0.2426	0.2783	0.2420	0.2650
Chemicals and chemical products	0.1980	0.2270	0.2399	0.2421	0.2349	0.2404	0.2503	0.2447	0.2375	0.2365	0.2330	0.2393	0.2428	0.2503	0.1980	0.2359
Coke and refined petroleum products	0.0240	0.0247	0.0250	0.0253	0.0331	0.0305	0.0355	0.0349	0.0319	0.0329	0.0324	0.0319	0.0321	0.0355	0.0240	0.0303
Computer, electronic and optical products	0.1860	0.2044	0.2046	0.1915	0.1970	0.1923	0.2091	0.2181	0.1957	0.1737	0.2364	0.2293	0.2389	0.2389	0.1737	0.2059
Electrical equipment	0.2152	0.2370	0.2312	0.2293	0.2488	0.2435	0.2455	0.2470	0.2421	0.2384	0.2324	0.2378	0.2455	0.2488	0.2152	0.2380
Fabricated metal products	0.1382	0.1509	0.1596	0.1626	0.1689	0.1666	0.1645	0.1677	0.1624	0.1600	0.1635	0.1617	0.1656	0.1689	0.1382	0.1609
Fish and other fishing products; aquaculture products;	0.1557	0.1727	0.1920	0.1833	0.1873	0.1796	0.1795	0.1742	0.1618	0.1546	0.1384	0.1321	0.1340	0.1920	0.1321	0.1650
Food, beverages and tobacco products	0.2492	0.2636	0.2769	0.2842	0.2876	0.2892	0.2878	0.2893	0.2893	0.2911	0.2936	0.2893	0.2841	0.2936	0.2492	0.2827
Furniture and other manufactured goods	0.3741	0.4103	0.4236	0.4268	0.4383	0.4446	0.4592	0.4450	0.4318	0.4201	0.4068	0.4023	0.4092	0.4592	0.3741	0.4225
Machinery and equipment n.e.c.	0.1063	0.1215	0.1305	0.1288	0.1334	0.1350	0.1403	0.1425	0.1377	0.1354	0.1106	0.1056	0.1161	0.1425	0.1056	0.1264
Mining and quarrying	0.0622	0.0720	0.0711	0.0417	0.0463	0.0561	0.0575	0.0597	0.0489	0.0443	0.0501	0.0430	0.0547	0.0720	0.0417	0.0544
Motor vehicles, trailers and semi-trailers	0.1680	0.1878	0.1980	0.1990	0.2050	0.1841	0.1874	0.1898	0.1959	0.1974	0.1978	0.2258	0.2742	0.2742	0.1680	0.2008
Other non-metallic mineral products	0.1762	0.1912	0.2027	0.2012	0.2096	0.2052	0.2108	0.2276	0.2226	0.2186	0.2191	0.2274	0.2200	0.2276	0.1762	0.2102
Other transport equipment	0.0457	0.0518	0.0572	0.0530	0.0539	0.0530	0.0587	0.0603	0.0525	0.0494	0.0460	0.0443	0.0408	0.0603	0.0408	0.0513
Paper and paper products	0.2379	0.2451	0.2613	0.2580	0.2644	0.2778	0.2810	0.2863	0.2916	0.2897	0.2843	0.2946	0.3034	0.3034	0.2379	0.2750
Products of agriculture, hunting and related services	0.1258	0.1409	0.1518	0.1631	0.1608	0.1553	0.1488	0.1493	0.1440	0.1444	0.1454	0.1389	0.1446	0.1631	0.1258	0.1472
Products of forestry, logging and related services	0.1746	0.1739	0.1777	0.1858	0.1819	0.1772	0.1707	0.1689	0.1622	0.1573	0.1602	0.1653	0.1663	0.1858	0.1573	0.1709
Rubber and plastic products	0.0829	0.0920	0.0992	0.0957	0.1002	0.0954	0.1034	0.1042	0.1085	0.1084	0.1106	0.1389	0.1383	0.1389	0.0829	0.1060
Textiles, wearing apparel, leather and related products	0.3620	0.3939	0.4163	0.4330	0.4475	0.4610	0.4709	0.4720	0.4730	0.4622	0.4716	0.4613	0.4620	0.4730	0.3620	0.4451
Wood and of products of wood and cork	0.0358	0.0397	0.0501	0.0340	0.0438	0.0423	0.0440	0.0459	0.0436	0.0442	0.0408	0.0447	0.0479	0.0501	0.0340	0.0428
Total	0.1896	0.2082	0.2171	0.2127	0.2196	0.2211	0.2273	0.2281	0.2201	0.2141	0.2195	0.2159	0.2247	0.2281	0.1896	0.2168

Source: Own Calculations. Data from the Supply Table at current price under the National Account in the EUROSTAT database. Industries are categorized under the statistical classification of products by activity (CPA).

Table B: The Estimated Distribution Margin Across Countries in 2008

Industries \ Countries	Austria	Belgium	Czech Republic	Estonia	Finland	France	Germany	Greece	Hungary	Ireland	Italy	Lithuania	Netherlands	Portugal	Romania	Slovenia	Sweden	United Kingdom	Max	Min	Average
Basic metals	0.0446	0.0804	0.0427	0.0644	0.0103	0.0618	0.0492	0.1526	0.0884	0.0661	0.0501	0.0804	0.0575	0.0880	0.0681	0.0484	0.0774	0.1715	0.1715	0.0103	0.0723
Basic pharmaceutical products and pharmaceutical preparations	0.3179	0.2139	0.1336	0.2173	0.2976	0.3507	0.2776	0.2985	0.1841	0.0539	0.2711	0.3264	0.2174	0.3370	0.1867	0.1690	NA	0.2514	0.3507	0.0539	0.2414
Chemicals and chemical products	0.1451	0.1180	0.1121	0.1503	0.1103	0.1860	0.1037	0.2588	0.0999	0.0989	0.1208	0.1162	0.0906	0.1593	0.1436	0.1003	0.1445	0.2393	0.2588	0.0906	0.1388
Coke and refined petroleum products	0.1669	0.0518	0.1229	0.1336	0.0579	0.1021	0.0636	0.1221	0.0930	0.1981	0.1089	NA	0.0415	0.1443	0.2336	0.1217	0.0798	0.0319	0.2336	0.0319	0.1102
Computer, electronic and optical products	0.1852	0.1976	0.0553	0.0890	0.0800	0.2143	0.1630	0.3012	0.0346	0.0043	0.2088	0.2646	0.1886	0.1273	0.1473	0.1897	0.0909	0.2293	0.3012	0.0043	0.1539
Electrical equipment	0.1190	0.1986	0.0624	0.0650	0.1219	0.1757	0.1312	0.2835	0.0671	0.0321	0.1124	0.0932	0.1914	0.1296	0.0544	0.0846	0.1086	0.2378	0.2835	0.0321	0.1260
Fabricated metal products	0.1059	0.1042	0.0712	0.0938	0.0817	0.1269	0.0898	0.2467	0.0798	0.0090	0.0893	0.2211	0.1150	0.0965	0.0643	0.1016	0.0777	0.1617	0.2467	0.0090	0.1076
Fish and other fishing products; aquaculture products;	0.2962	0.2635	NA	0.0892	0.1837	0.4529	0.0679	0.2729	0.1936	0.2993	0.3985	0.3882	0.3628	0.5059	0.6100	0.3260	0.1452	0.1321	0.6100	0.0679	0.2934
Food, beverages and tobacco products	0.2084	0.1761	0.1886	0.1800	0.2730	0.2429	0.2099	0.2527	0.2025	0.2084	0.2202	0.2801	0.1601	0.2480	0.2145	0.2348	0.2120	0.2893	0.2893	0.1601	0.2223
Furniture and other manufactured goods	0.2674	0.1775	0.1184	0.2108	0.3203	0.3640	0.2410	0.2711	0.2157	NA	0.2566	0.2191	0.2596	0.2831	0.2771	0.1826	0.2116	0.4023	0.4023	0.1184	0.2517
Machinery and equipment n.e.c.	0.1033	0.1697	0.0740	0.0982	0.0418	0.1837	0.0609	0.1682	0.0855	0.3553	0.0852	0.1744	0.1542	0.1123	0.0844	0.0870	0.1199	0.1056	0.3553	0.0418	0.1258
Mining and quarrying	0.0694	0.0475	0.0431	0.0861	0.0423	0.0439	0.1630	0.1750	0.0538	0.0418	0.0490	0.0535	0.0204	0.0167	0.1028	0.1682	0.0475	0.0430	0.1750	0.0167	0.0704
Motor vehicles, trailers and semi-trailers	0.0918	0.0996	0.0758	0.1529	0.1349	0.1638	0.0779	0.4620	0.0675	0.2590	0.1367	0.2083	0.1139	0.0819	0.1322	0.0892	0.0779	0.2258	0.4620	0.0675	0.1473
Other non-metallic mineral products	0.1552	0.1884	0.0908	0.2049	0.2015	0.2595	0.1610	0.1692	0.1305	0.1990	0.1428	0.2373	0.2309	0.0869	0.1407	0.2373	0.1469	0.2274	0.2595	0.0869	0.1784
Other transport equipment	0.0325	0.1278	0.0156	0.1410	0.0438	0.0273	0.0492	0.1872	0.1051	0.0385	0.0885	0.2515	0.0887	0.0645	0.0384	0.1051	0.0636	0.0443	0.2515	0.0156	0.0841
Paper and paper products	0.1096	0.1674	0.0983	0.1125	0.0727	0.1409	0.1181	0.2740	0.1565	0.1811	0.0958	0.1181	0.1785	0.1356	0.0752	0.0703	0.1026	0.2946	0.2946	0.0703	0.1390
Printing and recording services	0.0008	0.1474	0.0980	NA	0.1951	0.1237	0.0032	0.1539	0.0244	0.1050	0.0013	0.2130	0.0117	0.0096	0.0134	0.0013	0.0093	NA	0.2130	0.0008	0.0694
Products of agriculture, hunting and related services	0.1648	0.2123	0.0974	0.1506	0.1726	0.2119	0.1830	0.2216	0.1365	0.1664	0.2753	0.1408	0.1862	0.1804	0.0274	0.2079	0.2456	0.1389	0.2753	0.0274	0.1733
Products of forestry, logging and related services	0.1263	0.1852	NA	0.1398	0.1555	0.1999	0.0275	0.1526	0.1625	0.0272	0.3781	0.2758	0.1323	0.1155	0.0308	0.1163	0.0031	0.1653	0.3781	0.0031	0.1408
Rubber and plastic products	0.1500	0.1227	0.0406	0.0978	0.1570	0.1520	0.0810	0.2830	0.0613	0.2463	0.0936	0.1362	0.1708	0.1441	0.1151	0.0666	0.0850	0.1389	0.2830	0.0406	0.1301
Textiles, wearing apparel, leather and related products	0.3093	0.2426	0.2879	0.2299	0.3738	0.3123	0.3529	0.2691	0.2140	0.3250	0.2188	0.2952	0.3436	0.2074	0.2254	0.2240	0.3628	0.4613	0.4613	0.2074	0.2920
Wood and of products of wood and cork	0.1134	0.1437	0.0555	0.0984	0.1337	0.1725	0.1481	0.2815	0.1491	0.1129	0.1887	0.1211	0.2357	0.0879	0.1755	0.0697	0.0678	0.0447	0.2815	0.0447	0.1333
Total	0.1464	0.1356	0.0943	0.1362	0.1266	0.1883	0.1306	0.2337	0.0991	0.1360	0.1500	0.1752	0.1406	0.1596	0.1379	0.1298	0.1212	0.2118	0.2337	0.0943	0.1474

Source: Own Calculations. Data from the Supply Table at current price under the National Account in the EUROSTAT database. Industries are categorized under the statistical classification of products by activity (CPA).

Table C: The Estimated Short Run Bilateral ERPT Against the Sterling (GBP) Across Countries

	CPA_A01	CPA_A03	CPA_B	CPA_C10-C12	CPA_C13-C15	CPA_C16	CPA_C17	CPA_C18	CPA_C19	CPA_C20	CPA_C21	CPA_C22	CPA_C23	CPA_C24	CPA_C25	CPA_C26	CPA_C27	CPA_C28	CPA_C30	CPA_C31_C32
Finland	**-0.7728**	**0.8180**	**-0.1001**	**-0.1001**	**0.3959*** **	**0.5373***	**0.1630**	**0.5805**	**1.3094**	**0.1536**	**1.6518**	**0.4838**	**0.5192***	**0.5065**	**0.3474***	**-0.2649**	**0.6666** **	**0.2577**	**-0.4906**	**-0.0991**
	(1.032)	(0.546)	(0.229)	(0.229)	(0.149)	(0.276)	(0.239)	(0.437)	(0.944)	(0.481)	(1.481)	(0.358)	(0.313)	(0.323)	(0.202)	(0.223)	(0.321)	(0.245)	(0.500)	(0.228)
France	**0.5809***	**0.0639**	**0.4898*** **	**0.4890*** **	**0.3559*** **	**0.4195** **	**-0.0413**	**0.4829**	**-0.0869**	**0.2043**	**0.7910** **	**0.4177**	**-0.1638**	**0.1038**	**0.5168*** **	**0.3302** **	**0.4713***	**0.4540***	**0.1477**	**0.4872*** **
	(0.335)	(0.308)	(0.116)	(0.116)	(0.109)	(0.197)	(0.172)	(0.324)	(0.467)	(0.243)	(0.396)	(0.271)	(0.248)	(0.208)	(0.136)	(0.150)	(0.257)	(0.256)	(0.332)	(0.116)
Germany	**-0.0091**	**0.7859***	**0.4034*** **	**0.3948*** **	**0.3183*** **	**0.1089**	**-0.0413**	**-0.2639**	**-0.3447**	**0.2344**	**1.0045**	**0.4271***	**0.2259**	**-0.1500**	**0.3222***	**0.1640**	**0.2501**	**0.3044**	**-0.0447**	**0.3976*** **
	(0.451)	(0.470)	(0.131)	(0.130)	(0.126)	(0.177)	(0.227)	(0.440)	(0.624)	(0.301)	(0.630)	(0.253)	(0.242)	(0.308)	(0.180)	(0.220)	(0.311)	(0.270)	(0.330)	(0.130)
Greece	**2.8867*** **	**2.8785*** **	**-0.0906**	**-0.0906**	**-0.0955**	**-1.0431**	**0.3884**	**-0.1658**	**0.0384**	**0.1803**	**0.1132**	**0.1662**	**0.4073***	**1.0107*** **	**0.7908*** **	**0.7105*** **	**0.4203**	**-1.0076** **	**0.6560**	**-0.0840**
	(0.798)	(0.824)	(0.154)	(0.154)	(0.154)	(0.598)	(0.334)	(0.485)	(0.846)	(0.534)	(1.403)	(0.193)	(0.209)	(0.422)	(0.253)	(0.297)	(0.344)	(0.402)	(0.516)	(0.152)
Ireland	**-0.4686**	**0.2241**	**0.4406*** **	**0.4376*** **	**0.4305*** **	**0.0781**	**0.3779**	**0.7562**	**0.6409*** **	**0.2435**	**0.8652**	**-0.1747**	**0.6054*** **	**0.0791**	**0.3328** **	**0.1038**	**0.6562** **	**-0.1862**	**-0.1244**	**0.4804*** **
	(0.715)	(0.467)	(0.140)	(0.141)	(0.142)	(0.150)	(0.323)	(0.506)	(0.273)	(0.289)	(0.572)	(0.221)	(0.241)	(0.273)	(0.148)	(0.175)	(0.322)	(0.388)	(0.306)	(0.148)
Italy	**0.1897**	**0.3008**	**0.2229***	**0.2229***	**0.2386** **	**0.1516**	**0.0939**	**1.0375*** **	**-0.0614**	**0.2533**	**1.4378*** **	**0.3130**	**0.2449**	**0.4824**	**0.2482**	**0.1352**	**0.1651**	**-0.0116**	**0.6664**	**0.2228***
	(0.845)	(0.522)	(0.119)	(0.119)	(0.115)	(0.179)	(0.182)	(0.409)	(0.590)	(0.423)	(0.574)	(0.242)	(0.213)	(0.324)	(0.169)	(0.154)	(0.282)	(0.208)	(0.415)	(0.117)
Netherlands	**0.5206**	**-0.2219**	**0.6520*** **	**0.6577*** **	**0.6454*** **	**0.1327**	**0.0510**	**0.5771**	**0.8888*** **	**-0.0331**	**0.9770**	**0.2254**	**-0.0944**	**-0.0447**	**0.4493*** **	**0.2270**	**0.7556**	**0.5516***	**0.4054**	**0.6528*** **
	(0.592)	(0.456)	(0.236)	(0.236)	(0.253)	(0.259)	(0.240)	(0.462)	(0.321)	(0.260)	(0.608)	(0.291)	(0.192)	(0.234)	(0.189)	(0.197)	(0.521)	(0.297)	(0.369)	(0.234)
Portugal	**-1.1093**	**0.9495**	**0.1265**	**0.1265**	**0.2931*** **	**0.5961**	**-0.0705**	**0.2421**	**1.3627** **	**-0.0080**	**0.4752**	**0.1461**	**-0.1353**	**-0.3764**	**0.3695***	**0.2359**	**0.2267**	**0.1652**	**0.5633**	**0.1244**
	(1.215)	(0.592)	(0.189)	(0.189)	(0.121)	(0.369)	(0.189)	(0.339)	(0.664)	(0.257)	(0.595)	(0.435)	(0.486)	(0.410)	(0.200)	(0.247)	(0.231)	(0.411)	(0.618)	(0.187)
Spain	**3.0165*** **	**1.2233*** **	**0.2421** **	**0.2421** **	**0.1906***	**0.0822**	**0.1419**	**0.3281**	**-0.1681**	**0.5366**	**0.9621**	**0.0863**	**0.3165**	**-0.0899**	**0.2785***	**0.0950**	**0.2532**	**0.0981**	**-0.0385**	**0.2307** **
	(0.985)	(0.489)	(0.111)	(0.111)	(0.109)	(0.358)	(0.217)	(0.404)	(0.598)	(0.416)	(0.772)	(0.308)	(0.242)	(0.255)	(0.150)	(0.186)	(0.225)	(0.208)	(0.677)	(0.107)
Sweden	**-0.0773**	**0.8600**	**0.2232***	**0.2232***	**0.2314** **	**0.1606**	**0.3126**	**-0.1609**	**0.0642**	**1.7350*** **	**0.2539**	**0.0437**	**0.2443**	**0.4763***	**0.2361**	**0.3142***	**0.3511** **	**-0.3805**	**0.4584**	**0.2213***
	(0.894)	(0.543)	(0.117)	(0.117)	(0.105)	(0.200)	(0.196)	(0.410)	(0.513)	(0.501)	(0.613)	(0.242)	(0.198)	(0.249)	(0.168)	(0.162)	(0.174)	(0.273)	(0.424)	(0.117)

Standard errors are shown in the parentheses. *,**,*** indicate the null hypothesis of the short run ERPT coefficient $\beta = 0$ is rejected at 10%, 5% and 1% significant level respectively. The estimation result for long run ERPT and the detailed diagnostic statistics for each regression are avaiable upon request.

Source: Own Estimations. Import and Export price data from COMEXT (Eurostat) database. The bilateral exchange rate uses the monthly average spot exchange rate reported by Bank of England. The sample period ranges from 1995:01 to 2001:12 for all these ten countries expect Sweden for which the full sample period (from 1995:01 to 2004:12) is used. Industries are categorized under the statistical classification of products by activity (CPA).

The relevant codes for the CPA classification

CPA_A01	Products of agriculture, hunting and related services
CPA_A02	Products of forestry, logging and related services
CPA_A03	Fish and other fishing products; aquaculture products; support services to fishing
CPA_B	Mining and quarrying
CPA_C10-C12	Food, beverages and tobacco products
CPA_C13-C15	Textiles, wearing apparel, leather and related products
CPA_C16	Wood and of products of wood and cork, except furniture; articles of straw and plaiting materials
CPA_C17	Paper and paper products
CPA_C18	Printing and recording services
CPA_C19	Coke and refined petroleum products
CPA_C20	Chemicals and chemical products
CPA_C21	Basic pharmaceutical products and pharmaceutical preparations
CPA_C22	Rubber and plastic products
CPA_C23	Other non-metallic mineral products
CPA_C24	Basic metals
CPA_C25	Fabricated metal products, except machinery and equipment
CPA_C26	Computer, electronic and optical products
CPA_C27	Electrical equipment
CPA_C28	Machinery and equipment n.e.c.
CPA_C29	Motor vehicles, trailers and semi-trailers
CPA_C30	Other transport equipment
CPA_C31_C32	Furniture and other manufactured goods

Table D: International Correlation of Consumptions

Countries	Australia	Austria	Belgium	Canada	Denmark	Euro Area	Finland	France	Germany	Hungary	Italy	Japan	Luxembou	Mexico	Netherlan	New Zeala	Norway	Poland	Portugal	Sweden	Switzerlar	UK	US
Australia	1.00	0.16	0.16	0.57	0.59	0.48	0.62	0.38	0.14	0.46	0.37	0.20	0.24	0.59	0.31	0.53	0.53	0.08	0.28	0.61	0.22	0.63	0.72
Austria	0.16	1.00	0.43	0.13	0.13	0.65	0.22	0.60	0.45	-0.04	0.69	0.06	0.25	0.31	0.27	-0.01	0.15	-0.06	0.44	0.48	0.27	0.00	0.47
Belgium	0.16	0.43	1.00	0.36	0.13	0.65	0.21	0.50	0.25	-0.30	0.70	-0.04	0.21	0.20	0.59	-0.22	0.19	0.30	0.78	0.43	0.34	0.00	0.12
Canada	0.57	0.13	0.36	1.00	0.53	0.57	0.57	0.32	0.16	0.27	0.50	0.23	0.38	0.66	0.37	0.27	0.62	0.23	0.43	0.66	0.19	0.65	0.51
Denmark	0.59	0.13	0.13	0.53	1.00	0.32	0.64	0.13	-0.03	0.14	0.29	0.37	0.24	0.65	0.07	0.60	0.68	0.19	0.16	0.46	0.06	0.55	0.57
Euro Area	0.48	0.65	0.65	0.57	0.32	1.00	0.47	0.85	0.65	0.17	0.78	0.04	0.45	0.54	0.69	0.00	0.43	0.13	0.73	0.73	0.48	0.36	0.61
Finland	0.62	0.22	0.21	0.57	0.64	0.47	1.00	0.27	0.13	0.36	0.46	0.37	0.17	0.60	0.17	0.56	0.57	0.25	0.25	0.47	-0.06	0.58	0.57
France	0.38	0.60	0.50	0.32	0.13	0.85	0.27	1.00	0.54	0.25	0.54	0.04	0.47	0.32	0.53	-0.03	0.29	-0.11	0.59	0.67	0.50	0.29	0.52
Germany	0.14	0.45	0.25	0.16	-0.03	0.65	0.13	0.54	1.00	0.12	0.31	0.06	0.25	0.32	0.38	-0.17	0.14	-0.05	0.20	0.46	0.37	0.17	0.43
Hungary	0.46	-0.04	-0.30	0.27	0.14	0.17	0.36	0.25	0.12	1.00	-0.02	0.10	0.07	0.29	-0.01	0.35	0.05	-0.34	-0.13	0.30	-0.04	0.46	0.42
Italy	0.37	0.69	0.70	0.50	0.29	0.78	0.46	0.54	0.31	-0.02	1.00	-0.08	0.24	0.39	0.48	-0.02	0.24	0.23	0.71	0.54	0.21	0.10	0.40
Japan	0.20	0.06	-0.04	0.23	0.37	0.04	0.37	0.04	0.06	0.10	-0.08	1.00	0.14	0.34	-0.20	0.52	0.43	-0.01	-0.10	0.27	-0.05	0.53	0.34
Luxembourg	0.24	0.25	0.21	0.38	0.24	0.45	0.17	0.47	0.25	0.07	0.24	0.14	1.00	0.41	0.35	-0.08	0.31	-0.02	0.26	0.38	0.30	0.33	0.39
Mexico	0.59	0.31	0.20	0.66	0.65	0.54	0.60	0.32	0.32	0.29	0.39	0.34	0.41	1.00	0.30	0.46	0.65	-0.03	0.28	0.63	0.31	0.72	0.82
Netherlands	0.31	0.27	0.59	0.37	0.07	0.69	0.17	0.53	0.38	-0.01	0.48	-0.20	0.35	0.30	1.00	-0.26	0.18	0.26	0.71	0.41	0.47	0.17	0.29
New Zealand	0.53	-0.01	-0.22	0.27	0.60	0.00	0.56	-0.03	-0.17	0.35	-0.02	0.52	-0.08	0.46	-0.26	1.00	0.56	-0.08	-0.15	0.31	-0.14	0.61	0.51
Norway	0.53	0.15	0.19	0.62	0.68	0.43	0.57	0.29	0.14	0.05	0.24	0.43	0.31	0.65	0.18	0.56	1.00	0.24	0.22	0.58	0.21	0.66	0.59
Poland	0.08	-0.06	0.30	0.23	0.19	0.13	0.25	-0.11	-0.05	-0.34	0.23	-0.01	-0.02	-0.03	0.26	-0.08	0.24	1.00	0.28	-0.09	-0.14	-0.07	-0.13
Portugal	0.28	0.44	0.78	0.43	0.16	0.73	0.25	0.59	0.20	-0.13	0.71	-0.10	0.26	0.28	0.71	-0.15	0.22	0.28	1.00	0.43	0.31	0.06	0.22
Sweden	0.61	0.48	0.43	0.66	0.46	0.73	0.47	0.67	0.46	0.30	0.54	0.27	0.38	0.63	0.41	0.31	0.58	-0.09	0.43	1.00	0.43	0.63	0.72
Switzerland	0.22	0.27	0.34	0.19	0.06	0.48	-0.06	0.50	0.37	-0.04	0.21	-0.05	0.30	0.31	0.47	-0.14	0.21	-0.14	0.31	0.43	1.00	0.22	0.42
UK	0.63	0.00	0.00	0.65	0.55	0.36	0.58	0.29	0.17	0.46	0.10	0.53	0.33	0.72	0.17	0.61	0.66	-0.07	0.06	0.63	0.22	1.00	0.71
US	0.72	0.47	0.12	0.51	0.57	0.61	0.57	0.52	0.43	0.42	0.40	0.34	0.39	0.82	0.29	0.51	0.59	-0.13	0.22	0.72	0.42	0.71	1.00
Avg.	**0.40**	**0.27**	**0.27**	**0.42**	**0.34**	**0.49**	**0.39**	**0.38**	**0.24**	**0.13**	**0.37**	**0.16**	**0.26**	**0.44**	**0.30**	**0.19**	**0.39**	**0.05**	**0.32**	**0.48**	**0.22**	**0.38**	**0.46**

Note: Statistics are calculated based on logged and HP-filtered quarterly seasonal adjusted private final consumption expenditure at constant prices from 1995:1-2013:2. Data source: OECD Main Economic Indicators.

Table E: International Correlation of GDP

Countries	Australia	Austria	Belgium	Canada	Denmark	Finland	France	Germany	Ireland	Italy	Japan	uxembour	Mexico	letherland	lew Zealan	Norway	Portugal	Spain	Sweden	Switzerlan	UK	US	ECD - Total
Australia	1.00	0.43	0.36	0.35	0.37	0.48	0.37	0.32	0.38	0.22	0.30	0.45	0.42	0.34	0.36	0.47	0.21	0.40	0.40	0.36	0.54	0.53	0.47
Austria	0.43	1.00	0.89	0.81	0.83	0.89	0.94	0.89	0.83	0.87	0.52	0.84	0.75	0.88	0.17	0.45	0.64	0.92	0.86	0.89	0.81	0.75	0.87
Belgium	0.36	0.89	1.00	0.83	0.79	0.86	0.89	0.85	0.76	0.87	0.60	0.80	0.77	0.85	0.21	0.43	0.66	0.82	0.83	0.84	0.77	0.75	0.89
Canada	0.35	0.81	0.83	1.00	0.76	0.79	0.87	0.75	0.74	0.74	0.49	0.74	0.75	0.70	0.25	0.35	0.48	0.76	0.84	0.73	0.74	0.82	0.85
Denmark	0.37	0.83	0.79	0.76	1.00	0.83	0.81	0.81	0.80	0.80	0.64	0.70	0.81	0.76	0.27	0.53	0.54	0.79	0.81	0.76	0.82	0.79	0.89
Finland	0.48	0.89	0.86	0.79	0.83	1.00	0.87	0.88	0.75	0.87	0.66	0.77	0.81	0.81	0.21	0.50	0.62	0.85	0.86	0.81	0.87	0.76	0.90
France	0.37	0.94	0.89	0.87	0.81	0.87	1.00	0.89	0.81	0.89	0.52	0.85	0.72	0.83	0.19	0.40	0.64	0.89	0.89	0.88	0.80	0.78	0.88
Germany	0.32	0.89	0.85	0.75	0.81	0.88	0.89	1.00	0.75	0.92	0.63	0.75	0.70	0.85	0.18	0.42	0.64	0.84	0.83	0.84	0.78	0.63	0.85
Ireland	0.38	0.83	0.76	0.74	0.80	0.75	0.81	0.75	1.00	0.73	0.51	0.74	0.69	0.72	0.30	0.53	0.42	0.80	0.72	0.75	0.76	0.72	0.81
Italy	0.22	0.87	0.87	0.74	0.80	0.87	0.89	0.92	0.73	1.00	0.65	0.76	0.70	0.79	0.14	0.42	0.65	0.82	0.86	0.84	0.78	0.62	0.84
Japan	0.30	0.52	0.60	0.49	0.64	0.66	0.52	0.63	0.51	0.65	1.00	0.48	0.63	0.40	0.49	0.49	0.25	0.44	0.64	0.46	0.78	0.65	0.79
Luxembourg	0.45	0.84	0.80	0.74	0.70	0.77	0.85	0.75	0.74	0.76	0.48	1.00	0.60	0.72	0.25	0.33	0.61	0.79	0.77	0.78	0.71	0.69	0.78
Mexico	0.42	0.75	0.77	0.75	0.81	0.81	0.72	0.70	0.69	0.70	0.63	0.60	1.00	0.68	0.22	0.61	0.44	0.65	0.75	0.68	0.78	0.83	0.87
Netherlands	0.34	0.88	0.85	0.70	0.76	0.81	0.83	0.85	0.72	0.79	0.40	0.72	0.68	1.00	0.02	0.41	0.78	0.89	0.70	0.88	0.65	0.58	0.74
New Zealand	0.36	0.17	0.21	0.25	0.27	0.21	0.19	0.18	0.30	0.14	0.49	0.25	0.22	0.02	1.00	0.41	-0.18	0.16	0.27	0.01	0.47	0.41	0.40
Norway	0.47	0.45	0.43	0.35	0.53	0.50	0.40	0.42	0.53	0.42	0.49	0.33	0.61	0.41	0.41	1.00	0.16	0.41	0.38	0.42	0.58	0.53	0.57
Portugal	0.21	0.64	0.66	0.48	0.54	0.62	0.64	0.64	0.42	0.65	0.25	0.61	0.44	0.78	-0.18	0.16	1.00	0.62	0.56	0.71	0.37	0.34	0.49
Spain	0.40	0.92	0.82	0.76	0.79	0.85	0.89	0.84	0.80	0.82	0.44	0.79	0.65	0.89	0.16	0.41	0.62	1.00	0.76	0.87	0.78	0.65	0.79
Sweden	0.40	0.86	0.83	0.84	0.81	0.86	0.89	0.83	0.72	0.86	0.64	0.77	0.75	0.70	0.27	0.38	0.56	0.76	1.00	0.74	0.84	0.81	0.90
Switzerland	0.36	0.89	0.84	0.73	0.76	0.81	0.88	0.84	0.75	0.84	0.46	0.78	0.68	0.88	0.01	0.42	0.71	0.87	0.74	1.00	0.71	0.64	0.78
UK	0.54	0.81	0.77	0.74	0.82	0.87	0.80	0.78	0.76	0.78	0.78	0.71	0.78	0.65	0.47	0.58	0.37	0.78	0.84	0.71	1.00	0.84	0.93
US	0.53	0.75	0.75	0.82	0.79	0.76	0.78	0.63	0.72	0.62	0.65	0.69	0.83	0.58	0.41	0.53	0.34	0.65	0.81	0.64	0.84	1.00	0.92
OECD - Total	0.47	0.87	0.89	0.85	0.89	0.90	0.88	0.85	0.81	0.84	0.79	0.78	0.87	0.74	0.40	0.57	0.49	0.79	0.90	0.78	0.93	0.92	1.00
Avg.	**0.39**	**0.76**	**0.74**	**0.69**	**0.72**	**0.76**	**0.75**	**0.73**	**0.68**	**0.72**	**0.55**	**0.68**	**0.68**	**0.68**	**0.24**	**0.45**	**0.48**	**0.71**	**0.73**	**0.70**	**0.73**	**0.68**	**0.78**

Note: Statistics are calculated based on logged and HP-filtered quarterly seasonal adjusted gross domestic product at constant prices from 1963:3-2013:2. Data source: OECD Main Economic Indicators.

Table F: Cross Correlation Between GDP and Consumption

Countries	(-4)	(-3)	(-2)	(-1)	0	(+1)	(+2)	(+3)	(+4)
Australia	0.05	0.26	0.47	0.60	0.68	0.61	0.43	0.23	0.11
Austria	-0.27	-0.12	0.05	0.23	0.40	0.53	0.61	0.63	0.62
Belgium	-0.15	-0.05	0.14	0.36	0.50	0.46	0.30	0.12	0.00
Canada	0.02	0.15	0.35	0.52	0.62	0.61	0.45	0.17	-0.15
Denmark	0.11	0.25	0.36	0.48	0.68	0.67	0.58	0.51	0.30
Finland	-0.03	0.17	0.40	0.61	0.75	0.74	0.66	0.42	0.17
France	0.17	0.29	0.45	0.58	0.66	0.61	0.53	0.40	0.23
Germany	0.12	0.30	0.44	0.46	0.52	0.46	0.41	0.34	0.25
Italy	-0.36	-0.12	0.16	0.43	0.61	0.66	0.63	0.50	0.32
Japan	-0.20	-0.08	0.10	0.42	0.77	0.70	0.53	0.42	0.22
Luxembourg	0.24	0.35	0.46	0.51	0.54	0.36	0.24	0.16	0.03
Mexico	0.29	0.46	0.66	0.82	0.91	0.74	0.53	0.31	0.13
Netherlands	0.41	0.53	0.65	0.72	0.71	0.59	0.42	0.24	0.06
New Zealand	0.19	0.33	0.50	0.58	0.66	0.60	0.42	0.28	0.11
Norway	0.08	0.23	0.39	0.55	0.65	0.53	0.49	0.34	0.25
Portugal	0.04	0.25	0.49	0.70	0.83	0.71	0.60	0.46	0.27
Sweden	0.04	0.27	0.48	0.62	0.79	0.77	0.61	0.41	0.10
Switzerland	0.26	0.31	0.40	0.49	0.59	0.59	0.50	0.39	0.24
UK	0.12	0.34	0.59	0.78	0.86	0.82	0.66	0.41	0.13
US	0.46	0.63	0.78	0.88	0.91	0.85	0.70	0.51	0.31
Avg.	**0.08**	**0.24**	**0.42**	**0.57**	**0.68**	**0.63**	**0.51**	**0.36**	**0.18**

Note: Statistics are calculated based on logged and HP-filtered quarterly seasonal adjusted private final consumption expenditure at constant prices and gross domestic products at constant prices from periods 1995:1-2013:2. Data source: OECD Main Economic Indicators.

Table G: Autocorrelations of RER

Countries	(-1)	(-2)	(-3)	(-4)
Australia	0.79	0.55	0.37	0.20
Belgium	0.84	0.59	0.36	0.18
Canada	0.83	0.58	0.39	0.22
Denmark	0.79	0.50	0.25	0.03
Finland	0.88	0.69	0.51	0.31
France	0.80	0.55	0.36	0.19
Germany	0.83	0.58	0.32	0.07
Greece	0.68	0.46	0.32	0.23
Iceland	0.79	0.61	0.37	0.11
Ireland	0.79	0.48	0.24	0.06
Italy	0.84	0.60	0.38	0.17
Japan	0.83	0.59	0.45	0.28
Luxembourg	0.80	0.54	0.32	0.13
Mexico	0.80	0.56	0.33	0.05
Netherlands	0.83	0.58	0.32	0.08
New Zealan	0.84	0.64	0.45	0.29
Norway	0.71	0.35	0.05	-0.18
Portugal	0.78	0.51	0.28	0.04
Spain	0.83	0.60	0.38	0.18
Sweden	0.81	0.55	0.33	0.13
Switzerland	0.80	0.54	0.31	0.07
UK	0.80	0.52	0.32	0.16
US	0.81	0.55	0.36	0.19
Avg.	**0.80**	**0.55**	**0.34**	**0.14**

Note: Statistics are calculated based on logged and HP-filtered quarterly seasonal adjusted real effective exchange rate from 1980:1-2013:2. Data source: International Financial Statistics

Table H: Autocorrelations of RGDP

Countries	(-1)	(-2)	(-3)	(-4)
Australia	0.69	0.49	0.29	0.07
Austria	0.75	0.56	0.37	0.16
Belgium	0.84	0.56	0.25	0.01
Canada	0.85	0.63	0.42	0.21
Denmark	0.75	0.55	0.36	0.14
Finland	0.82	0.69	0.53	0.33
France	0.57	0.50	0.38	0.17
Germany	0.81	0.62	0.42	0.23
Ireland	0.83	0.70	0.49	0.29
Italy	0.82	0.56	0.27	0.00
Japan	0.79	0.57	0.35	0.11
Luxembourg	0.83	0.64	0.38	0.16
Mexico	0.87	0.67	0.43	0.24
Netherlands	0.62	0.48	0.34	0.16
New Zealan	0.75	0.54	0.42	0.21
Norway	0.67	0.52	0.35	0.23
Portugal	0.86	0.70	0.50	0.28
Spain	0.81	0.67	0.48	0.25
Sweden	0.71	0.54	0.35	0.20
Switzerland	0.84	0.66	0.50	0.28
UK	0.82	0.66	0.47	0.25
US	0.88	0.70	0.49	0.28
OECD - Tota	0.89	0.70	0.46	0.22
Avg.	**0.78**	**0.60**	**0.40**	**0.19**

Note: Statistics are calculated based on logged and HP-filtered quarterly seasonal adjusted GDP at constant prices from 1980:1-2013:2. Data source: OECD Main Economic Indicators.

Chapter 2

Decomposing Exchange Rate Pass Through – Evidence from Chinese Exporters

We develop a new approach to estimating demand-side exchange rate pass through, also known as pricing-to-market, for multi-product multi-destination exporting firms. We then examine the pass through of nominal exchange rate fluctuations and local CPI fluctuations into export prices of Chinese firms' products. A sequential fixed effects estimator is employed to examine the extent to which Chinese firms adjust their markups in different foreign markets in response to bilateral and global exchange rate movements. We assess how markups vary over time by firm structure and by product features. Finally, we develop a simple model to disentangle the importance of bilateral (renminbi to local currency) versus multilateral exchange rate fluctuations to firms' pricing behavior.

2.1 Introduction

In this paper[1], we use highly-disaggregated data on the universe of Chinese exports to examine exchange rate pass through (ERPT), the export price responsiveness to changes in the bilateral, nominal value of the renminbi (rmb) vis-a-vis the currencies of 161 importing countries over 2000-2011. We develop a new approach to identifying and estimating a specific component of ERPT, the price markup elasticity of exported goods (in exporter's currency) with respect to bilateral exchange rate movements. Our analysis makes three empirical contributions. First, using this new estimator on Chinese data, we find a markup elasticity for Chinese exporters of 5.6-6.6% over 2000-2011. This implies that, after accounting for marginal cost changes, 93-95% of bilateral exchange rate movements are passed through to import prices (in importer's currency). Second, we identify considerable heterogeneity across Chinese firms in the extent to which they engage in pricing to market across destinations, with some types of firms adjusting markups by more than 10%. Finally, we confirm that the Chinese export price to a destination (importing) country responds not only to the bilateral (rmb to importer currency) exchange rate movements but also to multilateral currency movements that are orthogonal to movements in the bilateral (rmb to importer) exchange rate. More importantly, although both bilateral and orthogonal multilateral movements contribute to export price changes, only the bilateral exchange rate impacts on the destination-specific markup elasticity.

Our analysis begins with the observation that the price of an imported good can be decomposed into three parts: (a) the good's production costs in the origin (exporting) country including any parts or components that were imported into that origin country, (b) a markup, i.e., an increase in the price charged over and above production costs that reflects how much market power an exporting firm has in a destination (importing) country, and (c) the costs associated with marketing and distributing a good to consumers in the destination (importing) country. At least since Corsetti and Dedola (2005) and Corsetti, Dedola and Leduc (2008), it has been well understood that movements in the exchange rate would impact on each of these components differently, depending not only on structural factors, such as the intensity of market competition in the destination (importing) country, but also on the the source of the shocks that caused the movement in the exchange rate. Empirical models that fail to recognise the need to distinguish the differential effect of exchange rates on each component of prices and ignore the underlying source of the shocks to the exchange rate yield estimates of ERPT that are of limited usefulness in guiding policymakers.

The approach we take focuses on micro, firm-level data. Among the leading firm-level studies to date are Berman, Martin and Mayer (2012) and Amiti, Itskhoki, and Konings (2014), both of which find evidence consistent with the theoretical predictions of Corsetti

[1] This chapter is sourced from my working paper with Dr. Meredith Crowley at University of Cambridge and Associate Professor Huasheng Song at Zhejiang University.

and Dedola (2005).[2] Berman, Martin and Mayer (2012) find that larger, more productive French firms adjust their markups more than smaller, less productive firms in response to exchange rate changes (pass through is lower); Amiti, Itskhoki, and Konings (2014) use data on Belgian firms to show that firms with higher shares of imported components tend to adjust their export prices more than firms with less reliance on imported parts. While both papers break important new ground, both need to rely on imperfect controls for changes in production costs of individual producers, without which it is not possible to obtain accurate estimates of the mark-up elasticity. In this paper, we develop and implement a methodology to improve the empirical accuracy of this essential parameter.

The novel methodology consists of a sequential fixed effects estimator applied to a large, high-dimensional unbalanced panel dataset, covering the universe of firm and product level trade transactions for a country. Motivated by the observation that 71% of Chinese exporters export their products to more than one destination, we develop and employ a Gagnon and Knetter (1995) type identification strategy which exploits export price information from multi-destination firms. Because the same product of an exporter which is exported to multiple destinations should have the same marginal cost, we could exploit price differentials across destinations in response to an exchange rate shock to estimate the markup elasticity. However, we show that simply applying a conventional method of adding destination fixed effects to highly disaggregated firm-level data would produce biases in estimates of the markup elasticity due to endogenous changes in a firm's product-level trade pattern. We demonstrate that when the purpose of the study is to estimate markup elasticities by controlling for the effect of unobserved marginal cost, the sequence in which fixed effects are introduced matters. We construct a sequential fixed effects estimator at the firm-product level to eliminate the time-varying marginal cost component from prices in order to recover the destination specific markup and estimate its responsiveness to exchange rate movements.[3]

The estimator is designed to deliver a more accurate estimate of the responsiveness of the markup component arising from country-specific changes in import demand. We argue that this estimator is better at controlling for time-varying marginal cost at the firm-product level than alternative controls like time-varying firm-level TFP. Moreover, because our methodology has a low data requirement (i.e., it does not require the detailed firm-level data used in the computation of TFP), there is a distinct practical advantage to our estimator.[4] In practice, the estimator's low data requirement facilitates the estimation of the markup elasticity for groups of firms for which national manufacturing censuses do not collect data. In providing guidance to policy, we argue that this methodology is superior to other, more

[2] An alternative approach to this question has aimed to capture the differential impact of exchange rate shocks on components of prices with macro time series models (Forbes, et. al., 2015) and disaggregated analysis by sectors that attempt to reconcile sectoral and aggregate estimates (Lewis, 2016).

[3] Our estimator is robust to Gopinath, Itskhoki, and Rigobon (2010) style of S-period differences and enables us to estimate the markup elasticity conditional on price changes.

[4] For example, our estimator can be applied to customs transactions data which is available for many countries.

conventional approaches, which average out the effect of different shocks, and/or may rely on information which is aggregated and imprecise as regards production costs.

Economic theory suggests that there are cost-side and demand-side factors which contribute to incomplete ERPT. On the cost side, an appreciation of rmb against other currencies would be expected to reduce Chinese export prices because Chinese firms that use imported inputs in production would experience a decrease in their marginal production costs. On the demand side, profit-maximizing firms which face less elastic demand and command more market power in destination (importing) countries would be expected to reduce their markups in the face of an rmb appreciation. Together, these two effects could generate reductions in export prices (in exporter's currency) and incomplete ERPT.

First, using a conventional estimator of ERPT (Gopinath and Rigobon (2008)), we find the ERPT of China's exporters is high. Conditional on a price change, the rmb price of Chinese exports responds to nominal bilateral exchange rate shocks by 20%; this means that 80% is passed through to the price denominated in the local destination currency. Although our estimated ERPT is higher than those estimates obtained from sector or industry-level price indices[5], our finding is consistent with recent studies on exporters using firm-level data (e.g. Amiti, Itskhoki, and Konings (2014)).

Second, with our estimator, we isolate the markup adjustments of Chinese firms and find that, conditional on price changes, markup adjustments account for one-third to one-half of the total ERPT for China's exporters over 2000-2011. There is substantial markup responsiveness to exchange rate movements by all Chinese exporters of 5.6-6.6% over 2000-2011. This finding is important because a traditional estimator of ERPT (e.g. Gopinath, Itskhoki, and Rigobon (2010)) finds evidence of substantially incomplete pass through; that is, firms adjust export prices to dampen the transmission of exchange rate shocks to consumers. However, this traditional estimator cannot inform us as to whether export price adjustment occurs via the passing through of a firm's gain/loss on the cost side versus an adjustment of its markup.[6]

Third, we document heterogeneity in the markup responsiveness to bilateral exchange rate changes. For example, over 2005-2011, 25% of Chinese export value was exported by "trading firms."[7] We estimate the markup elasticity of these firms as ranging from 7.4-13.5% (2000-2011), roughly 2-4 times higher than other Chinese exporters. Further, we estimate higher markup responsiveness for firms exporting to a set of seven high-income countries

[5] Knetter (1989), Knetter (1993), Goldberg and Campa (2006).

[6] Fitzgerald and Haller (2014) take a different methodological approach to estimating markup adjustments and find that Irish firms adjust prices to maintain markups in local currency.

[7] See the appendix for a definition of trading firms. Chinese trading firms, which are somewhat unique in terms of their history and structure, might directly manufacture goods, but they also serve as trade intermediaries. Our approach enables estimation of the markup elasticity for firms of this type for which the construction of timing varying production costs is not feasible.

that use freely floating major currencies other than the US dollar of 8.0% (all firms, 2000-2011) and 13.5% (trading firms, 2000-2011).

Fourth, we evaluate the effect of the switch of the exchange rate regime in China in 2005 from a peg to a managed peg/float and find evidence for a unique role of the dollar as an international currency [Gopinath (2015)]. We first estimate how export prices (inclusive of production costs) change in response to bilateral exchange rate movements and the orthogonal component of the importing country's multilateral exchange rate against currencies other than the rmb. This approach verifies that, after controlling for the bilateral rmb-importer currency, the orthogonal component of the importer's multilateral exchange rate has a significant effect on exporters' prices. In other words, as suggested by Dornbusch (1987) and Atkeson and Burstein (2008) and empirically verified by Auer and Schoenle (2016), competitive pressure from other exporters in a destination influences Chinese prices. More importantly, with our estimator, we are able to distinguish the contribution of the orthogonal multilateral exchange rate movement to the Chinese export price (inclusive of costs and global demand factors) versus the destination markup. Our results suggest that the competition effect on the destination markup arising from idiosyncratic exchange rate shocks between the trading partner and its trading partners is small.

Our final contribution is to integrate a canonical partial equilibrium model of firm-level export prices with shocks from macro data series to study the sensitivity of estimated ERPT to the explanatory variables that our reduced form estimates suggest are important. Specifically, we study three possible causes of incomplete ERPT apart from nominal rigidities. The first category explains incomplete ERPT by considering oligopolistic markets in which the optimal adjustment of the markup depends on the curvature of demand and the market structure.[8] The second group emphasizes the importance of the local component of the consumer price, e.g., the importance of local distribution costs[9]. The third group is represented by the recent work of Amiti, Itskhoki, and Konings (2014) which focuses on the impact of the exchange rate on marginal cost through imported inputs. Our model confirms the importance of the firm's import share, the pass-through of exchange rates into the prices of imported inputs and the role of local distribution cost in ERPT estimation. We also reaffirm our estimator is robust to various specifications of firm-specific cost factors. Finally, our simulation results suggest that firm-specific factors account for around half of the general equilibrium effect of price adjustments due to exchange rate movements.

The rest of the paper is organized as follows. Section 2.2 presents a brief review of recent contributions. Section 2.3 presents the general pricing equation for a firm exporting to multiple destinations. Section 2.4 discusses our empirical identification strategy. Section 2.5 summarizes the database and presents statistics on Chinese exporters. Section 2.6 and 2.7 present our empirical results on bilateral and multilateral ERPT respectively. Section 2.8

[8] Dornbusch (1987), Atkeson and Burstein (2007), Atkeson and Burstein (2008).

[9] Burstein, Neves, and Rebelo (2003), Corsetti and Dedola (2005)

builds a numerical model to understand how changing key parameters of the firm's pricing equation affects the estimates of ERPT. Section 2.9 concludes.

2.2 Literature Review

A consistent finding of the exchange rate pass through (ERPT) literature is that export prices denominated in local currency do not react one-to-one to bilateral exchange rate movements. The literature suggests that the lack of pass through is mainly driven by three channels: (a) nominal rigidities, where the export price is rigid in local currency; (b) a marginal cost channel, where marginal cost of a product could be correlated with exchange rate movements through direct channels such as imported inputs or indirect channels through change of the production scale as in Marston (1990); (c) a pricing to market channel, where an exporter optimally chooses to stabilize its export price denominated in local currency due to an oligopolistic market structure and the existence of a local distribution cost.

As discussed in Goldberg and Knetter (1997) and Corsetti, Dedola, and Leduc (2008a), controlling for the marginal cost is of particular importance for understanding how exporters pass through exchange rate shocks to their prices. At the aggregate level, the literature has used various proxies to control for the marginal cost. The most commonly used two are the nominal wage index[10] and the PPI of the exporting country[11]. Although industrial cost measures might be good proxies for the average cost, they fail to capture the true marginal cost.

A number of recent studies of ERPT[12] have used firm-level data and emphasized the role of markup adjustments as an impediment to the transmission of exchange rate shocks into international prices. Amiti, Itskhoki, and Konings (2014), for example, have improved the approximation accuracy of the marginal cost by taking into account the estimated productivity of the exporter and the cost of its total imported inputs. Their method gives the most precise proxy of the firm's marginal cost so far. However, the productivity estimation often involves complicated matching process across datasets. Due to the frequency limitations of annual industrial survey, only annual productivity estimation can be used. Another limitation is that an exporter may simultaneously export many products. It may be problematic to just assume all products share the same marginal cost, while dropping multi-product firms

[10] e.g. Goldberg and Campa (2005)

[11] e.g. Marazzi, Sheets, Vigfusson, Faust, Gagnon, Marquez, Martin, Reeve, and Rogers (2005)

[12] Berman, Martin, and Mayer (2012) find that French exporters raise export prices in response to a depreciation of their currency against a destination currency. Moreover, this markup adjustment is larger for more productive firms, implying that PTM by the most productive and largest exporters contributes to low pass-through into import prices. Amiti, Itskhoki, and Konings (2014) examine the role of markup adjustment and imported inputs in pass through for Belgian firms. They find that firms with greater shares of imported inputs and larger export sales have lower pass through.

involves a significant drop in data points and loss of information[13]. In addition, it is often not clear which imported input is used to produce the exported product.

The most relevant paper in terms of the methodology is Fitzgerald and Haller (2014) which studies the markup adjustments of Irish exporters and uses their domestic sale price as the reference to control for unobserved cost shifts. They find strong evidence of pricing to market by Irish firms (i.e., stable prices in consumer currency implying full pass through of markup adjustments to producer currency conditional on price changes). Interestingly, this is qualitatively different from what we observe for Chinese firms. While Fitzgerald and Haller (2014) find that conditional on a price change, Irish firms adjust markups in exporter currency by almost 100%, we find that Chinese firms adjust their markups in exporter currency by 5-10%. This implies that although the price of the good in consumer's currency is rising with an rmb appreciation, this price increase is dampened by only a small reduction in markup charged in the currency of the consumer. One possible explanation for these observed differences between Irish and Chinese firms might be due to the firm's choice of an invoicing currency. The Irish firms studied by Fitzgerald and Haller (2014) invoiced their export transactions in the destination market currency - sterling.

2.3 A General Pricing Equation

We consider the problem of a firm f located in China that can sell its output i domestically or in N foreign countries which are denoted $d_1, d_2, ..., d_N$. For destination market d_1, the Chinese firm's problem is to set a price in rmb given the bilateral exchange rate rate between the rmb and the destination country d_1's currency, the economic environment in China, including macroeconomic conditions and firm and product-specific variables X_{ift}, the economic environment in destination d_1, X_{ifd_1t}, including preferences for firm f's product i, the bilateral exchange rates of all other countries whose firms compete with Chinese firms in market d_1, $e_{\frac{d_2}{d_1},t}, ..., e_{\frac{d_N}{d_1},t}$, and features of the economic environments in countries d_2 through d_N that impact on the costs of firms that compete against Chinese firms in destination d_1, $X_{d_2t}, ..., X_{d_Nt}$.

$$P^{rmb}_{ifd_1t} = f\left(e_{\frac{CHN}{d_1},t}, e_{\frac{d_2}{d_1},t}, ..., e_{\frac{d_N}{d_1},t}, X_{ift}, X_{ifd_1t}, X_{d_2t},, ..., X_{d_N,t}, \varepsilon_{fhd_1t}\right) \tag{2.1}$$

$$P^{d_1}_{ifd_1t} = \frac{P^{rmb}_{ifd_1t}}{e_{\frac{CHN}{d_1},t}} \tag{2.2}$$

where $P^{rmb}_{ifd_1t}$ stands for the rmb price of the product i sold by firm f to destination d_1 at time t and $P^{d_1}_{ifd_1t}$ is the price of that same good in country d_1's currency; CHN stands for China.

[13] Amiti, Itskhoki, and Konings (2014) report 95% of Belgian firms are multi-product exporters.

Without loss of generality, we can separate the effect of changing exchange rates and economic fundamentals into supply and demand components associated with China and destination d_1 ($\mathcal{S}_{ifd_1t}$ and $\mathcal{D}_{ifd_1t}$) and with destination d_1 and its trading partners, $d_2,...,d_N$, ($\mathcal{S}_{-d_1t}$ and $\mathcal{D}_{-d_1t}$) respectively.

$$\mathcal{S}_{ifd_1t} \equiv g_{\mathcal{S},d_1}(e_{\frac{CHN}{d_1},t}, X_{ift}) \tag{2.3}$$

$$\mathcal{D}_{ifd_1t} \equiv g_{\mathcal{D},d_1}(e_{\frac{CHN}{d_1},t}, X_{ifd_1t}) \tag{2.4}$$

$$\mathcal{S}_{-d_1t} \equiv g_{\mathcal{S},-d_1}(e_{\frac{d_2}{d_1},t},...,e_{\frac{d_N}{d_1},t}, X_{d_2,t},...,X_{d_N,t}) \tag{2.5}$$

$$\mathcal{D}_{-d_1t} \equiv g_{\mathcal{D},-d_1}(e_{\frac{d_2}{d_1},t},...,e_{\frac{d_N}{d_1},t}, X_{d_2,t},...,X_{d_N,t}) \tag{2.6}$$

Therefore, the pricing equation can be simplified to

$$P^{rmb}_{ifd_1t} = f_{transformed}\left(\mathcal{S}_{ifd_1t}, \mathcal{D}_{ifd_1t}, \mathcal{S}_{-d_1t}, \mathcal{D}_{-d_1t}, \varepsilon_{ifd_1t}\right) \tag{2.7}$$

We assume that the pricing equation (3.25) is log linear in its components[14] and rewrite it as:

$$\begin{aligned} log(P^{rmb}_{ifd_1t}) = \beta_{0,d_1} + \beta_{1,d_1} log(\mathcal{S}_{ifd_1t}) + \beta_{2,d_1} log(\mathcal{D}_{ifd_1t}) \\ + \beta_{3,d_1} log(\mathcal{S}_{-d_1t}) + \beta_{4,d_1} log(\mathcal{D}_{-d_1t}) + log(\varepsilon_{ifd_1t}) \end{aligned} \tag{2.8}$$

Traditional approaches to estimating ERPT regress the exporter's price in exporter's currency on the bilateral exchange rate between the origin and destination. From equation (2.8), it is clear that any bilateral exchange rate movements (rmb/d_1) can operate either through destination-specific demand-side factors ($\mathcal{D}_{ifd_1t}$) or origin and firm-specific supply side factors ($\mathcal{S}_{ifd_1t}$). Our primary objective is to develop an estimator that can identify and decompose the impact of bilateral exchange rate movements on price changes into components operating through idiosyncratic demand factors in each destination and firm-product specific cost. A secondary objective of our analysis is to examine what role, if any, currency movements of other countries play in the firm's pricing decision. Third country currency movements could affect the costs of a Chinese firm via an imported input channel or they could affect the intensity of competition the firm faces in a destination country.[15]

We develop an estimator that can identify the responsiveness of the export price markup and the demand-side component of prices to bilateral and multilateral exchange rate movements. To simplify the exposition of our problem, we begin by presenting the identification problem under the assumption that supply and demand factors associated with countries $d_2,...,d_N$ do not influence the pricing decision of the firm in country d_1. We then present

[14] As shown in section 2.4, our empirical strategy also works under the existence of possible interactive effects.

[15] Dornbusch (1987), Atkeson and Burstein (2007), Atkeson and Burstein (2008).

results associated with the bilateral estimator. We will relax this assumption in section 2.7 and report the more general results on the impact of bilateral and multilateral exchange rate movements on prices.

2.4 Empirical Strategy

To illustrate how identification works with our estimator, we present (2.8) as a parsimonious pricing equation and discuss existing approaches for estimating ERPT, the biases that can arise under these approaches, and the advantages of our estimator. We write the price of a Chinese exporting firm f selling a good i in destination d in year t as a function of these four factors - firm, product, destination and year - and interactions between these factors.[16]

$$\begin{aligned} p_{fhdt} \quad = \quad & \mathcal{F}_i + \mathcal{F}_f + \mathcal{F}_d + \mathcal{F}_t \\ & + \mathcal{F}_{if} + \mathcal{F}_{id} + \mathcal{F}_{it} + \mathcal{F}_{fd} + \mathcal{F}_{ft} + \mathcal{F}_{dt} \\ & + \mathcal{F}_{fdt} + \mathcal{F}_{idt} + \mathcal{F}_{ift} + \mathcal{F}_{ifd} \\ & + \mathcal{F}_{ifdt} + \varepsilon_{ifdt} \end{aligned} \tag{2.9}$$

where the coefficients in front of factors are omitted for clarity. Some of the factors governing pricing, for example, $\mathcal{F}_d$ and $\mathcal{F}_{id}$ are best understood as demand-side factors, like destination specific tastes for all goods and for good i, respectively. Other factors, $\mathcal{F}_f$ and $\mathcal{F}_{ft}$ are firm-level supply factors. The term $\mathcal{F}_{ifd}$ captures a time-invariant match of supply and demand, i.e., destination d's idiosyncratic preference for the product i manufactured by firm f. Time-varying factors common to all Chinese firms (Chinese GDP growth, Chinese inflation) are captured by $\mathcal{F}_t$.

The most challenging component for econometricians interested in ERPT is the time-varying unobserved marginal cost of the product i produced within the firm, $\mathcal{F}_{ift}$. Importantly, this factor is not only unobserved by the econometrician, but it is also potentially unobservable to the agents in a multi-product firm because the allocation of some firm-level costs across products is not conceptually well-defined. The key object of interest, the bilateral exchange rate between China and country d is captured by the factor $\mathcal{F}_{dt}$ which also includes macro variables for country d like CPI and GDP growth.

We emphasize two additional time-varying factors that reflect important heterogeneity in ERPT across firms. The term $\mathcal{F}_{fdt}$ embodies any variation across firms in the extent of ERPT associated with variables like a firm's market share in a destination which, in turn, reflects the firm's market power in a destination. Similarly, $\mathcal{F}_{ifdt}$ captures any interactions between the bilateral exchange rate between China and destination d that impact on the firm's marginal cost in producing good i. From Berman, Martin, and Mayer (2012), who

[16] We will return to a richer specification that includes multilateral competition later.

find that larger, more productive French firms have lower pass-through than smaller, less productive firms, and Amiti, Itskhoki, and Konings (2014), who find similar differences in ERPT between large and small firms in Belgium, we know that these interaction terms are quantitatively significant in ERPT.

From (2.9), it is straightforward to see a high-dimensional (product-firm-destination) fixed effect achieves a similar result as an S-period difference in terms of eliminating unobserved confounding factors. After including product-firm-destination (ifd) fixed effects, the equation simplifies to:

$$\begin{aligned}
\widetilde{p}_{ifdt} = \; & \mathcal{F}_f + \mathcal{F}_i + \mathcal{F}_d + \widetilde{\mathcal{F}}_t \\
& + \mathcal{F}_{if} + \mathcal{F}_{id} + \mathcal{F}_{fd} + \widetilde{\mathcal{F}}_{it} + \widetilde{\mathcal{F}}_{ft} + \widetilde{\mathcal{F}}_{dt} \\
& + \widetilde{\mathcal{F}}_{fdt} + \widetilde{\mathcal{F}}_{idt} + \widetilde{\mathcal{F}}_{ift} + \mathcal{F}_{ifd} \\
& + \widetilde{\mathcal{F}}_{ifdt} + \widetilde{\varepsilon}_{ifdt}
\end{aligned} \tag{2.10}$$

in which the tilde sign above a variable indicates that it is the deviation in the underlying variable from the ifd fixed effect, i.e., $\widetilde{x_j} \equiv x_j - \sum_j x_j / n_J \; \forall j \in \{t, it, ft, dt, fdt, idt, ift, ifdt\}$

Alternatively, taking an S-period difference over time yields:

$$\begin{aligned}
\Delta_s p_{fhdt} = \; & \mathcal{F}_i + \mathcal{F}_f + \mathcal{F}_d + \Delta_s \mathcal{F}_t \\
& + \mathcal{F}_{if} + \mathcal{F}_{id} + \mathcal{F}_{fd} + \Delta_s \mathcal{F}_{it} + \Delta_s \mathcal{F}_{ft} + \Delta_s \mathcal{F}_{dt} \\
& + \Delta_s \mathcal{F}_{fdt} + \Delta_s \mathcal{F}_{idt} + \Delta_s \mathcal{F}_{ift} + \mathcal{F}_{ifd} \\
& + \Delta_s \mathcal{F}_{ifdt} + \Delta_s \varepsilon_{ifdt}
\end{aligned} \tag{2.11}$$

where $\Delta_s x_{j,t} \equiv x_{j,t} - x_{j,t-s} \; \forall j \in \{f, i, f, d, fd, id, if, ifd\}$.

The problem with estimating either (2.10) or (2.11) directly, is that neither approach controls for the product specific marginal cost within the firm, $\mathcal{F}_{ift}$, which is unobserved and correlated with the key variable of interest, the bilateral exchange rate, $\mathcal{F}_{dt}$. In an effort to control for the missing unobserved variable, a typical ERPT estimator is constructed by taking first differences of (2.9) and adding firm-product-year dummies. This yields as an estimating equation

$$\begin{aligned}
\widetilde{\Delta p_{ifdt}} = \; & \widetilde{\Delta \mathcal{F}_{dt}} \\
& + \widetilde{\Delta \mathcal{F}_{fdt}} + \widetilde{\Delta \mathcal{F}_{idt}} + \widetilde{\Delta \mathcal{F}_{ifdt}} + \widetilde{\Delta \varepsilon_{ifdt}}
\end{aligned} \tag{2.12}$$

in which $\widetilde{\Delta x_j} \equiv x_j - \sum_j \Delta x_j / n_J \; \forall j \in \{dt, fdt, idt, ifdt\}$.

We make two comments here. First, the conventional method of ERPT estimation is essentially a sequential partition process applying two sets of high-dimensional fixed effects. That is, (2.12) is equivalent to sequentially adding ifd dummies and ift dummies to (2.9). In a balanced panel, if the covariances between the bilateral exchange rate and the terms in the second line of (2.12) are zero, then this procedure recovers an unbiased estimate of ERPT.

However, more generally, when bilateral ERPT varies systematically with firm-level variables, like firm-level TFP, covariances between the exchange rate and terms in the second line of (2.12) are non-zero (e.g., $cov(\widetilde{\Delta\mathcal{F}_{dt}}, \widetilde{\Delta\mathcal{F}_{ifdt}}) \neq 0$).[17] Further, datasets of firm-product export prices to destinations are unbalanced because the set of destinations served by a firm changes endogenously with exchange rate movements. As we show below, these two features of data on firm-product prices combine to generate biases in ERPT under the approach described by (2.12).

2.4.1 Identification from orthogonal dimensions with an interaction term

We propose an estimator of the markup adjustment in export prices to bilateral exchange rate movements which allows for interactions between unobserved firm-product-time specific factors and bilateral exchange rates in an unbalanced panel. The estimator is free of the bias inherent in the standard approach in the literature. We argue that the presence of an interaction term between the rmb-destination exchange rate and firm-product unobserved variables in the firm's pricing equation implies that the order of partition in a sequential fixed effects procedure matters. If controlling for unobserved firm-product marginal costs, $\mathcal{F}_{ift}$, is the primary concern, destination fixed effects (rather than firm-product-year dummies or S-period differences) need to be applied first.

The estimation procedure is as follows:

1. For each set of firm-product-year observations, demean all variables by destination.
2. For each firm-product-year triplet, formulate a string that records the triplet's associated trade pattern, i.e., the set of destinations for that triplet, e.g., VN-KR-JP.
3. Run a regression of the destination demeaned variables with the trade pattern fixed effect.

$$\widetilde{p}_{ifdt} = \kappa_0 + \kappa_{1,t}\widetilde{e}_{dt} + +\widetilde{X}'_{dt}\kappa_2 + TP_{ift} + \widetilde{u}_{ifdt}$$

where e_{dt} is the bilateral exchange rate (rmb/d_1), X_{dt} is a vector of destination-specific macro variables including local CPI and real GDP, $\widetilde{x_{jt}} \equiv x_{jt} - \sum_j x_{jt}/n_J \; \forall j \in \{d, ifd\}$, and TP_{ift} is a string variable of the set of destinations to which firm f exports product i in year t. Intuitively, the trading pattern strings facilitate apples-to-apples comparisons across sets of firm-product prices in different periods and prevent the introduction of bias associated with endogenously changing trade patterns.

[17] As noted earlier, Berman, Martin, and Mayer (2012) document variation in ERPT as a function of firm-level productivity for French firms.

In the next sections and appendix 2.B.3, we show that even if there are interactions between the exchange rate and unobservable marginal cost, the estimate of $\kappa_{1,t}$ is the price responsiveness to the exchange rate evaluated at the mean of exporters' marginal costs, in other words, the price markup elasticity.

Consistency of the estimator in a balanced panel

The essence of our estimation procedure is to take advantage of multi-dimensional panel data and utilize orthogonal dimensions between variables to exactly identify the parameter of interest. To illustrate how the estimator works, we present some simple stylized examples.

First consider a simple pricing equation in which price varies cross-sectionally by firm and by destination and an interaction between these components is included, but in which there is no common time factor driving the cross-sectional variables, the destination-specific bilateral exchange rate (e_d) and the firm's marginal cost (mc_f).

$$p_{fd} = \beta_0 + \beta_1 e_d + \beta_2 mc_f + \beta_3 e_d * mc_f + u_{fd} \tag{2.13}$$

If we regress p_{fd} on e_d,

$$p_{fd} = \gamma_0 + \gamma_1 e_d + u_{fd}$$

the optimal $\widehat{\gamma}_1$ is given by

$$\begin{aligned}
\widehat{\gamma}_1 &= \frac{\sum_f \sum_d (e_d - \overline{e})(\beta_1 e_d + \beta_2 mc_f + \beta_3 e_d * mc_f)}{\sum_f \sum_d (e_d - \overline{e})^2} \\
&= \beta_1 + \frac{\beta_3 \sum_f \left[mc_f \sum_d (e_d - \overline{e})^2 \right]}{\sum_f \sum_d (e_d - \overline{e})^2} \\
&= \beta_1 + \beta_3 \overline{mc}
\end{aligned}$$

$\widehat{\gamma}_1$ gives the price markup elasticity, i.e., ERPT when marginal cost is evaluated at its mean value across all firms. The estimator controls for the first order effect of cross-sectional variation in mc_f and incorporates the second order interaction effect. The estimated $\widehat{\gamma}_1$ is ERPT evaluated at the mean of the exporters' marginal costs. Our estimator captures the true markup adjustment regardless of whether or not the true underlying pricing equation includes an interaction term. However, if there is an interaction term and conventional methods are applied, the estimate of ERPT will be biased.

Next, we present a more realistic example in which firm and destination variables share a common time-varying factor. This example illustrates that our estimator can cope with the cross-destination comparison problem, i.e., cross-destination comparisons of macro variables including nominal exchange rates and CPI are meaningless. Essentially, we want to use cross destination variation in prices to control for time-varying unobserved marginal costs, but still exploit the intertemporal variation in nominal exchange rate and CPI series to identify the price markup elasticity.

Consider a more realistic pricing equation in which the unobserved variable mc_{ft} shares a common dimension t with the policy variable of interest e_{dt}.

$$p_{fdt} = \beta_0 + \beta_1 e_{dt} + \beta_2 mc_{ft} + \beta_3 e_{dt} * mc_{ft} + u_{fdt}$$

Partitioning out firm-time fixed effects, i.e., demeaning variables along the destination dimension, yields

$$\widetilde{p}_{fdt} = \beta_0 + \beta_1 \widetilde{e}_{dt} + \beta_3 \widetilde{e}_{dt} * mc_{ft} + \widetilde{u}_{fdt}.$$

Taking a second partition for destination fixed effects, i.e., demeaning variables along the firm-time dimension, yields

$$\widetilde{p}^{SFE}_{fdt} = \beta_0 + \beta_1 \widetilde{e}^{SFE}_{dt} + \beta_3 (\widetilde{e}_{dt} * mc_{ft})^{SFE} + \widetilde{u}^{SFE}_{fdt}$$

At this stage, it is important to possess the capability to separate the term $(\widetilde{e}_{dt} * mc_{ft})^{SFE}$ into two components, the variation of our observed variable $\widetilde{e}_{dt}$ and that of the unobserved variable mc_{ft}. As we will show in the next subsection, an incorrect partition order will make the term $(\widetilde{e}_{dt} * mc_{ft})^{SFE}$ non-separable in an unbalanced panel. We have:

$$\begin{aligned}
(\widetilde{e}_{dt} * mc_{ft})^{SFE} &= \frac{1}{n_F n_T n_D}\left[n_F n_T (n_D e_{dt} - \sum_d e_{dt}) mc_{ft} - \sum_t \sum_i (n_D e_{dt} - \sum_d e_{dt}) mc_{ft}\right] \\
&= \frac{1}{n_T n_D}\left[n_T (n_D e_{dt} - \sum_d e_{dt}) mc_{ft} - \sum_t (n_D e_{dt} - \sum_d e_{dt}) \overline{mc_t}\right] \\
\sum_f (\widetilde{e}_{dt} * mc_{ft})^{SFE} &= \frac{n_F}{n_T n_D}\left[n_T (n_D e_{dt} - \sum_d e_{dt}) \overline{mc_t} - \sum_t (n_D e_{dt} - \sum_d e_{dt}) \overline{mc_t}\right] \\
&= \frac{n_F}{n_T n_D}\left[n_T (n_D e_{dt} - \sum_d e_{dt}) - \sum_t (n_D e_{dt} - \sum_d e_{dt})\right] \overline{mc_t}
\end{aligned}$$

If we run the regression $\widetilde{p}^{SFE}_{fdt} = \gamma_0 + \gamma_{1,t} \widetilde{e}^{SFE}_{dt} + \widetilde{u}^{SFE}_{fdt}$, then we are able to estimate the price markup elasticity evaluated at the mean of the unobserved variable, marginal cost, for each time period $\gamma_{1,t} = \beta_1 + \beta_3 \overline{mc_t}$[18]. For example, we can run a regression with time interaction factors or run regressions separately for each time period to obtain the desired estimator.

We next highlight another important property of our estimator. Under our proposed partition order, simply regressing $\widetilde{p}^{SFE}_{fdt}$ on $\widetilde{e}^{SFE}_{dt}$ will yield a consistent estimator of $\gamma =$

[18] The level effect has been differenced out and the interaction effect is evaluated at the average value of the unobserved variable for each time period.

$\beta_1 + \beta_3\overline{mc}$ as long as the volatility of the time component of e_{dt} is not correlated with the time component of the unobserved variable mc_{ft}.

Suppose we run the following regression in which the parameter of interest is assumed to be time invariant, $\widetilde{p}^{SFE}_{fdt} = \gamma_0 + \gamma_1\widetilde{e}^{SFE}_{dt} + \widetilde{u}^{SFE}_{fdt}$. Then the optimal $\widehat{\gamma_1}$ is given by

$$\begin{aligned}
\widehat{\gamma_1} &= \frac{\sum_i\sum_d\sum_t\left\{(\widetilde{e}^{SFE}_{dt} - \overline{\widetilde{e}^{SFE}})\left[\beta_1\widetilde{e}^{SFE}_{dt} + \beta_3(\widetilde{e}_{dt} * mc_{ft})^{SFE}\right]\right\}}{\sum_i\sum_d\sum_t(\widetilde{e}^{SFE}_{dt} - \overline{\widetilde{e}^{SFE}})^2} \\
&= \beta_1 + \beta_3\frac{\sum_i\sum_d\sum_t\left\{(\widetilde{e}^{SFE}_{dt} - \overline{\widetilde{e}^{SFE}})(\widetilde{e}_{dt} * mc_{ft})^{SFE}\right\}}{\sum_i\sum_d\sum_t(\widetilde{e}^{SFE}_{dt} - \overline{\widetilde{e}^{SFE}})^2} \\
&= \beta_1 + \beta_3\frac{\sum_d\sum_t\left\{(\widetilde{e}^{SFE}_{dt} - \overline{\widetilde{e}^{SFE}})^2\overline{mc_t}\right\}}{\sum_d\sum_t(\widetilde{e}^{SFE}_{dt} - \overline{\widetilde{e}^{SFE}})^2}.
\end{aligned}$$

Note that the process e_{dt} can be approximated as $e_{dt} = v_t + v_d + v_t * v_d$. Under this approximation, we could rewrite $\widehat{\gamma_1}$ as[19]

$$\begin{aligned}
\widehat{\gamma_1} &= \beta_1 + \beta_3\frac{\sum_d\sum_t\left\{[(v_t - \frac{\sum_t v_t}{n_T})(v_d - \frac{\sum_d v_d}{n_D})]^2\overline{mc_t}\right\}}{\sum_d\sum_t[(v_t - \frac{\sum_t v_t}{n_T})(v_d - \frac{\sum_d v_d}{n_D})]^2} \\
&= \beta_1 + \beta_3\frac{\sum_t(v_t - \frac{\sum_t v_t}{n_T})^2\overline{mc_t}}{\sum_t(v_t - \frac{\sum_t v_t}{n_T})^2}
\end{aligned}$$

Note that

$$\mathbb{E}[(v_t - \frac{\sum_t v_t}{n_T})^2]\mathbb{E}[\overline{mc_t}] = \mathbb{E}[(v_t - \frac{\sum_t v_t}{n_T})^2\overline{mc_t}] - cov[(v_t - \frac{\sum_t v_t}{n_T})^2, \overline{mc_t}] \tag{2.14}$$

If $cov[(v_t - \frac{\sum_t v_t}{n_T})^2, \overline{mc_t}] = 0$, we can obtain the estimator evaluated at the mean of the unobserved variable $\mathbb{E}[\widehat{\gamma_1}] = \beta_1 + \beta_3\mathbb{E}[mc]$. In the context of our problem, this condition is generally satisfied because the volatility of the exchange rate is not systematically correlated with the level change in firms' marginal costs. We estimate specifications with and without time interaction dummies. Empirically, the difference between these two specifications is small.

[19]
$$\begin{aligned}
\textit{where} \quad \widetilde{e}^{SFE}_{dt} - \overline{\widetilde{e}^{SFE}} &= \frac{1}{n_F n_T n_D}\left[n_F n_T n_D e_{dt} - n_F n_T\sum_d e_{dt} - n_D\sum_t\sum_i e_{dt} + \sum_i\sum_t\sum_d e_{dt}\right] \\
&= \frac{1}{n_T n_D}\left[n_D(n_T e_{dt} - \sum_t e_{dt}) - \sum_d(n_T e_{dt} - \sum_t e_{dt})\right]
\end{aligned}$$

Consistency in an unbalanced panel

Having discussed how our proposed procedure works in a balanced panel, we construct numerical examples to show how and explain why the order of partition matters in an unbalanced panel with interactive unobserved variables.

We construct a three-dimensional numerical example in which the price p_{fdt} is determined by three unobserved factors, $v_{1,f}, v_{1,d}, v_{1,t}$, each varying along a particular dimension (f, d, and t), the bilateral nominal exchange rate, e_{dt}, and the unobserved marginal cost, mc_{ft}. We further allow for interactions between the unobserved marginal cost and the exchange rates. $\mathcal{I}$ is an indicator variable that takes values of 0 or 1. The data generating process for export prices is given below:

$$\begin{aligned} p_{fdt} &= \mathcal{I}_1 v_{1,f} + \mathcal{I}_2 v_{1,d} + \mathcal{I}_3 v_{1,t} + \beta_1 e_{dt} + mc_{ft} + \mathcal{I}_4 \beta_2 e_{dt} * mc_{ft} + u_{fdt} \qquad (2.15)\\ e_{dt} &= v_{1,d} + v_{1,t} + v_{2,d} + v_{2,t} + v_{3,d} * v_{3,t} + v_{4,d} * v_{4,t}\\ mc_{ft} &= v_{1,f} + v_{1,t} + v_{2,f} + v_{2,t} + v_{5,f} * v_{5,t} + v_{4,f} * v_{4,t}\\ v_{j,k} &\sim N(0,1) \quad \forall j \in \{1,2,3,4,5\}, k \in \{f,d,t,fdt\} \end{aligned}$$

In this example, bilateral exchange rates, e_{dt}, and firm specific marginal costs, mc_{ft}, co-move for three reasons. The first macro shock, $v_{1,t}$, directly affects all variables, p_{fdt}, e_{dt}, and mc_{ft}. The second macro shock, $v_{2,t}$, affects exchange rates and firm marginal costs directly; this shock only influences prices through the exchange rate and the marginal cost.[20] The destination-specific shock, $v_{4,d}$, has a cross-sectional impact on exchange rates and marginal costs. Interactions between factors, such as $v_{4,d} * v_{4,t}$ and $v_{4,f} * v_{4,t}$, ensure that variables p_{fdt}, e_{dt}, and mc_{ft} are correlated with each other in all possible dimensions. Finally, the bilateral exchange rate and the firm-level marginal cost each contains an idiosyncratic interaction, $v_{3,d} * v_{3,t}$ and $v_{5,f} * v_{5,t}$, respectively.

The objective is to estimate ERPT given that marginal cost is unobserved; i.e., we want to estimate $\beta_1 + \beta_2 \overline{mc_t}$ where $\overline{mc_t} \equiv \sum_f mc_{ft}/n_F$ when we do not observe $v_{j,k}$, nor do we know the data generating process of e_{dt} and mc_{ft}.

We simulate data for 200 firms, 10 destinations and 10 time periods to compare the results obtained from our estimator and those from the conventional methods, as can be seen from table 2.1. Estimates from the conventional method used in Gopinath and Rigobon (2008), referred to hereafter as the GR methodology, are calculated by first taking S-period differences at the firm-destination level and then adding year fixed effects. We also compare our results (denoted as CHS) to the high-dimensional fixed effects estimator proposed by Correia (2016) which, as shown below, addresses the biases in an unbalanced balance associated with the variation along a single dimension, but cannot correct for the bias associated with the presence of interactions between dimensions. These are reported under

[20] For example, an aggregate origin-country productivity shock.

the column "reghdfe". For the unbalanced panel experiment, we construct a missing data pattern similar to what we observe in the Chinese customs database. In particular, for each firm-year combination, we randomly generate 3 missing values (out of 10) for price p_{fdt}. We repeat this process for firm-destination combinations, and generating 3 missing values among the remaining observations[21].

Table 2.1 Comparison of estimators on simulated data

	Balanced Panel			Unbalanced Panel			Theoretical
$\mathcal{I}_1$ $\mathcal{I}_2$ $\mathcal{I}_3$ $\mathcal{I}_4$	GR	reghdfe	CHS	GR	reghdfe	CHS	
0 0 0 0	1.00***	1.00***	1.00***	1.28***	1.00***	1.00***	1.00
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.03)	
1 1 1 0	1.00***	1.00***	1.00***	1.34***	1.00***	1.00***	1.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	
0 0 0 1	0.93***	0.88***	0.96***	1.37***	0.88***	0.96***	0.96
	(0.04)	(0.03)	(0.06)	(0.04)	(0.04)	(0.11)	
1 1 1 1	1.05***	1.07***	1.03***	1.44***	1.06***	1.03***	1.03
	(0.03)	(0.03)	(0.06)	(0.04)	(0.04)	(0.10)	

Note: Estimated coefficients are calculated from the average of 100 simulations. Theoretical estimates are calculated by $\beta_1 + \beta_2 \overline{mc}$. Results of reghdfe are estimated using the command "`reghdfe p e, absorb(firm_year destination)`". S equals 1 in the balanced panel case.

In table 2.1, the far left column indicates which sources of variation in the data generating process are active in our price equation (2.15). For example, in the first row, by setting all indicator variables to zero, the price is determined by the shocks that drive the exchange rate and the marginal cost. In the second row, firm, destination and time-varying shocks directly impact on prices. Both rows (1) and (2) show that for a balanced panel, the three estimators return the correct estimate of the true ERPT parameter (listed in the last column). However, the conventional method of estimating ERPT, referred to as GR, shows a strong upward bias in an unbalanced panel. In row (3), we simulate the price series with a destination-firm time-varying interacted factor. Finally, row (4) turns on each dimension of the variation and their interactions. In the last row, only the CHS procedure is capable of returning the correct theoretical parameter from the data-generating process. Our simulation shows that one needs to be careful in applying multiple fixed effects in an unbalanced panel with interacted unobserved variables. The order of applying fixed effect matters. If there exists orthogonal dimensions between the observed explanatory variable and the unobserved variable(s), a unique sequence of fixed effects can obtain the true parameter.

We compare analytical decompositions of our method and the conventional method. We evaluate by first employing the conventional approach where the S-period time difference is

[21]The advantage of using two separate processes compared to a random drop at the firm level is that the former allows the structure of missing values to differ at the time and the destination dimension.

taken and then some additional fixed effects are added.

$$\Delta_s^{fd} p_{fdt} = \mathcal{I}_3 \Delta_s^{fd} v_{1,t} + \beta_1 \Delta_s^{fd} e_{dt} + \Delta_s^{fd} mc_{ft} + \mathcal{I}_4 \beta_2 \Delta_s^{fd} (e_{dt} * mc_{ft}) + \Delta_s^{fd} u_{fdt} \tag{2.16}$$

where

$$\begin{aligned}
\Delta_s^{fd} v_{1,t} &= v_{1,t} - v_{1,t-s} \\
\Delta_s^{fd} e_{dt} &= v_{1,t} - v_{1,t-s} + v_{2,t} - v_{2,t-s} + v_{3,d} * (v_{3,t} - v_{3,t-s}) + v_{4,d} * (v_{4,t} - v_{4,t-s}) \\
\Delta_s^{fd} mc_{ft} &= v_{1,t} - v_{1,t-s} + v_{2,t} - v_{2,t-s} + v_{5,f} * (v_{5,t} - v_{5,t-s}) + v_{4,f} * (v_{4,t} - v_{4,t-s}) \\
\Delta_s^{fd} (mc_{ft} * e_{dt}) &= (v_{1,d} + v_{1,t} + v_{2,d} + v_{2,t} + v_{3,d} * v_{3,t} + v_{4,d} * v_{4,t}) \\
&\quad * (v_{1,f} + v_{1,t} + v_{2,f} + v_{2,t} + v_{5,f} * v_{5,t} + v_{4,f} * v_{4,t}) \\
&\quad - (v_{1,d} + v_{1,t-s} + v_{2,d} + v_{2,t-s} + v_{3,d} * v_{3,t-s} + v_{4,d} * v_{4,t-s}) \\
&\quad * (v_{1,f} + v_{1,t-s} + v_{2,f} + v_{2,t-s} + v_{5,f} * v_{5,t-s} + v_{4,f} * v_{4,t-s}) \\
&\neq \Delta_s^{fd} mc_{ft} * \Delta_s^{fd} e_{dt}
\end{aligned}$$

This illustrates that $\Delta_s^{fd} e_{dt}$ is correlated with three unobserved regressors $\Delta_s^{fd} v_{1,t}$, $\Delta_s^{fd} mc_{ft}$ and $\Delta_s^{fd} (mc_{ft} * e_{dt})$. As mc_{ft} is unobserved, it is important that the interaction term can be decomposed into two terms[22] such that we could exploit the relationship similar to (2.14). The resulting estimator is in general biased and the direction of bias cannot be determined.[23] This problem is not limited to S-period difference related estimations but occurs in all cases where multiple fixed effects are needed in an unbalanced multidimensional panel with missing factors interacting with observed explanatory variables. As we have discussed above, taking first differences achieves the same result in terms of eliminating unobserved components as a high level fixed effect. We show in the appendix that fixed effects partitioned in an incorrect order will lead to the same problem.

Reversing the order of these two partitions helps to identify the correct EPRT estimator given the existence of orthogonal dimensions between the unobserved variable mc_{ft} and the

[22]The two terms are in the form of an observed $\Delta_s^{fd} e_{dt}$ interacted with an unknown term varying along firm and time dimensions.

[23]While we could control for two unobserved regressors $\Delta_s^{fd} v_{1,t}$ and $\Delta_s^{fd} mc_{ft}$ by adding firm-time specific fixed effects, we cannot add further interactive fixed effects to control for $\Delta_s^{fd} (mc_{ft} * e_{dt})$ as the S-period lag is firm destination specific.

variable of interest e_{dt}. If we first demean at the destination dimension, we will get

$$\begin{aligned}
\widetilde{p_{fdt}}^{ft} &= \mathcal{I}_2\widetilde{v_{1,d}}^{ft} + \widetilde{e_{dt}}^{ft} + \mathcal{I}_4\widetilde{(e_{dt} * mc_{ft})}^{ft} + \widetilde{u_{fdt}}^{ft} \\
\widetilde{v_{1,d}}^{ft} &= v_{1,d} - \frac{\sum_{d\in\tau^{ft}} v_{1,d}}{n_D^{ft}} \\
\widetilde{e_{dt}}^{ft} &= e_{dt} - \frac{\sum_{d\in\tau^{ft}} e_{dt}}{n_D^{ft}} = \widetilde{v_{1,d}}^{ft} + \widetilde{v_{2,d}}^{ft} + \widetilde{v_{3,d}}^{ft} * v_{3,t} + \widetilde{v_{4,d}}^{ft} * v_{4,t} \\
\widetilde{e_{dt} * mc_{ft}}^{ft} &= (e_{dt} - \frac{\sum_{d\in\tau^{ft}} e_{dt}}{n_D^{ft}}) * mc_{ft} = \widetilde{e_{dt}}^{ft} * mc_{ft} \\
&= \widetilde{e_{dt}}^{ft}(v_{1,f} + v_{1,t} + v_{2,f} + v_{2,t} + v_{5,f} * v_{5,t} + v_{4,f} * v_{4,t})
\end{aligned} \tag{2.17}$$

This operation will lead to exchange rates $\widetilde{e_{dt}}^{fd}$, indexed by fd. However, as a nature of macro variables, e_{dt} cannot vary along the firm dimension. Therefore, we can separate destination varying components of the exchange rate e_{dt} from the missing factor mc_{ft}. The essence of this idea is that most macro variables do not vary at all dimensions, which can be exploited in a micro dataset to construct an estimator of a particular interest.

At this point, it may seem impossible to estimate (2.17) as we have a missing factor $\widetilde{v_{1,d}}^{ft}$ varying at all three dimensions. However, $\widetilde{v_{1,d}}^{ft}$ is special as its variation is limited and depends on the total number of destinations. For example, in the case of 10 destinations, there will be maximum $10 \times 10 = 100$ variations of $\widetilde{v_{1,d}}^{ft}$. If the number of firms is larger than the number of destinations[24], which is generally true in a disaggregated trade dataset, the model (2.17) will still be identifiable.

We propose a trade pattern dummy to control for $\widetilde{v_{1,d}}^{ft}$. The procedure involves three steps. First, for each firm and time combination, record a list of exported destination names (or indices). Second, join each list as a string in a alphabetic (or numerical) order. Third, for each firm and time combination, prefix the destination name (or number) to the combined string in step two for each destination within this combination. This string indicator can be used to create dummies or execute the second partition.

Notably, in the context of exchange rate pass through estimation, we do not interpret the conventional method as a biased estimator. Especially, directly estimating regression specification (2.16) will capture how price generally reacts to exchange rate shocks. As the first three factors simultaneously change other variables in the pricing equation, the coefficient on S-period exchange rate will be a combination of effects of exchange rates and other variables related to factors that also affect exchange rates. Our estimator separates the destination specific shock component and evaluates the ERPT at the mean value of the firm specific factor.

[24] Another practical reason is that we do not observe firms randomly change the set of countries they exported to. The theoretical maximum number is unlikely to be reached in an empirical dataset.

2.4.2 Mapping to the general pricing equation

In the following two subsections, we compare the conventional estimation method with our estimation method under the general pricing equation framework proposed in section 2.3.

Bilateral ERPT

From the general pricing equation (2.8), we have[25]

$$\Delta log(P^{rmb}_{CHN,dt}) = \beta_{1,d}\Delta log(\mathcal{S}_{CHN,dt}) + \beta_{2,d}\Delta log(\mathcal{S}_{CHN,-dt}) + \beta_{3,d}\Delta log(\mathcal{D}_{CHN,dt}) + \beta_{4,d}\Delta log(\mathcal{D}_{CHN,-dt}) + \Delta log(\varepsilon_{CHN,dt})$$

The conventional method controls for observed economic fundamentals X_{dt} using destination CPI, real GDP and import-to-GDP ratio and runs the following regression pooling all destinations.

$$\Delta log(P^{rmb}_{CHN,dt}) = \gamma_0 + \gamma_1 \Delta log(e_{\frac{CHN}{d},t}) + \gamma_2 \Delta log(X_{dt}) + v_{dt} \tag{2.18}$$

The obtained γ_1 can be interpreted as the general effect of bilateral exchange rate movements.

$$\gamma_1 = \sum_{d \neq CHN} \frac{1}{n_D}\left[\beta_{1,d}\frac{\partial log(\mathcal{S}_{CHN,dt})}{\partial log(X_{CHN,t})}\frac{\partial log(X_{CHN,t})}{\partial log(e_{\frac{CHN}{d},t})} + \beta_{3,d}\frac{\partial log(\mathcal{D}_{CHN,dt})}{\partial log(e_{\frac{CHN}{d},t})}\right] + Bias \tag{2.19}$$

Without any cost shocks and demand shocks, the export price denominated in the producer's currency should not move with changes of nominal bilateral exchange rates.[26] The bias component reflects the fact that we cannot perfectly control for all demand factors in the destination for two reasons: (a) we do not observe price changes of local competitors; (b) we do not observe price changes of competitors of exporters from other countries.

Demand-side bilateral ERPT

Similarly, we can derive the destination demeaned version of equation (2.8) as

$$\widetilde{log(P^{rmb}_{CHN,dt})} = \widetilde{\beta_{0,d}} + \beta_{1,d}\widetilde{log(\mathcal{S}_{CHN,dt})} + \beta_{2,d}\widetilde{log(\mathcal{S}_{CHN,-dt})} + \beta_{3,d}\widetilde{log(\mathcal{D}_{CHN,dt})} + \beta_{4,d}\widetilde{log(\mathcal{D}_{CHN,-dt})} + \widetilde{log(\varepsilon_{CHN,dt})} \tag{2.20}$$

[25]We drop the subscript if for clarity. CHN indicates the exporting country is China.
[26]After controlling for demand shocks, there is no incentive for pricing-to-market.

where

$$log(\widetilde{\mathcal{S}_{CHN,dt}}) \equiv log(\mathcal{S}_{CHN,dt}) - \frac{1}{N}\sum_{d\neq CHN}\frac{\beta_{1,d}}{\beta_{1,d}}log(\mathcal{S}_{CHN,dt}) \tag{2.21}$$

$$log(\widetilde{\mathcal{S}_{CHN,-d_1 t}}) \equiv log(\mathcal{S}_{CHN,-d_1 t}) - \frac{1}{N}\sum_{d\neq CHN}\frac{\beta_{2,d}}{\beta_{2,d_1}}log(\mathcal{S}_{CHN,-dt}) \tag{2.22}$$

The key relationship our estimator exploits is

$$\beta_{1,d}log(\mathcal{S}_{CHN,dt}) + \beta_{2,d}log(\mathcal{S}_{CHN,-dt}) = \beta_{1,d}log(\mathcal{S}_{CHN,dt}) + \beta_{2,d}log(\mathcal{S}_{CHN,-dt}) \quad \forall d \neq CHN \tag{2.23}$$

That is, the weight of each bilateral supply effect may differ but the total supply effect must be the same for the same product. For our empirical analysis, we assumed that at firm-product-custom-level, the quality of the good sold to different countries is not systematically different.

Equation (2.23) implies that the term $\beta_{1,d}log(\widetilde{\mathcal{S}_{CHN,dt}}) + \beta_{2,d}log(\widetilde{\mathcal{S}_{CHN,-dt}})$ must be zero. Therefore, we controlled for both bilateral and multilateral supply shocks.

We now introduce our estimator that exploits the cross destination variations.

$$log(\widetilde{P^{rmb}_{CHN,dt}}) = \gamma_0 + \gamma_1 log(\widetilde{e_{\frac{CHN}{d},t}}) + \gamma_2 log(\widetilde{X_{dt}}) + FE_d + v_{dt} \tag{2.24}$$

where

$$log(\widetilde{e_{\frac{CHN}{d},t}}) \equiv log(e_{\frac{CHN}{d},t}) - \frac{1}{N}\sum_{d\neq CHN}log(e_{\frac{CHN}{d},t}) \tag{2.25}$$

To see what our estimator captures, consider the case of a general appreciation of the currency k against all other currencies, such that $dlog(e_{\frac{d}{k},t}) = dlog(e_{\frac{CHN}{k},t}) = dlog(e_{\frac{j}{k},t}) \quad \forall j \neq k$.

$$\frac{\partial log(\widetilde{e_{\frac{CHN}{d},t}})}{\partial log(e_{\frac{CHN}{k},t})} = \begin{cases} -\frac{1}{N} & if \quad d \neq k \\ 1 & if \quad d = k \end{cases} \tag{2.26}$$

$$\frac{dlog(\widetilde{P^{rmb}_{CHN,dt}})}{dlog(e_{\frac{CHN}{k},t})} = \beta_{3,d}\frac{\partial log(\widetilde{\mathcal{D}_{CHN,dt}})}{\partial log(e_{\frac{CHN}{k},t})} + \beta_{4,d}\frac{dlog(\widetilde{\mathcal{D}_{CHN,-dt}})}{dlog(e_{\frac{CHN}{k},t})} \tag{2.27}$$

$$= \begin{cases} -\frac{1}{N}\beta_{3,k}\frac{\partial log\left[g_{\mathcal{D},k}(e_{\frac{CHN}{k},t}, X_{d_k,t})\right]}{\partial log(e_{\frac{CHN}{k},t})} + \beta_{4,d}\frac{dlog(\widetilde{\mathcal{D}_{CHN,-dt}})}{dlog(e_{\frac{CHN}{k},t})} & if \quad d \neq k \\ \beta_{3,k}\frac{\partial log\left[g_{\mathcal{D},k}(e_{\frac{CHN}{k},t}, X_{k,t})\right]}{\partial log(e_{\frac{CHN}{k},t})} + \beta_{4,d}\frac{dlog(\widetilde{\mathcal{D}_{CHN,-dt}})}{dlog(e_{\frac{CHN}{k},t})} & if \quad d = k \end{cases} \tag{2.28}$$

The derivation of the last step is given in the appendix by equation (2.47). If we ignore the second term, our estimator will capture exactly the demand effect of bilateral exchange rate movements, i.e., $\gamma_{1,k} = \beta_{3,k} \frac{\partial log\left[g_{\mathcal{D},k}(e_{\frac{CHN}{k},t}, X_{k,t})\right]}{\partial log(e_{\frac{CHN}{k},t})}$.

In general, our estimator γ_1 will be an observation weighted average of country specific pass through $\gamma_{1,d}$ plus a bias due to the fact that we cannot perfectly control for price changes of competitors from other countries. We will discuss this bias in more detail in section 2.7.

$$\gamma_1 = \sum_{d \neq CHN} \frac{1}{n_D} \left\{ \beta_{3,d} \frac{\partial log\left[g_{\mathcal{D},d}(e_{\frac{CHN}{d},t}, X_{d,t})\right]}{\partial log(e_{\frac{CHN}{d},t})} \right\} + Bias \tag{2.29}$$

2.5 Data

To construct the dataset in this paper, we merge the Chinese customs database, the universe of import and export transactions for China from 2000 to 2011, with macroeconomic data from the World Bank and variables derived from the BIS's nominal effective exchange rate series.

The Chinese custom database covers all entries of China's exports and imports at the firm and harmonized system (HS08) 8-digit product level annually from 2000 to 2011.[27] The data reported include the export value and quantity, reported by the Chinese authorities in US dollars and Chinese-language quantity classifiers, respectively. We convert annual export values in US dollars to rmb. Because the dataset does not report the transaction-level price, the analysis of ERPT uses the unit value in rmb by firm-product-destination as the export price.[28] In addition, the database contains information on the firm's name, a unique numerical identifier for each firm, the location of production, the mode of shipment (cargo ship, air, etc.), and a Chinese government production classification. The database also reports information on the classification of Chinese firms by capital formation.[29]

Products in the dataset are indexed using HS08 codes with around 7,000 identifiers each year. The total number of active exporters had increased dramatically over the period from 62,770 in 2000 to 253,893 in 2011. We track the total number of actively traded products by counting unique product-exporter pairs and find this measure increases roughly at the same pace as the number of exporters from roughly 904 thousand in 2000 to 4.128 million in 2011. The total exported value measured in dollars had increased tenfold from 2000 to 2011, while the total quantity traded had grown at a slightly slower rate implying a gradual increase in

[27] The database consists of monthly transactions by firm-product-destination for 2000-2006 but only reports annual data for 2007-2011. We aggregate the monthly data for 2000-2006 to the annual frequency in this study.

[28] We discuss the quantity classifiers used to construct unit values in detail in appendix 2.C.1.

[29] Categorical information on capital formation includes private firms, state-owned firms, Chinese-foreign joint ventures, foreign-invested enterprises, etc. Unfortunately, this detailed information is only provided up until 2006.

the average unit value of goods exported from China. See table 2.C.2 in appendix 2.C for additional details.

The key to identifying price responses to exchange rate movements for our estimator relies on cross-destination market variation in prices. Following Mayer, Melitz, and Ottaviano (2014), we document in table 2.1 that a "happy few" exporters are responsible for most of China's exports. The top panel provides a breakdown of the number of export transactions by the count of products and destinations served by a Chinese exporting firm. The bottom panel presents the respective shares of export value by firms that differ by the count of exported products and foreign markets reached. Overall, we see that multi-destination exporters account for two-thirds of export transactions (row 5 of the top panel of table 2.1, 33.19 +13.61+ 20.27%) and are responsible for 92% of export value (row 5 of the bottom panel of table 2.1, 13.54+10.86+67.72%).[30] These statistics highlight two important facts: (1) the identification scheme based on multi-destination exporters uses observations from those firms that are most important to China's trade and (2) the vast majority of firms are not single-product exporters.[31] The shares of export transactions and export value by the count of products and destination markets are relatively stable across years in our sample period. As our identification strategy relies on multiple destinations, we drop all observations on products that were exported to only one destination by a firm. Further, we drop observations for products exported to the United States and to Hong Kong because China's rmb was pegged to the US dollar (and by extension to the HK dollar) from 2000 until 2005.

An advantage of using Chinese customs data to study price changes is that the use of information-conveying count and mass classifiers to record quantity is a feature of the Chinese language. Relative to previous studies which construct price as a unit value (export value/export quantity) from data in which quantity is measured by weight (Berman, Martin, and Mayer (2012)) or in a combination of weight and units (Amiti, Itskhoki, and Konings (2014)), the Chinese customs authority reports 30 distinct types of quantity classifiers that exist in the Chinese language. In practice, this implies that the calculated unit value is the price per unit for 42 percent of transactions in our dataset. Furthermore, for the remaining 58 percent of transactions that use value per kilogram or value per liter, etc. as the price, the detailed information on Chinese quantity classifiers helps to ensure a better proxy for prices

[30]Conversely, we see that transactions by single-destination firms account for a small share of total Chinese export value. In the top left cell of the top panel of table 2.1, we observe that 17.32% of observations on exports in the Chinese customs dataset were articles exported to a single destination by a single product firm. However, these transactions comprised only 1.51% of Chinese export value in 2003. The bottom row of the top panel shows that one-third of export transactions in 2003 were products exported by a firm to a single destination. However, the last row of the bottom panel indicates that the value of these transactions by single-destination exporters was only 7.88% of total Chinese exports.

[31] This implies that the calculation of firm-level productivity that relies on input shares from single-product firms is using information from the type of firms that is relatively rare empirically and is not representative of the typical exporting firm.

Table 2.1 Multi-product, multi-destination exporters (2003)

	No. of Products	Number of Countries 1	2-5	6-10	10+	Total
by Share of Exporters	1	17.32	7.75	1.88	1.3	28.25
	2-5	11.02	17.28	5.88	5	39.18
	6-10	2.33	4.73	2.83	3.48	13.37
	10+	2.26	3.42	3.01	10.5	19.19
	Total	32.94	33.19	13.61	20.27	100
by Share of Exports	1	1.51	1.75	1.31	1.49	6.06
	2-5	2.57	5.19	4.25	8.42	20.43
	6-10	1.47	2.85	2.11	8.06	14.49
	10+	2.34	3.75	3.19	49.75	59.01
	Total	7.88	13.54	10.86	67.72	100

Note: Each cell in the top panel shows the percentage of observations in the Chinese customs data in 2003 that fall under the relevant description. The bottom panel presents the corresponding value of exports.

using unit values. See appendix 2.C.1 for details of the Chinese quantity measures in our data.

We calculate the price change (unit value change) at the firm-product-destination level between t and $t+1$ and summarize the relevant statistics across industries in table 2.2. First, the mean price change across industries is slightly above one percent per year, indicating a gradual increase in Chinese export prices over the sample period. However, the median price change is zero, which is consistent with the observation of a high level of price stability of internationally traded goods from other datasets. There is substantial heterogeneity in the distribution of price changes across industries. More sophisticated products, such as machinery appliances, and luxury products, such as precious stones and leather articles, have a greater dispersion of price changes compared to commodities and raw materials. More information on the frequency of price changes and the magnitude of price changes at different frequencies can be found in appendix 2.C. We drop observations with changes of prices denominated in rmb less than $\pm 5\%$.

Macroeconomic variables on real GDP in constant 2005 US dollars, the import to GDP ratio, and CPI in all destination countries and in China (normalized so that CPI=100 in 2010 for all series) come from the World Bank. We construct the bilateral nominal exchange rate in rmb per unit of destination currency from China's official exchange rate (rmb per US$) and each destination country's official exchange rate in local currency units per US$ (all series are the yearly average). Similarly, the bilateral real exchange rate in rmb per local currency unit comes from the World Bank. These variables are available for the 161 destination countries in our sample. When we examine the impact of multilateral exchange rate movements, we incorporate a variable derived from the broad nominal effective exchange rate (NEER) series provided by the BIS. These series are geometric weighted averages of bilateral exchange rates adjusted by relative consumer prices. As this measure is only available for 42 destinations in our dataset during our sampling period, the analysis of multilateral exchange rate movements is restricted to these destinations.

In our empirical analysis, we use the bilateral nominal exchange rate, the local destination CPI and a variable derived from the nominal effective exchange rate as the main variables of interest. This approach is motivated by our desire to decompose price changes into factors that are largely destination market-specific (i.e., CPI) versus those that respond to fundamental shocks originating in two or more locations (i.e., nominal bilateral and multilateral exchange rates). A concern with estimating ERPT across countries is that nominal series, such as nominal bilateral exchange rates and CPI indices, provide meaningful information about changes *within a country* over time, but cannot be directly compared across countries. The typical ERPT regression which uses the real bilateral exchange rate implicitly imposes a linear relationship between each nominal bilateral exchange rate (origin currency/destination currency) and the ratio of CPI indices between destination and origin with a coefficient equal to one.[32] An alternative approach to estimating ERPT using the real exchange rate involves taking a first or S-period time difference; as discussed in section 2.4.1, this procedure introduces potential biases in an unbalanced multi-dimensional panel. Our proposed estimator is robust to this type of mis-specification error and allows us to recover the unbiased impact of nominal exchange rates and CPI movements on export prices.[33]

[32]Real exchange rate series which embed this restriction are highly correlated with nominal exchange rates. Because nominal exchange rate series are significantly more volatile over time than the ratio of CPI indices, this means that movements in the real exchange rate are primarily driven by fluctuations in nominal exchange rates. It is not clear if restricting these two variables with significantly different volatilities to have a one-to-one linear relationship when estimating ERPT is a good assumption.

[33]The compatibility problem in data series across destinations (i.e., the cross-destination comparability of series) is addressed in our approach by assuming that each bilateral exchange rate series can be written as a compatible series which consists of the observed bilateral series and an unobserved component that varies only along the destination dimension, i.e., $e_{dt}^{compatible} = e_{dt}^{nominal} - \mu_d$. Under our approach, the destination specific component μ_d, which makes the different bilateral series compatible across markets, is absorbed into the trade pattern dummies.

Table 2.2 Price changes across sectors (percentage)

Category	HS code	N	mean	s.d.	p25	p50	p75
Live animals; animal products	01-05	84,863	1.1	87.7	-22.9	0.0	25.1
Vegetable products	06-14	302,530	1.3	65.9	-23.9	0.0	26.6
Animal/vegetable fats, oils	15	6,736	1.2	53.7	-19.9	0.0	23.4
Prepared foodstuffs	16-24	245,712	0.9	57.0	-17.0	0.0	18.8
Mineral products	25-27	79,844	1.3	67.2	-22.3	0.0	25.0
Chemical products	28-38	1,256,653	0.9	85.7	-22.8	0.0	24.6
Plastics/rubber articles	39-40	1,673,590	1.0	103.2	-32.9	0.0	35.2
Rawhides/leather articles, etc.	41-43	721,354	2.6	123.3	-52.2	0.3	57.5
Wood and articles of wood	44-46	372,609	0.8	72.6	-26.5	0.0	28.2
Pulp of wood/other fibrous cellulosic material	47-49	510,673	1.0	101.5	-35.3	0.0	37.5
Textile and textile articles	50-63	4,662,971	1.2	75.3	-29.1	0.0	31.5
Footwear, headgear, etc.	64-67	775,858	1.9	87.6	-31.9	0.0	35.5
Misc. manufactured articles	68-70	902,709	-0.3	102.1	-34.2	0.0	35.2
Precious or semi-precious stones	71	91,301	1.6	140.8	-52.3	0.0	55.5
Base metals and articles of base metals	72-83	2,583,361	1.2	85.6	-29.6	0.1	31.9
Machinery and mechanical appliances, etc.	84-85	4,603,775	0.9	110.5	-34.0	0.0	35.7
Vehicles, aircraft, etc.	86-89	601,186	1.0	81.6	-26.0	0.1	27.7
Optical, photographic, etc.	90-92	842,866	0.8	123.4	-35.6	0.0	37.3
Arms and ammunition	93	2,237	1.3	89.1	-34.6	0.8	36.0
Articles of stone, plaster, etc.	94-96	2,439,015	1.5	105.8	-37.5	0.0	40.2
Others	97+	13852	0.0	156.8	-58.3	0.0	58.3
Total		2.28E+07	1.1	96.4	-31.5	0.0	33.7

2.6 Results on Bilateral Exchange Rate Movements

In this section, we present our estimates of price adjustments in response to bilateral exchange rate movements. We begin by presenting results conditional on price changes between periods t and $t+s$ that exceed $\pm 5\%$ in order to focus on pass-through in the absence of nominal rigidities. This approach closely follows Gopinath and Rigobon (2008) and facilitates a comparison across methodologies. We then proceed by presenting estimates of the markup elasticity of export prices to exchange rate movements and decompose the effects of exchange rate shocks by comparing the estimates under different methodologies.

Table 2.1 presents estimates of ERPT conditional on price changes, which we refer to as the Gopinath-Rigobon (GR) method. These estimates represent the average export price adjustment in response to a bilateral exchange rate shock given that Chinese exporters have changed their prices.[34] Beginning with column (1), we find a 100% depreciation of the rmb

[34] Both Gopinath and Rigobon (2008) and Fitzgerald and Haller (2014) emphasize the role of nominal rigidities in price stability in response to exchange rate movements. To provide a comparable analysis to their work, we begin by applying a data filter to the unbalanced panel of Chinese export data. For each product-firm-destination

against China's trading partner will increase the price denominated in rmb by 14.5%, which implies a pass through of 86.5% into the local destination currency price. While this estimate is substantially lower than the estimates using aggregate and disaggregated price indices[35], these estimates are consistent with recent findings using firm level data [e.g., Berman, Martin, and Mayer (2012) Amiti, Itskhoki, and Konings (2014)]. The coefficients on real GDP and the import share of GDP represent the destination market-specific demand effect and have the expected positive signs. Finally, Chinese firms raise their destination market prices in response to local price growth. A one percent increase in destination market CPI is passed through as a 0.2 percent increase in the export price.

The second column estimates the impact of a bilateral real exchange rate movement on prices. Surprisingly, the pass through of the bilateral real exchange rate is negative, -1.2%, in contrast to the small positive estimates in Berman, Martin, and Mayer (2012), Chatterjee, Dix-Carneiro, and Vichyanond (2013), Chen and Juvenal (2016).

Turning to columns (3) and (4), we split our data sample into two time periods, 2000-2005 and 2006-2011. These periods correspond to the period in which the rmb was pegged to the US dollar and the period in which the rmb was a managed float. The estimated response of export prices to bilateral nominal exchange rate movements is around 0.2 in both periods. This implies exchange rate pass through in both periods is around 80 percent. More interestingly, we observe a different coefficient on destination CPI before and after the rmb is unpegged from the US dollar. The smaller coefficient on destination CPI before the release of the peg of a 0.07 percent increase in export prices in response to a 1 percent increase in destination CPI relative to the substantial 0.25 percent increase in export prices after the release of the peg suggests that exporting firms switched their pricing behavior away from a uniform price based on the rmb/dollar rate toward a pricing strategy based on destination market conditions.

combination, we filter out absolute price changes <5%. This implies that pass-through estimates will be based on S-period differences in prices and the corresponding cumulated change in the exchange rate and other macro variables over the same period. The length of the S-period difference can vary within a firm-product-destination triplet and across these triplets. That is, for a single firm-product-destination triplet, we might observe S-period differences of 2 years, 4 years and 3 years within the 12 year panel. Our application of both the GR methodology and the CHS estimator will use all of these differences to estimate price responsiveness. As noted previously, a strength of the CHS estimator is that it works well in highly-unbalanced panels.

[35] Knetter (1989), Knetter (1993), Goldberg and Campa (2006)

Table 2.1 Estimating pass through conditional on price changes

	(1) Full Sample	(2) Full Sample	(3) Year ≤ 2005	(4) Year > 2005
Bilateral nominal exchange rate	0.145***		0.185***	0.201***
	(0.00398)		(0.00703)	(0.00521)
Destination CPI	0.191***		0.0742***	0.258***
	(0.00726)		(0.0124)	(0.00907)
Destination real GDP	0.0588***	0.194***	0.309***	0.00566
	(0.00764)	(0.00711)	(0.0194)	(0.00861)
Import-to-GDP ratio	0.205***	0.190***	0.201***	0.211***
	(0.00300)	(0.00304)	(0.00811)	(0.00327)
Bilateral real exchange rate		-0.0123***		
		(0.00401)		
Observations	4,722,665	4,722,665	1,075,599	3,647,066
S Period Difference	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes
Conditional on Price Change	Yes	Yes	Yes	Yes

Notes: The dependent variable and all regressors are the S-difference of the logged level. The bilateral exchange rate is defined as rmb per unit of foreign (destination) currency. Standard errors are clustered at the product-firm-destination level reported in parentheses. Statistical significance at the 1, 5 and 10 percent level indicated by ***, **, and *. The significance of results is robust to different levels of clustering.

More generally, we interpret our estimates in table 2.1, based on the GR methodology with S-period price differences, as the impact on exporter prices of exchange rate movements coming through both demand and cost channels. A bilateral shock between the Chinese rmb and China's trading partner's currency will have two direct impacts, through demand and through costs, on the profitability of exporters. On the demand side, if the exporter held its price in rmb fixed under an appreciation of the rmb, then this choice to engage in complete pass-through would transmit the exchange rate shock to foreign consumers as an increase in the foreign currency price in the foreign market. In turn, this foreign currency price increase would reduce the quantity demanded and the firm's profit. Thus, the choice of how much to adjust the export price is a firm response to destination market demand conditions. On the cost side, the impact of an rmb appreciation directly reduces the cost of imported inputs if foreign exporters of these products pass through the exchange rate shock.

At the same time, there are general equilibrium effects which depend upon the composition of the bilateral exchange rate shocks and the origin of the shock. In order to understand the general equilibrium effects, we need to decompose the types of shocks that drive the exchange rate change. For example, in column (1) the coefficient on the destination market inflation measure has the same sign and a similar magnitude as the coefficient of the nominal bilateral exchange rate. That is, holding real output, the import share, and the bilateral exchange rate unchanged, a change in foreign market inflation would have a similar effect on export prices from China as a change in the exchange rate. One could think of two scenarios in a standard two-country open economy model in which the above condition is satisfied. First, suppose the foreign country suffers from a temporary negative productivity shock and its central bank chooses to stabilize the output gap. The foreign central bank takes an expansionary monetary policy which boosts inflation and demand. As a result, domestic firms in the country and Chinese exporters choose to increase their prices. Alternatively, suppose Chinese exporters suffer from an exogenous increase in marginal costs. *Ceteris paribus*, Chinese exporters could choose to pass through their cost shocks into export prices for goods sold in foreign markets. The difficulty for understanding exchange ERPT is that neither CPI nor bilateral exchange rate is necessarily an exogenous variable as recently highlighted by Forbes, Hjortsoe, and Nenova (2015).

Moreover, the exporters' pass through will depend on how other competitors in the destination market react. For example, the ERPT of US import prices will depend on the movement of bilateral exchange rates with other countries and the market share of the firm as estimated in Auer and Schoenle (2016). Therefore, the composition of the bilateral exchange rate between the destination country and its trading partners other than China should matter for Chinese export prices. We return to this problem in section 2.7.

Table 2.2 offers a comparison of the results based on the GR methodology and our estimator (denoted CHS). Our estimator exploits cross destination country variation in prices and differences out those factors that are common to all destinations. Under the assumption that the HS08 product exported by a firm to different destinations has the same marginal production cost, then our estimator eliminates the direct and indirect effect of supply-side shocks as well as any demand shock that is common to all of the firm's destination markets.[36] That is, our estimator presents the effect of exchange rate changes on export prices that operates through time variation in destination-specific demand. Columns (1) and (2) of 2.2 are identical to columns (1) and (2) of table 2.1 and are included for comparison purpose. Column (3) reports the markup elasticity of the exporter price of 5.6% in response to a destination currency appreciation relative to the rmb. This is about 40% of the total exchange rate pass through estimated with the GR method. Conditional on a price change in rmb, exchange rate shocks driven by the trading partner-specific demand imply a change in the export price markup of only 5.6% implying that 94% of the currency movement passes through to local currency prices. Interestingly, we also estimate a small adjustment of the export price to local CPI growth; a one percent increase in local CPI implies only a 0.05 percent increase in the price to the destination. The difference in the estimated coefficients on CPI in columns (1) versus (3) arises because the CHS estimator removes the global trend in the exporter's price associated with global CPI movements and isolates the

[36]More precisely, as long as any marginal cost difference for an HS08 product exported by a firm to different destination markets is not systematically correlated with movements in exchange rates, then our estimator is consistent.

local component. Turning to column (4), we see that the price markup response to the real exchange rate of 5% is as one would expect given the elasticities of export price markups to the nominal exchange rate and to local CPI are 5% in column (3).

Table 2.2 Total ERPT compared with demand-driven ERPT

	(1) GR	(2) GR	(3) CHS	(4) CHS
Bilateral nominal exchange rates	0.145***		0.0559***	
	(0.00398)		(0.00824)	
Destination CPI	0.191***		0.0516***	
	(0.00726)		(0.00972)	
Destination real GDP	0.0588***	0.194***	0.0257**	0.0235**
	(0.00764)	(0.00711)	(0.0110)	(0.0105)
Import-to-GDP ratio	0.205***	0.190***	0.00447	0.00446
	(0.00300)	(0.00304)	(0.00667)	(0.00667)
Bilateral real exchange rates		-0.0123***		0.0548***
		(0.00401)		0.0548***
Observations	4,722,665	4,722,665	10,348,597	10,348,597
S Period Difference	Yes	Yes	No	No
Clustered SE	Yes	Yes	Yes	Yes
Conditional on Price Change	Yes	Yes	Yes	Yes
Sequential FE			Yes	Yes

Notes: For GR, the dependent variable and all regressors are the S-difference of the logged level. For CHS, the dependent variable and all regressors are in logs. Application of the same 5% price change filter implies the two datasets contain identical information, with the GR estimation dataset comprised of observations of t to $t+s$ differences while the CHS method records twice as many observations because both t and $t+s$ observations appear in the logged level. The bilateral exchange rate is defined as rmb per unit of foreign (destination) currency. Standard errors are clustered at the product-firm-destination level reported in parentheses. Statistical significance at the 1, 5 and 10 percent level indicated by ***, **, and *. The significance of results is robust to different levels of clustering.

2.6.1 Heterogeneity in price markup elasticities

In this section, we present results from the CHS estimator to document heterogeneity in the price markup elasticity to bilateral exchange rate movements (also referred to as demand-side ERPT) for different types of firms and products.

We begin by examining the heterogeneity across Chinese exporting firms according to a method of classification that is somewhat unique to China.[37] Unlike the naming conventions for businesses in English speaking countries, firm names in Chinese include Chinese characters that directly provide information about the types of business activities in which the firms engage. For example, firms which conduct specialized import and export

[37]See appendix 2.C.2 for a detailed description of the classification method. Previous work using this classification includes Ahn, Khandelwal, and Wei (2011), Crowley, Song, and Meng (2016), and Lu, Tao, and Zhang (2013)

activities and serve as intermediaries for smaller manufacturing firms often distinguish themselves by including the characters for import and export in their names. We filter these firms using a keyword list described in appendix 2.C.2 and label these firms as “trading” firms and label the rest as “non-trading” firms. Although we use the term “trading firms” to describe those firms that engage in trading and intermediary activities, some of these firms are manufacturers who export their own output. The term “non-trading” refers to firms that manufacture their own products for export. Many of these firms can be directly linked to their production data in the Chinese Industrial Survey, but smaller manufacturing firms, which are not surveyed every year, are also included in the group of “non-trading firms.” As noted previously, an advantage of the CHS estimator is that it can be used to estimate price adjustments to exchange rate movements for trading firms that do not produce their own output and for smaller firms that do not report production data to the government.

Table 2.3 reports that the export prices of trading firms are more responsive to exchange rate movements than those of non-trading firms. The results suggest that trading firms—specialized exporters that reach a larger number of foreign destinations than non-trading firms—appear to use their extensive information on local market conditions to optimally set prices. The extent of pass through into import prices in destination currency is more incomplete (lower) compared to non-trading firms. Column (1) reproduces the results from table 2.2 column (3) for comparability. In column (2), we report that for trading firms, the export price in rmb falls 0.7 percent in response to a 10 percent appreciation of the rmb. This is slightly more than twice the magnitude of the price responsiveness of non-trading manufacturing firms reported in column (3). In addition to reducing markups more in response to an appreciation of the rmb, trading firms raise their export prices more in response to local CPI growth. In column (2), we report that a one percent CPI growth is associated with an increase in the export price of 0.08 percent. This response to destination market conditions is considerably larger than the elasticity of 0.025 for non-trading firms reported in column (3).

Columns (4) - (7) document two important facts about the price responsiveness of Chinese firms. First, price markups were largely unresponsive to bilateral exchange rate movements when China maintained a fixed exchange rate against the dollar. Second, the difference in the elasticity of price markups for trading versus non-trading firms is even larger post-2005 when China began to allow the rmb to float against the US dollar. Columns (4) and (6) report estimates for trading and non-trading firms, respectively, over the dollar-peg regime of 2000-2005. For both groups of firms, the estimated markup elasticity is close to zero and not statistically different from zero. We emphasize that all estimation samples excluded exports to the U.S. and Hong Kong. Thus, even during the fixed exchange rate regime period, there is substantial variation in the bilateral rmb-euro, rmb-yen, rmb-sterling, etc. rates. Interestingly, these exchange rate changes do not materialize into price markup adjustments that vary across destinations. One possible explanation for this lack of price variation across destinations might be that Chinese firms invoiced their transactions to all destinations in US dollars.[38] Because of the peg with the US dollar, prior to 2005 Chinese exporters practiced an extreme form of pure producer currency pricing.

Interestingly, a pricing strategy of destination-specific markup adjustments in response to bilateral currency changes seems to have become important after 2005. Column (5) shows that in response to a 100% rmb appreciation, trading firms cut their prices by 12.3%. The idea that firms should optimally cut markups to maintain their market shares in the face of a depreciation of the importing (destination) country's currency has long been suggested, but attempts to document it empirically have yielded mixed results. Finally, column (7) indicates that this increased price responsiveness is not limited to trading firms; the export price of non-trading manufacturers also became more responsive to bilateral exchange rate movements after 2005. Their prices remained less responsive than those of trading firms, consistent with the idea that they might have less destination-specific information available to them for developing their pricing rules.[39]

Table 2.3 Heterogeneity in price markup elasticities by firm organizational structure

	(1) Full Sample	(2) Trading	(3) Non-trading	(4) Trading ≤ 2005	(5) Trading > 2005	(6) Non-trading ≤ 2005	(7) Non-trading > 2005
Bilateral nominal exchange rates	0.0559***	0.0744***	0.0357***	0.0135	0.123***	0.0203	0.0697***
	(0.00824)	(0.0127)	(0.0120)	(0.0245)	(0.0205)	(0.0258)	(0.0200)
Destination CPI	0.0516***	0.0799***	0.0254*	0.0356	0.0344	0.0258	0.0502
	(0.00972)	(0.0146)	(0.0146)	(0.0316)	(0.0316)	(0.0318)	(0.0312)
Destination real GDP	0.0257**	0.0229	0.0282*	-0.000984	0.0128	0.0923*	-0.0116
	(0.0110)	(0.0167)	(0.0165)	(0.0567)	(0.0275)	(0.0510)	(0.0291)
Import-to-GDP ratio	0.00447	0.0122	-0.00162	-0.00591	0.0427***	0.00883	0.0309*
	(0.00667)	(0.0101)	(0.00983)	(0.0285)	(0.0164)	(0.0271)	(0.0171)
Observations	10,348,597	4,487,372	5,861,225	1,314,394	3,172,978	1,576,958	4,284,267
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Conditional on Price Change	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sequential FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The bilateral exchange rate is defined as rmb per unit of foreign (destination) currency. Standard errors reported in parentheses are clustered at the product-firm-destination level. Statistical significance at the 1, 5 and 10 percent level is indicated by ***, **, and *. The significance of results is robust to different levels of clustering.

[38] Unfortunately, no information on the currency of invoicing is reported in the Chinese customs dataset so we cannot infer if this is true.

[39] In appendix 2.A we present corresponding estimates for each column of table 2.3 using the GR method. As expected, the estimates of ERPT are larger, implying that changes in the price of imported inputs have an impact on the export price denominated in rmb.

Table 2.4 provides a comparison of results from the GR methodology alongside those from the CHS estimator broken down by (1) the type of contractual arrangement under which the goods were produced in the top panel and (2) the product's Rauch classification in the bottom. These results document that the CHS estimator is effective at capturing the price markup adjustment. In the top panel of table 2.4, specifications (1) - (6), present results for subsamples by different official custom's classifications of trade mode. These modes refer to the type of contractual arrangement between input suppliers and manufacturers that determine how imported inputs are taxed at the border. Among these classifications, we use three: (i) general trade (GT), (ii) assembling supplied materials (ASM), and (iii) processing imported materials (PIM). Together, these three modes account for over 90% of transaction records in our sample over 2000-2006.[40] An export transaction is classified as GT if the producing firm receives no special tariff treatment for any imported inputs and is free to sell its output domestically or abroad. In other words, production takes place without any special contract over pricing. We expect GT export prices to respond optimally to bilateral exchange rate movements. Export transactions classified as ASM imply the manufacturing firm produces the output under a contract with a pre-determined price and receives a fee from the components supplier for its services. Therefore, the price should be insensitive to bilateral exchange rate shocks and changes in destination CPI. Finally, export transactions under which firms process imported materials (PIM) involve an obligation for the firm to export the merchandise produced from imported inputs that receive duty-free tariff treatment. Because these firms purchase imported inputs, their costs can fluctuate with exchange rate movements. However, because they are processing firms which provide little value-added, they have little scope to engage in pricing to market or markup adjustments.

In columns (1) and (4), we present estimates of price responsiveness to bilateral exchange rate movements under the GR and CHS methods. The finding that the CHS estimate of 0.075 is strictly smaller than the GR estimate of 0.099 is consistent with our claim that the GR method incorporates movements due to input price fluctuations and markup adjustments whereas the CHS method isolates the markup adjustment. More persuasively, we show in columns (2) and (3) that, using the GR method, export prices of ASM and PIM producers fall slightly as the rmb appreciates, consistent with what we would expect for producers reliant on imported inputs. Interestingly, when we apply the CHS estimator that differences out the product-level time-varying marginal cost of ASM and PIM producers, there is no price adjustment. These results serve to validate that the CHS estimator is acting in the way we expected—finding markup adjustments when they are feasible and not finding them for contractual arrangements that make them implausible.

We further validate the CHS estimator by comparing it to the GR method for products grouped under the Rauch classification. Under this classification, products traded on open

[40]These detailed trade modes are only reported for transactions until 2006.

exchanges (OE) are generally regarded as commodities whose prices are expected to fluctuate with global supply and demand. Reference price (RP) products are list price goods which compete somewhat directly by offering goods for sale at a published price reported in an industry trade publication. The producers of these goods are thought to have a very limited ablility to exploit market power in pricing. Finally, differentiated goods are characterized by non-public negotiated prices which indicate limited direct competition among firms and greater scope for charging a markup.

Because only differentiated goods should exhibit a mark-up adjustment, we expect that the CHS estimator would only yield a statistically significant elasticity for differentiated goods. This is exactly what we observe in specifications (10) through (12). In contrast, the GR specifications (7) - (9) indicate that export prices of all types of goods move in response to bilateral exchange rate movements. Taken together, these results allow us to decompose these price movements into markup adjustments and marginal cost movements. For OE and RP products, all of the substantial changes in the export prices appear to be due to changes in the cost of imported inputs. In contrast, for differentiated goods, almost 40% of the total exchange rate pass through estimated by the GR method is due to price markup adjustments that vary across destinations.

Table 2.4 Heterogeneity in price markup elasticities by contractual arrangement and product type

	GR	GR	GR	CHS	CHS	CHS
	(1)	(2)	(3)	(4)	(5)	(6)
Contractual Arrangement	GT	ASM	PIM	GT	ASM	PIM
Bilateral nominal exchange rate	0.0986***	0.0547*	0.0731***	0.0750***	-0.0439	0.0178
	(0.00547)	(0.0320)	(0.0180)	(0.00723)	(0.0603)	(0.0364)
Destination CPI	0.0853***	0.0759	0.0314	0.0696***	-0.0206	-0.0672
	(0.00857)	(0.0490)	(0.0262)	(0.00893)	(0.105)	(0.0450)
Observations	2,811,221	76,684	134,084	6,377,126	159,403	311,490
	(7)	(8)	(9)	(10)	(11)	(12)
Rauch Classification	Differentiated	RP	OE	Differentiated	RP	OE
Bilateral nominal exchange rates	0.128***	0.348***	0.262***	0.0482***	0.0453	-0.0316
	(0.00579)	(0.0169)	(0.0475)	(0.0157)	(0.0343)	(0.109)
Destination CPI	0.118***	0.422***	0.419***	0.0282	0.0573	0.00465
	(0.0109)	(0.0251)	(0.0735)	(0.0178)	(0.0366)	(0.133)
Observations	2,178,012	211,811	16,197	4,661,095	458,392	31,957
S-Period Difference	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes
Conditional on Price Change	Yes	Yes	Yes	Yes	Yes	Yes
Sequential FE	No	No	No	Yes	Yes	Yes

Notes: CPI, real GDP and M/GDP are included in each regression, but estimates are not reported. The bilateral exchange rate is defined as rmb per unit of foreign (destination) currency. Standard errors are clustered at the product-firm-destination level reported in parentheses. Statistical significance at the 1, 5 and 10 percent level is indicated by ***, **, and *. The significance of results is robust to different levels of clustering.

Previous research (Manova and Zhang (2012)) has suggested that price differences across destination markets can reflect differences in the quality of the product exported by a firm to different destinations. Because our estimator is based on changes in destination-specific markups that move with the exchange rate and not the time-invariant level of the markup (which would reflect a quality differential), it should be robust to small differences in the level of product quality across destinations. In table 2.5, we address the concern that our estimator might simply be capturing the differences in quality by evaluate price markup adjustments to bilateral exchange rates for a small set of high-income economies which we expect have similar preferences for quality. Therefore, we believe it is reasonable to assume that all destinations in this set receive goods of the same quality from individual Chinese firms. Table 2.5 presents the markup elasticities for Chinese firms that export to Australia, Canada, France, Germany, Japan, the Netherlands and the United Kingdom. In this restricted sample of exports to a set of high-income countries with similar GDP per-capita, we find slightly higher markup elasticities than our baseline results in table 2.2. We take these results as confirming that our estimator captures the responsiveness of markups to bilateral exchange rate movements rather than time-invariant cross-sectional differences in prices.

Table 2.5 Markup elasticities for exports to high-income countries

	(1) Full Sample	(2) Trading	(3) Non-trading	(4) Year ≤ 2005	(5) Year > 2005
Bilateral nominal exchange rates	0.0796***	0.135***	0.0363*	-0.0731	0.0929***
	(0.0165)	(0.0255)	(0.0215)	(0.0609)	(0.0339)
Destination CPI	-0.00878	0.0323	-0.0470	0.234	-0.126
	(0.0495)	(0.0765)	(0.0649)	(0.189)	(0.107)
Destination Real GDP	0.175***	0.260***	0.103	0.274**	-0.0207
	(0.0497)	(0.0761)	(0.0655)	(0.121)	(0.0819)
Import-to-GDP ratio	0.0610***	0.109***	0.0240	0.0513	0.114***
	(0.0184)	(0.0292)	(0.0236)	(0.0554)	(0.0368)
Observations	470,798	210,827	259,971	156,468	314,330
Clustered SE	Yes	Yes	Yes	Yes	Yes
Conditional on Price Change	Yes	Yes	Yes	Yes	Yes
Sequential FE	Yes	Yes	Yes	Yes	Yes

Notes: The bilateral exchange rate is defined as rmb per unit of foreign (destination) currency. Standard errors are clustered at the product-firm-destination level reported in parentheses. Statistical significance at the 1, 5 and 10 percent level is indicated by ***, **, and *. The significance of results is robust to different levels of clustering.

2.7 Evidence for the International Pricing System

Having shown that we can estimate a destination-specific price markup elasticity in response to bilateral exchange rate movements and destination CPI growth, we turn to the question: what effect do third country currency movements have on the prices set by Chinese exporters? We construct two measures of multilateral exchange rate movements and explore the effect of the exchange rate movements between third country trading partners ($d_2, d_3, ..., d_N$) and the destination country, d_1. We expect that an appreciation of the destination d_1 currency relative to its trading partners $d_2, d_3, ..., d_N$ raises the competitive pressure on Chinese firms exporting to d_1 and should lead them to reduce their prices. Our analysis of multilateral currency movements uses the global currency invoicing matrix[41] provided by Gopinath (2015) and examines how the competition effect changes with the proportion of dollar invoicing in a destination country.[42] Our results provide evidence that multilateral exchange rate movements affect prices from other exporters; further, variation in this effect

[41] A shortcoming is the absence of invoicing data for Chinese exporters. Our analysis provides a new perspective on how Chinese exporters vary their pricing according to the dollar-invoicing competitive environment.

[42] Gopinath (2015) proposed the idea of an international price system characterized by two features."First, the overwhelming share of world trade is invoiced in very few currencies, with the dollar being the dominant currency. Second, international prices, in their currency of invoicing, are not very sensitive to exchange rate

by the proportion of the transactions invoiced in dollars supports Gopinath's hypothesis on the existence of an international price system.

2.7.1 Construction of a multilateral exchange rate measure orthogonal to the rmb

The nominal exchange rate between China and destination d_1 is correlated with the bilateral exchange rates between d_1 and its trading partners $d_2, d_3, ..., d_N$. Thus, to conduct an analysis of bilateral (rmb/d_1) and multilateral ($d_2/d_1, d_3/d_1, ..., d_n/d_1$) exchange rate movements, we must first extract the component of third country exchange rate movements that are orthogonal to the rmb/d_1 rate. We start from the measure of the nominal effective exchange rate (NEER) of country d_1 which is constructed as[43]

$$log(neer_{d_1 t}) = w_{rmb,t} log(e_{\frac{CHN}{d_1},t}) + w_{d_2,t} log(e_{\frac{d_2}{d_1},t}) + ... + w_{d_N,t} log(e_{\frac{d_N}{d_1},t}) \tag{2.30}$$

where $w_{d_2,t}$ represents the trade weight of d_2 in country d_1. Firstly, we fit the following OLS regression for each destination d and extract $\widehat{u_{dt}}$.

$$\Delta log(neer_{dt}) = \widehat{a_d} + \widehat{b_d} \Delta[w_{d,rmb,t} log(e_{\frac{CHN}{d},t})] + \widehat{u_{dt}} \tag{2.31}$$

If changes in the bilateral exchange rate between the destination country vis-à-vis China and the destination country vis-à-vis its other trading partners are orthogonal[44], then the residual from (2.31) is the trade-weighted average of third country exchange rates.

$$\widehat{u_{dt}} = \Delta[w_{d,d_2,t} log(e_{\frac{d_2}{d_1},t})] + ... + \Delta[w_{d,d_N,t} log(e_{\frac{d_N}{d_1},t})]$$

If the orthogonality condition does not hold, then the regression in (2.31) will separate out the common components from $\widehat{u_{dt}}$. For example, if $\Delta[w_{d,d_2,t} log(e_{\frac{d_2}{d_1},t})] = \Delta[w_{d,rmb,t} log(e_{\frac{CHN}{d_1},t})]$,

$$\widehat{u_{dt}} = \Delta[w_{d,d_3,t} log(e_{\frac{d_3}{d_1},t})] + ... + \Delta[w_{d,d_N,t} log(e_{\frac{d_N}{d_1},t})].$$

fluctuations at horizons of up to two years." The existence of an international price system would imply a single indicator, the invoicing currency statistics for a country, is a powerful predictor for ERPT into import prices.

[43] We take the nominal effective exchange rate index (board) from Bank for International Settlements as they provide details on the trade weight matrix used to construct the NEER for each country.

[44] i.e., $Cov\left[\Delta log(e_{\frac{CHN}{d_1},t}), \Delta log(e_{\frac{d_k}{d_1},t})\right] = 0 \quad \forall d_k \neq rmb$

We construct our main empirical measure of the orthogonal component of the destination's multilateral exchange rate, the orthogonal NEER, as follows:

$$\begin{aligned}
log(oneer_{d,2000}) &= log(100) \\
log(oneer_{dt}) &= log(oneer_{dt-1}) + \widehat{u_{dt}} \quad \forall t = 2001, ..., 2011 \\
&= log(100) + \sum_{\tau=2001}^{t} \widehat{u_{d\tau}} \\
oneer_{dt} &= oneer_{dt} / oneer_{d,2010}
\end{aligned}$$

The last step re-indexes all series to a base year of 2010. It can be easily verified that $\Delta log(oneer_{dt}) = \widehat{u_{dt}}$.

2.7.2 Results on bilateral and multilateral exchange rate movements

Having constructed an orthogonal multilateral exchange rate measure, we estimate an extended GR ERPT regression:

$$\Delta log(P^{rmb}_{CHN,dt}) = \gamma_0 + \gamma_1 \Delta log(e_{\frac{CHN}{d},t}) + \gamma_2 \Delta log(X_{dt}) + \gamma_3 \Delta log(oneer_{dt}) + v_{dt} \tag{2.32}$$

where

$$\gamma_3 = \beta_2 \frac{\partial log(\mathcal{S}_{CHN,dt})}{\partial log(oneer_{dt})} + \beta_4 \frac{\partial log(\mathcal{D}_{CHN,dt})}{\partial log(oneer_{dt})} \tag{2.33}$$

and γ_3 captures the composition of supply and demand effects of bilateral exchange rate movements that are orthogonal to the bilateral exchange rate between China and the destination country.

The sign in front of the supply effect β_2 is expected to be zero or negative. Holding the bilateral exchange rate between China and destination d fixed, an appreciation of *oneer* implies a depreciation of (at least one of) the destination's trade partner countries with respect to China. The change in prices of imported inputs from those countries whose currencies have depreciated against the rmb is less than or equal to zero, depending on the degree of ERPT into imported input prices from that country. After receiving a negative cost shock, Chinese exporters will optimally reduce their export prices denominated in rmb.

The sign in front of the demand effect β_3 is also expected to be zero or negative. Exporters from third countries whose currencies, like China's, appreciated also receive cheaper imported inputs and are likely to reduce their export prices in destination d. Note that the local destination currency price of goods from those countries whose currencies depreciated will fall if some of the bilateral exchange rate movements is passed through to import prices. Therefore, all competitors in destination d optimally choose to charge a lower (or the same) price. The competition effect implies downward pressure on prices and Chinese exporters are likely to react to this competition effect by cutting their prices. Therefore, a non-positive demand effect and a non-positive supply effect together imply a zero or negative γ_3.

Table 2.1 reports the contribution to export prices of both bilateral and the orthogonal component of multilateral exchange rate movements. Notably, table 2.1 shows a significant structural break of ERPT for multilateral exchange rate movements associated with the change in China's exchange rate regime from a peg to a managed float. In columns (3) and (5), prior to 2006, we observe a significant negative coefficient on the orthogonal component of the NEER, implying the existence of a local competition effect on Chinese export prices. Post-2005, estimates in columns (4) and (6) show that this orthogonal component has no statistically significant impact on prices. Our results contribute to the literature by providing the first evidence on this competition effect from the perspective of exporting firms.[45]

Further, in columns (5) and (6), we document that the importance of the dollar invoicing share in destination imports changed after the rmb abandoned its peg to the US dollar. The coefficient on the interaction term between the orthogonal component of the multilateral exchange rate and the share of the destination country's trade invoiced in dollars is positive and significantly different from zero. This is consistent with the findings of Gopinath, Itskhoki, and Rigobon (2010) and Gopinath (2015). A positive interaction term means the competition effect is smaller as the proportion of dollar invoicing increases. Prior to 2006, the fact that rmb was pegged to dollars gives us the unique opportunity to identify this competition effect.[46] Larger proportion of dollar invoicing implies a high proportion of competitors' prices is stable in dollars.

Interestingly, estimates show that the multilateral effect diminished after 2006 when rmb was unpegged to dollar.

[45] Auer and Schoenle (2016) provide empirical evidence on this competition effect from the perspective of importers.

[46] Before 2006, our *oneer* measures destination multilateral exchange rate shocks that are orthogonal to dollar movements.

Table 2.1 ERPT of bilateral and orthogonal multilateral exchange rate shocks (GR)

	(1) Full Sample	(2) Full Sample	(3) ≤ 2005	(4) > 2005	(5) ≤ 2005	(6) > 2005
Bilateral nominal exchange rates	0.125***	0.118***	0.215***	0.192***	0.211***	0.263***
	(0.00462)	(0.00463)	(0.00896)	(0.00611)	(0.0101)	(0.00689)
Orthogonal Destination NEER		0.179***	-0.0652**	-0.00362	-0.360***	0.0426
		(0.00859)	(0.0277)	(0.0102)	(0.0716)	(0.0444)
Dollar Invoicing Share					0.00553	- 0.0358***
					(0.00346)	(0.00188)
Orthogonal Destination NEER $*$ Dollar Invoicing Share					0.599***	-0.0492
					(0.117)	(0.0742)
Destination CPI	0.270***	0.278***	0.0752***	0.443***	0.108***	0.757***
	(0.0108)	(0.0108)	(0.0188)	(0.0150)	(0.0217)	(0.0210)
Destination real GDP	0.0547***	0.0500***	0.318***	- 0.0623***	0.243***	-0.395***
	(0.0114)	(0.0114)	(0.0285)	(0.0133)	(0.0404)	(0.0195)
Import-to-GDP ratio	0.226***	0.243***	0.311***	0.223***	0.334***	0.284***
	(0.00388)	(0.00396)	(0.0134)	(0.00421)	(0.0145)	(0.00514)
Observations	3,773,153	3,773,153	886,387	2,886,766	770,640	2,439,715
S Period Difference	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes
Con Price Change	Yes	Yes	Yes	Yes	Yes	Yes

Notes: An increase in the orthogonal component of destination NEER means destination currency appreciation. The bilateral exchange rate is defined as rmb per unit of foreign (destination) currency. Standard errors reported in parentheses are clustered at the product-firm-destination level. Statistical significance at the 1, 5 and 10 percent level is indicated by ***, **, and *. The significance of results is robust to different levels of clustering.

2.7.3 Demand-side multilateral controls

We now introduce the construction of the destination demeaned orthogonal NEER measure for our method.

$$\widetilde{log(oneer_{dt})} \equiv log(oneer_{dt}) - \frac{1}{N} \sum_{j \neq d, rmb} log(oneer_{jt}) \tag{2.34}$$

The destination demeaned measure can be written as[47]

$$\widetilde{log(oneer_{dt})} - \frac{1}{T} \sum_{t} \widetilde{log(oneer_{dt})} = \vartheta_{dt} - \frac{1}{T} \sum_{\tau=2001}^{2011} \vartheta_{d\tau}$$

[47]See subsection 2.B.2 for derivations.

where

$$\vartheta_{dt} \equiv \sum_{j \neq d,rmb} [(\frac{N-1}{N} w_{d,j,t} + \frac{1}{N} w_{j,d,t}) log(e_{\frac{j}{d},t})] + \frac{1}{N} \sum_{k \neq d,rmb} \sum_{j \neq k} [w_{k,j,t} log(e_{\frac{k}{j},t})] \tag{2.35}$$

ϑ_{dt} can be decomposed into two elements. The first element represents changes of exchange rates and trade weights that are relevant to the destination d. The second element represents the bilateral exchange rate movements among countries excluding China and destination d.

The first element can be further decomposed into changes of trade weights and destination d currency movements. Consider the case of a general appreciation of destination d's currency against all of its trading partners. If trade weights $w_{d,j}$ are relatively stable, the first part of ϑ_{dt} will be proportional to changes of bilateral exchange rates between China and the destination country, i.e.

$$d\vartheta_{dt} = \sum_{j \neq d,rmb} [(\frac{N-1}{N} w_{d,j} + \frac{1}{N} w_{j,d}) dlog(e_{\frac{j}{d},t})] \propto dlog(e_{\frac{CHN}{d},t}) \tag{2.36}$$

Conditional on the bilateral exchange rate movement $e_{\frac{CHN}{d},t}$, variations in ϑ_{dt} are mainly driven by changes of trade weights and additional orthogonal exchange rate movements among trading partners excluding China and destination d. Therefore, in the following specification,

$$log(\widetilde{P^{rmb}_{CHN,dt}}) = \gamma_0 + \gamma_1 log(\widetilde{e_{\frac{CHN}{d},t}}) + \gamma_2 log(\widetilde{X_{dt}}) + \gamma_3 log(\widetilde{oneer_{dt}}) + v_t + v_{dt} \tag{2.37}$$

We interpret γ_1 as the demand effect of destination country exchange rate movements and γ_3 as the demand effect due to changes of trade weights and additional orthogonal exchange rate movements among trading partners excluding China and destination d.

Estimation results are shown in table 2.2. The main conclusion here is that adding additional multilateral controls does not affect our bilateral estimates. This additional measure and its associated interaction with dollar invoicing share are not significantly different from zero in all specifications.

Table 2.2 Markup responsiveness to bilateral and orthogonal multilateral shocks (CHS)

	(1) Full Sample	(2) Full Sample	(3) ≤ 2005	(4) > 2005	(5) ≤ 2005	(6) > 2005
Bilateral nominal exchange rates	0.0645***	0.0658***	0.0128	0.104***	0.0156	0.117***
	(0.00879)	(0.00907)	(0.0246)	(0.0152)	(0.0282)	(0.0170)
Orthogonal Destination NEER		-0.0122	0.0624	0.00135	0.0868	0.0115
		(0.0188)	(0.0597)	(0.0261)	(0.0993)	(0.0599)
Orthogonal Destination NEER $*$ Dollar Invoicing Share				-0.0360	-0.00809	
					(0.161)	(0.106)
Destination CPI	0.0581***	0.0589***	0.0413*	0.0474*	0.0458	0.0561*
	(0.0113)	(0.0114)	(0.0248)	(0.0269)	(0.0279)	(0.0333)
Destination real GDP	0.0218*	0.0220*	0.107**	-0.0203	0.122***	-0.0238
	(0.0128)	(0.0128)	(0.0419)	(0.0236)	(0.0464)	(0.0287)
Import-to-GDP ratio	0.00420	0.00412	0.00997	0.0348***	0.0229	0.0459***
	(0.00761)	(0.00761)	(0.0230)	(0.0132)	(0.0260)	(0.0164)
Observations	8,048,660	8,048,660	2,298,021	5,750,639	1,968,364	4,735,842
S Period Difference	No	No	No	No	No	No
Sequential FE	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes
Con Price Change	Yes	Yes	Yes	Yes	Yes	Yes

Notes: An increase in the orthogonal component of destination NEER means destination currency appreciation. The bilateral exchange rate is defined as rmb per unit of foreign (destination) currency. Standard errors reported in parentheses are clustered at the product-firm-destination level. Statistical significance at the 1, 5 and 10 percent level is indicated by ***, **, and *. The significance of results is robust to different levels of clustering.

2.8 A Canonical Pricing Model: Numerical Simulations of ERPT

In this section, we investigate how changing key parameters of firm's pricing strategy affects the estimated ERPT.

First, we simulate firms based on the productivity distributions estimated from China Industrial Survey Data (1999-2007). Next, we merge the simulated firms with real data of macro series and construct prices of the simulated firms using a pricing equation which incorporates three key ingredients of real rigidities, i.e., markup adjustment due to local competition, the wedge effect of local distribution cost, and the imported input channel[48]. We then apply GR and our own proposed method and record the estimated values and standard errors. We test a wide range of combinations of parameter values and regress the obtained coefficients on parameter values to get the sensitivity measure.

[48] As our estimates are conditional on prices changes, we focus on real rigidities rather than nominal rigidities.

The exporter's optimal price P^{rmb}_{fdt} can be derived as follows:

$$P^{rmb}_{fdt} = \frac{\theta_{fdt}}{\theta_{fdt}-1}\left[mc^{CHN}_{ft} + \frac{\eta P^*_{dt}}{\theta_{fdt}} e_{dt}\right] \tag{2.38}$$

$$mc^{CHN}_{ft} = \frac{Z_t + \gamma_f M_t/(e^{CHN}_t)^{\psi}}{\Omega_{ft}} \tag{2.39}$$

where θ_{fdt} is the firm-destination specific demand elasticity; mc^{CHN}_{ft} is the firm specific marginal cost denominated in producer's currency rmb; η is the local distribution margin as in Corsetti and Pesenti (2005); P^*_{dt} is the destination price level; e_{dt} is the bilateral nominal exchange rate between China and its trading partner, an increase in it means an rmb appreciation; Z_t and M_t stand for the price of local and imported input respectively; γ_f is the ERPT to imported inputs[49,50]; e^{CHN}_t is the nominal effective exchange rate of rmb[51], an increase of which means an rmb appreciation; ψ is the pass through for imported inputs; Ω_{ft} is the firm specific productivity shock.

To keep our analysis as tractable as possible, we make the following two simplifications. (a) the logarithm of the elasticity of demand is a linear process; (b) the firm's productivity follows an $AR(1)$ process.

$$log(\theta_{fdt} - \theta^0) = \theta_d - \beta_{mc} log(mc^{CHN}_{ft}) - \beta_e log(e_{dt}) - \beta_{oneer} log(oneer_{dt}) - \beta_p log(P^*_{dt}) + v_{fdt} \tag{2.40}$$

where θ^0 gives the average demand elasticity across all firms, destinations and time periods; θ_d represents heterogeneous competition environments in the destination markets; β_{mc} represents the effect of productivity on the competitiveness of a firm subject to a certain market structure.[52] ; β_e represents the direct impact of bilateral exchange rate movements; β_{oneer} represents the impact of orthogonal components of nominal effective exchange rate movements at destination d; β_p represents the effect of changing local consumer price level.

Firms productivity shocks are assumed to take a form specified in equation (2.41). We estimate parameters of equation (2.41) and the productivity distribution of Chinese firms using annual data of China Industrial Survey (1999-2007) following the method of Olley and Pakes (1996).

[49] We do not aim to separate the income effect and substitution effect here. γ_f in our model is the composition of these two effects.

[50] We do have the import data for each firm in our datasets. However, we cannot directly control for imported inputs as we cannot identify for most firms if the imported inputs are used for domestic sales or for inputs to produce exported products. Another related problem is that a typical firm both imports and exports multiple products, we do not know which imported input is used in which exported product.

[51] Here we assume that the cost of imported inputs is only affected by general rmb movements but not country specific exchange rate movement.

[52] A positive β_{mc} means that a higher responsiveness of the demand elasticity to marginal cost increases the estimated ERPT. In a Cournot competition framework, the demand elasticity is a function of firm's market share which is in turn a function of firm's marginal cost.

$$log(\Omega_{ft}) = \rho_{\Omega} log(\Omega_{ft-1}) + \xi_{ft}^{\Omega} + \zeta_{t}^{\Omega} \tag{2.41}$$

In the last step, we want to generate missing observations that resemble the pattern we have observed in our dataset. We use minimal demand-side settings to generate the correct weight of trade flows to each country.[53] Under this setup, the destination country will accept the price and purchase the product if the price converted into local currency is smaller than the cutoff value of a local CPI indexed relative price. The price can be dropped from our dataset for two reasons: (a) the potential price charged by the exporter is higher than the maximum price the destination market could accept; (b) the exporter is subject to a certain type of nominal rigidity and cannot change their price. We do not try to distinguish between these two effects. δ_{dt} is calibrated to match the observed trade weight (in terms of the number of observations) in each time period.

$$P_{fdt}^{rmb} e_{dt} < \delta_{dt} * P_{dt}^{*} \tag{2.42}$$

We simulate our model to examine all possible combinations of parameter values in table 2.1. Our simulation procedure is as follows.

1. Start with productivity data of 500 sampled firms and build series of Ω_{ft} from 2000 to 2011 by adding firm specific shocks ξ_{ft}^{Ω} randomly drawn from a standard normal distribution and aggregate shocks ζ_{t}^{Ω} estimated from our macro series according to equation (2.41).

2. Merge simulated firms with the dataset of macro series. Construct the demand elasticity θ_{fdt} using equation (2.40) with the local preference parameter θ_d randomly drawn from a standard normal distribution. Construct marginal cost mc_{ft}^{CHN} with Z proxied by the average nominal wage in China and M normalised to 100 using equation (2.39). Construct the price P_{fdt}^{rmb} using equation (2.38).

3. For each time period, compute the percentile of the price P_{fdt}^{rmb} charged by firm f in destination d and drop the price if equation (2.42) is not satisfied.

4. Collect the simulated data, apply and record the value and standard errors of estimates using GR and our method.

5. Repeat 1-4 for all possible combinations of parameter calibrations in table 2.1.

6. For each variable of interest, regress values of the estimated coefficient on values of calibrated parameters weighted by the inverse of the standard errors of the estimator.

[53] Note that our main objective is to understand the pricing behaviour. We do not aim to precisely match the extensive margin. The technical reason for our setup is to obtain a realistic weight of exchange rate shocks that resembles the weight implied in our empirical estimation.

Table 2.1 Calibration

Parameter	Testing Range
β_e	0, 0.25, 0.5, 0.75, 1
β_{oneer}	0, 0.25, 0.5, 0.75, 1
β_{P^*}	0, 0.25, 0.5, 0.75, 1
ψ	0, 0.25, 0.5, 0.75, 1
η	0, 0.2, 0.5
γ	0, 0.5, 1
θ^0	2.5

Table 2.2 summarises our simulation results. Each row represents the sensitivity of the estimated coefficients to the change of calibrated parameters holding other calibrated parameters fixed, i.e.,

$$\frac{\partial \text{ estimated coefficient of variable } x \text{ using method } m}{\partial \text{ calibration of parameter } para}$$

where $x \in \{\text{Bilateral NER, ONEER, CPI}\}$, $m \in \{\text{GR, CHS}\}$, $para \in \{\beta_e, \beta_{oneer}, \beta_{P^*}, \psi, \eta, \gamma\}$.

The last row ("constant") of each method represents the estimates when all PTM parameters are set to zero. In this case, simulated firms will charge a constant markup and only change their price denominated in rmb due to marginal cost shocks. The estimated EPRT of the GR method is high and significant (0.169), reflecting the correlation between movements of firms' nominal marginal cost and exchanges rates. As we expected, the estimated coefficients of Bilateral NER, ONEER and CPI show strong responses to the change of calibrated values of $\beta_e, \beta_{oneer}, \beta_{P^*}$ respectively for the GR panel. The import share γ plays a non-trivial role in ERPT and CPI coefficients using the GR method. A higher import share reduces ERPT and a higher import pass through strengthens this effect. The change in the calibration of import pass through ψ is quantitatively small.

After differencing out firm level cost shocks, the row "constant" suggests that estimates from our method give the correct markup elasticity, i.e., the estimated coefficients are not different from zero when all PTM parameters are set to zero. The responsiveness of estimated coefficients of Bilateral NER, ONEER and CPI to diagonal demand side parameters $\beta_e, \beta_{oneer}, \beta_{P^*}$ is highly significant and similar to that obtained in the GR method.[54] Changing values of supply-side parameters γ, ψ has no effect on the estimated coefficients using our method, confirming its capability of controlling for firm-specific cost-side confounding factors.

[54]For some numerical reasons, the off-diagonal coefficients are also significantly different from zero. These values are small in magnitude though.

Table 2.2 Sensitivity tests

	GR			CHS		
	Bilateral NER	Orthogonal NEER	CPI	Bilateral NER	Orthogonal NEER	CPI
β_e	1.000***	-4.66e-06	2.93e-06	0.998***	-0.000979**	-0.000883***
	(4.79e-06)	(5.62e-06)	(7.63e-06)	(0.000346)	(0.000391)	(0.000315)
β_{oneer}	-7.45e-07	1.000***	1.92e-07	-0.000984***	0.997***	-0.00106***
	(4.16e-06)	(6.17e-06)	(7.70e-06)	(0.000308)	(0.000539)	(0.000332)
β_{P^*}	-1.63e-06	9.63e-06*	1.000***	-0.000832***	-0.000918**	0.998***
	(4.12e-06)	(5.62e-06)	(8.14e-06)	(0.000294)	(0.000389)	(0.000379)
η	1.943***	-3.045***	2.363***	2.004***	0.00504***	2.004***
	(0.0573)	(0.0512)	(0.123)	(0.000865)	(0.00119)	(0.000939)
γ	-0.140***	0.0553	-0.514***	5.39e-05	5.14e-05	4.86e-05
	(0.0374)	(0.0359)	(0.0671)	(0.000239)	(0.000349)	(0.000258)
ψ	3.72e-05***	-7.14e-05***	9.62e-05***	-2.56e-05	-5.63e-05	-7.84e-06
	(4.38e-06)	(6.32e-06)	(8.24e-06)	(0.000213)	(0.000307)	(0.000230)
Constant	0.169***	1.467***	0.332***	-0.000293	-0.000401	-0.000321
	(0.0240)	(0.0251)	(0.0263)	(0.000209)	(0.000296)	(0.000226)

We find a large quantitative effect of the change in distribution margins on pass through estimates using both GR and our method. As pointed out by Corsetti and Dedola (2005), the existence of local distribution margin drives a wedge between local consumer prices and the dock price, resulting in an incomplete ERPT not only for the consumer price but also the price at the dock. Our simulation results imply that a higher local distribution margin increases the responsiveness of prices denominated in rmb and lowers the pass through to prices denominated in the local currency.

Table 2.3 presents the results when estimations are run on two sub-samples, the pre and post dollar peg period. From the upper panel, GR results show that estimated average ERPT (the "constant" row) and the sensitivity of changing parameters ψ and η change significantly, providing evidence for a structural break when rmb was unpegged from dollar. However, our estimates suggest a stable demand side relationship over time.

Table 2.3 Sensitivity tests (pre and post dollar peg era)

	≤ 2005			> 2005		
	Bilateral NER	Orthogonal NEER	CPI	Bilateral NER	Orthogonal NEER	CPI
GR						
β_e	1.000***	-1.16e-07	6.61e-06	1.000***	-3.42e-06	6.14e-07
	(5.70e-06)	(9.18e-06)	(8.06e-06)	(4.44e-06)	(5.10e-06)	(9.50e-06)
β_{oneer}	-5.81e-07	1.000***	-5.48e-07	-9.28e-07	1.000***	7.02e-07
	(4.96e-06)	(9.73e-06)	(8.14e-06)	(4.01e-06)	(5.47e-06)	(9.60e-06)
β_{P^*}	4.67e-06	-1.70e-05*	1.000***	-2.33e-06	6.01e-06	1.000***
	(4.90e-06)	(9.10e-06)	(8.79e-06)	(3.98e-06)	(5.08e-06)	(9.82e-06)
η	-2.684***	10.96***	5.197***	2.343***	-1.750***	3.780***
	(0.0625)	(0.0982)	(0.0655)	(0.0581)	(0.0763)	(0.169)
γ	-0.154***	0.342***	-0.173***	-0.157***	0.273***	-0.741***
	(0.0397)	(0.0552)	(0.0431)	(0.0374)	(0.0454)	(0.0882)
ψ	-3.59e-05***	0.000117***	0.000102***	4.81e-05***	-6.84e-05***	0.000142***
	(6.66e-06)	(1.41e-05)	(9.09e-06)	(3.82e-06)	(4.85e-06)	(9.14e-06)
Constant	2.496***	-5.822***	-1.425***	-0.0148	0.603***	-0.149***
	(0.0245)	(0.0252)	(0.0280)	(0.0236)	(0.0246)	(0.0257)
CHS						
β_e	0.998***	-0.00138**	-0.00120**	0.998***	-0.00101***	-0.00101***
	(0.000414)	(0.000609)	(0.000526)	(0.000333)	(0.000331)	(0.000317)
β_{oneer}	-0.00119***	0.996***	-0.00185***	-0.000961***	0.998***	-0.001000***
	(0.000373)	(0.000846)	(0.000582)	(0.000291)	(0.000415)	(0.000318)
β_{P^*}	-0.00106***	-0.00173***	0.997***	-0.00101***	-0.00106***	0.998***
	(0.000360)	(0.000635)	(0.000674)	(0.000295)	(0.000336)	(0.000385)
η	2.004***	0.00783***	2.007***	2.004***	0.00526***	2.005***
	(0.00102)	(0.00183)	(0.00156)	(0.00104)	(0.00124)	(0.00117)
γ	7.51e-05	1.00e-04	0.000100	8.68e-05	9.65e-05	7.05e-05
	(0.000237)	(0.000460)	(0.000379)	(0.000493)	(0.000597)	(0.000553)
ψ	-2.00e-05	-7.68e-05	-1.77e-05	-2.08e-05	-2.34e-05	-1.53e-05
	(0.000224)	(0.000431)	(0.000356)	(0.000292)	(0.000349)	(0.000326)
Constant	-0.000303	-0.000544	-0.000479	-0.000610	-0.000733	-0.000674
	(0.000216)	(0.000403)	(0.000343)	(0.000399)	(0.000480)	(0.000440)

2.9 Conclusions

This paper studies how China's exporters react to exchange rate shocks by employing firm-level data from China's custom office from 2000 to 2011. Using multi-destination exporters, we carefully decompose the effect of exchange rate shocks into demand and supply components.

We propose a sequential fixed effect estimator using orthogonal dimensions in a multi-dimensional panel to control for unobserved firm specific supply shocks. We demonstrate that if the dataset is unbalanced, the order for including various fixed effects matters. Different orders of inclusion eliminate different potential interaction terms between unobserved factors and observed explanatory variables. There may be a unique order that yields the best estimate for the question of interest.

We document four empirical findings on ERPT concerning China's exporters. First, we find an unconditional general ERPT estimate of 0.2, which suggests that Chinese exporters pass through 80% of exchange rate shocks to their trading partners. Second, adding time fixed effects give an unconditional pass through of bilateral exchange rate shocks of 0.14. Third, differencing out destination invariant factors leaves an estimate of country specific demand driven ERPT of 0.06, meaning that about one-half of the total pass-through is due to PTM or demand-side factors, which provides evidence for pricing-to-market. Fourth, the responsiveness to bilateral exchange rate changes is larger for trading firms vs. manufacturers; the flexible exchange rate regime vs. pegged exchange rate regime; general trade firms vs. processing and assembly firms; differentiated goods vs. reference price or open exchange goods.

Chinese export prices exhibit responsiveness to bilateral as well as global (US dollar) currency movements. We find a significant additional dollar effect on the pass-through of exchange rate and local price shocks. The dollar effect is sensitive to the proportion of dollar invoicing in the destination market, adding evidence for the existence of the international price system.

We also simulate a numerical model integrating data of real macro series and experiment on a wide range of combinations of parameters on the pricing equation of exporters. We confirm the importance of firm's imported input share, ERPT of imported inputs and the local distribution margin in ERPT estimation and reaffirm that our estimator is robust to changes in specifications of firm-specific cost factors.

Appendix

Appendix 2.A Additional Results on ERPT Heterogeneity

Table 2.A.1 presents estimates of ERPT for trading and non-trading (manufacturing) export prices using the GR method. The estimated ERPT is higher for non-trading exporters compared to trading exporters. Further, the effect of changing to a floating exchange rate regime is more noticeable for non-trading exporters.

Table 2.A.1 Trading v.s. non-trading exporters (GR)

	(1) Trading	(2) Non-trading	(3) Trading ≤ 2005	(4) Trading > 2005	(5) Non-trading ≤ 2005	(6) Non-trading > 2005
Bilateral nominal exchange rates	0.111***	0.162***	0.0836***	0.120***	0.0845***	0.187***
	(0.00799)	(0.00699)	(0.0148)	(0.00953)	(0.0127)	(0.00847)
Destination CPI	0.105***	0.0883***	0.0114	0.152***	0.0410**	0.104***
	(0.0123)	(0.0103)	(0.0208)	(0.0155)	(0.0172)	(0.0129)
Destination real GDP	0.158***	0.0776***	0.0555	0.151***	0.146***	0.0639***
	(0.0138)	(0.0111)	(0.0370)	(0.0152)	(0.0291)	(0.0122)
Import-to-GDP ratio	0.0555***	0.114***	0.0108	0.0652***	0.0700***	0.122***
	(0.00627)	(0.00516)	(0.0142)	(0.00706)	(0.0121)	(0.00574)
Constant	0.0400***	0.0477***	0.0127***	0.0484***	0.0198***	0.0559***
	(0.000767)	(0.000622)	(0.00197)	(0.000888)	(0.00156)	(0.000726)
Observations	1,375,019	1,991,442	307,601	1,067,418	433,754	1,557,688
S Period Difference	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effect	Year	Year	Year	Year	Year	Year
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes
Conditional on a Price Change	YES	YES	YES	YES	YES	YES

Notes: The bilateral exchange rate is defined as rmb per unit of foreign (destination) currency. Standard errors reported in parentheses are clustered at the product-firm-destination level. Statistical significance at the 1, 5 and 10 percent level is indicated by ***, **, and *. The significance of results is robust to different levels of clustering.

Table 2.A.2 presents a breakdown of ERPT estimates under the GR versus CHS method by Rauch classification and product type. As expected, the CHS estimator documents substantial adjustments to the export price markup in rmb for differentiated goods exported by both trading and non-trading firms. Further, the markup adjustment is larger for trading

firms. There is no markup adjustment for reference price goods using this estimator. The GR estimates of total ERPT show larger export price adjustments for reference price goods than for differentiated goods, indicating that export prices respond to changes in the cost of imported inputs.

Table 2.A.2 Rauch classification for trading vs. non-trading exporters

	Differentiated Products		Reference Priced	
	Bilateral NER	CPI	Bilateral NER	CPI
GR				
Trading	0.11***	0.11***	0.38***	0.42***
	(0.01)	(0.02)	(0.04)	(0.06)
Non-trading	0.16***	0.13***	0.36***	0.46***
	(0.01)	(0.01)	(0.02)	(0.03)
CHS				
Trading	0.08**	0.05	-0.00	0.10
	(0.03)	(0.04)	(0.08)	(0.09)
Non-trading	0.05*	0.02	0.05	0.07
	(0.02)	(0.03)	(0.05)	(0.05)

Notes: The bilateral exchange rate is defined as rmb per unit of foreign (destination) currency. Standard errors are clustered at the product-firm-destination level reported in parentheses. Statistical significance at the 1, 5 and 10 percent level indicated by ***, **, and *. The significance of results is robust to different levels of clustering.

Appendix 2.B Analytical Derivations

2.B.1 Multilateral results of our estimator

$$log(\widetilde{\mathcal{D}_{CHN,d_1t}}) \equiv log(\mathcal{D}_{CHN,d_1t}) - \frac{1}{N}\sum_{d\neq CHN}\frac{\beta_{3,d}}{\beta_{3,d_1}}log(\mathcal{D}_{CHN,dt}) \tag{2.43}$$

$$log(\widetilde{\mathcal{D}_{CHN,-d_1t}}) \equiv log(\mathcal{D}_{CHN,-d_1t}) - \frac{1}{N}\sum_{d\neq CHN}\frac{\beta_{4,d}}{\beta_{4,d_1}}log(\mathcal{D}_{CHN,-dt}) \tag{2.44}$$

Unlike our results of the supply side, to see which part has been differenced out, we need to make two additional simplification assumptions on the functional form of $g_{\mathcal{D},-d_1}(.)$. First, we assume that the demand effect is log-linearly separable by countries, i.e.,

$$\begin{aligned}log(\mathcal{D}_{CHN,-d_1t}) &= log\left[g_{\mathcal{D},-d_1}(e_{\frac{d_2}{d_1},t},...,e_{\frac{d_N}{d_1},t},X_{d_2t},...,X_{d_Nt})\right]\\ &= w_{d_2,d_1}log\left[h_{\mathcal{D},d_2}(e_{\frac{d_2}{d_1},t},X_{d_2t})\right]+...+w_{d_N,d_1}log\left[h_{\mathcal{D},d_N}(e_{\frac{d_N}{d_1},t},X_{d_Nt})\right]\end{aligned}$$

Second, we assume the functional form $h_{\mathcal{D},d_1}(.)$ does not differ across countries, i.e., $h_{\mathcal{D},d_1}(.) = h_{\mathcal{D},d}(.)\ \forall d \neq CHN$. That is, the competition effect on Chinese exporters from country j at destination d only depends on changes of exchange rates $e_{\frac{j}{d},t}$, economics fundamentals $X_{j,t}$ and a destination specific weight $w_{j,d}$.

$$log(\mathcal{D}_{CHN,-d_1t}) = \sum_{d\neq CHN,d_1} w_{d,d_1}log\left[h_{\mathcal{D}}(e_{\frac{d}{d_1},t},X_{dt})\right]$$

Equation (2.44) can be rewritten as

$$\begin{aligned}log(\widetilde{\mathcal{D}_{CHN,-d_1t}}) &= \sum_{d\neq CHN,d_1} w_{d,d_1}log\left[h_{\mathcal{D}}(e_{\frac{d}{d_1},t},X_{dt})\right] - \frac{1}{N}\sum_{d\neq CHN,d_1}\sum_{j\neq CHN,d}\frac{\beta_{4,d}}{\beta_{4,d_1}}w_{j,d}log\left[h_{\mathcal{D}}(e_{\frac{j}{d},t},X_{jt})\right]\\ &= \sum_{d\neq CHN,d_1}\left\{w_{d,d_1}log\left[h_{\mathcal{D}}(e_{\frac{d}{d_1},t},X_{dt})\right] - \frac{1}{N}\frac{\beta_{4,d}}{\beta_{4,d_1}}\sum_{j\neq CHN,d}w_{j,d}log\left[h_{\mathcal{D}}(e_{\frac{j}{d},t},X_{jt})\right]\right\}\end{aligned} \tag{2.45}$$

$$\frac{dlog(\widetilde{\mathcal{D}_{CHN,-d_1t}})}{dlog(e_{\frac{d_k}{d_1},t})} = w_{d_k,d_1}\frac{dlog\left[h_{\mathcal{D}}(e_{\frac{d_k}{d_1},t},X_{d_1t})\right]}{dlog(e_{\frac{d_k}{d_1},t})} - \frac{1}{N}\frac{\beta_{4,d_k}}{\beta_{4,d_1}}w_{d_1,d_k}\frac{dlog\left[h_{\mathcal{D}}(e_{\frac{d_1}{d_k},t},X_{d_kt})\right]}{dlog(e_{\frac{d_1}{d_k},t})} \tag{2.46}$$

$$\frac{dlog(\widetilde{\mathcal{D}_{CHN,d_1t}})}{dlog(e_{\frac{d_k}{d_1},t})} = -\frac{1}{N}\frac{\beta_{3,d_k}}{\beta_{3,d_1}}\frac{dlog\left[g_{\mathcal{D},d_k}(e_{\frac{CHN}{d_k},t},X_{d_kt})\right]}{dlog(e_{\frac{CHN}{d_k},t})} \tag{2.47}$$

2.B.2 Derivation of the empirical destination demeaned orthogonal NEER

Adding destination fixed effects gives

$$
\begin{aligned}
&\widetilde{log(oneer_{dt})} - \frac{1}{T}\sum_t \widetilde{log(oneer_{dt})} = \\
&\qquad log(oneer_{dt}) - \frac{1}{T}\sum_t log(oneer_{dt}) - \frac{1}{N}\sum_{j\neq CHN} log(oneer_{jt}) - \frac{1}{NT}\sum_{j\neq CHN}\sum_t log(oneer_{jt})
\end{aligned}
$$

$$
\begin{aligned}
log(oneer_{dt}) - \frac{1}{T}\sum_t log(oneer_{dt}) &= \sum_{\tau=2001}^{t} \widehat{u_{d\tau}} - \frac{1}{T}\sum_{k=2001}^{2011}\sum_{\tau=2001}^{k} \widehat{u_{d\tau}} \\
&= \sum_{j\neq d,CHN} [w_{d,j,t} log(e_{\frac{j}{d},t})] - \frac{1}{T}\sum_{\tau=2001}^{2011}\sum_{j\neq d,CHN} [w_{d,j,\tau} log(e_{\frac{j}{d},\tau})]
\end{aligned}
$$

2.B.3 Consequences of different partition orders

A different order of partition integrates the time components of the exchange rate e_{dt} and the unobserved variable M_{ft}. However, in a balanced panel, the change of partition order does not matter even if there exists interaction terms between unobserved variables and the variable of interest. As the demeaned time components are the same for all firm and destinations, simply adding the second firm-time fixed effect (demeaned at destination level) will separate destination varying components and firm-time varying components.

$$
\begin{aligned}
\widetilde{p_{fdt}} &= \mathcal{I}_2\widetilde{v_{1,t}} + \widetilde{e_{dt}} + \mathcal{I}_4(\widetilde{e_{dt} * M_{ft}}) + \widetilde{u_{fdt}} \\
\widetilde{v_{1,t}} &= v_{1,t} - \frac{\sum_t v_{1,t}}{n_T} \\
\widetilde{e_{dt}} &= e_{dt} - \frac{\sum_t e_{dt}}{n_T} = \widetilde{v_{1,t}} + \widetilde{v_{2,t}} + v_d^3 * \widetilde{v_{3,t}} + v_d^4 * \widetilde{v_{4,t}}
\end{aligned}
$$

$$
\begin{aligned}
\widetilde{e_{dt} * M_{ft}} &= (v_{1,d} + v_{1,t} + v_{2,d} + v_{2,t} + v_{3,d} * v_{3,t} + v_{4,d} * v_{4,t})(v_{1,f} + v_{1,t} + v_{2,f} + v_{2,t} + v_{5,f} * v_{5,t} + v_{4,f} * v_{4,t}) \\
&\quad - \sum_t (v_{1,d} + v_{1,t} + v_{2,d} + v_{2,t} + v_{3,d} * v_{3,t} + v_{4,d} * v_{4,t})(v_{1,f} + v_{1,t} + v_{2,f} + v_{2,t} + v_{5,f} * v_{5,t} + v_{4,f} * v_{4,t})/n_T \\
&= (v_{1,f} + v_{2,f})\widetilde{e_{dt}} + \widetilde{e_{dt}v_{1,t}} + \widetilde{e_{dt}v_{2,t}} + v_{5,f} * \widetilde{e_{dt}v_{5,t}} + v_{4,f} * \widetilde{e_{dt}v_{4,t}} \\
&= (v_{1,f} + v_{2,f})(\widetilde{v_{1,t}} + \widetilde{v_{2,t}} + v_d^3 * \widetilde{v_{3,t}} + v_d^4 * \widetilde{v_{4,t}}) \\
&\quad + (v_{1,d}\widetilde{v_{1,t}} + \widetilde{v_{1,t}v_{1,t}} + v_{2,d}\widetilde{v_{1,t}} + \widetilde{v_{2,t}v_{1,t}} + v_{3,d}\widetilde{v_{1,t}v_{3,t}} + v_{4,d}\widetilde{v_{1,t}v_{4,t}}) \\
&\quad + (v_{1,d}\widetilde{v_{2,t}} + \widetilde{v_{1,t}v_{2,t}} + v_{2,d}\widetilde{v_{2,t}} + \widetilde{v_{2,t}v_{2,t}} + v_{3,d}\widetilde{v_{2,t}v_{3,t}} + v_{4,d}\widetilde{v_{2,t}v_{4,t}}) \\
&\quad + (v_{1,d}v_{3,d}\widetilde{v_{3,t}} + v_{1,t}v_{3,d}\widetilde{v_{3,t}} + v_{2,d}v_{3,d}\widetilde{v_{3,t}} + \widetilde{v_{2,t}v_{3,t}}v_{3,d} + v_{3,d}v_{3,d}\widetilde{v_{3,t}v_{3,t}} + v_{3,d}v_{4,d}\widetilde{v_{3,t}v_{4,t}}) \\
&\quad + (v_{1,d}v_{4,f}\widetilde{v_{4,t}} + v_{4,f}\widetilde{v_{1,t}v_{4,t}} + v_{2,d}v_{4,f}\widetilde{v_{4,t}} + v_{4,f}\widetilde{v_{2,t}v_{4,t}} + v_{3,d}v_{4,f}\widetilde{v_{3,t}v_{4,t}} + v_{4,d}v_{4,f}\widetilde{v_{4,t}v_{4,t}})
\end{aligned}
$$

If the dataset is unbalanced, the integrated term is firm and destination specific. Adding the second fixed effect no longer separates destination varying components from firm-time varying components. The estimator will in general be biased and the direction of bias is not always clear.

$$
\begin{aligned}
\widetilde{p_{fdt}}^{fd} &= \mathcal{I}_2 \widetilde{v_{1,t}}^{fd} + \widetilde{e_{dt}}^{fd} + \mathcal{I}_4 \widetilde{(e_{dt} * M_{ft})}^{fd} + \widetilde{u_{fdt}}^{fd} \\
\widetilde{v_{1,t}}^{fd} &= v_{1,t} - \frac{\sum_{t\in\tau^{fd}} v_{1,t}}{n_T^{fd}} \\
\widetilde{e_{dt}}^{fd} &= e_{dt} - \frac{\sum_{t\in\tau^{fd}} e_{dt}}{n_T^{fd}} = \widetilde{v_{1,t}}^{fd} + \widetilde{v_{2,t}}^{fd} + v_d^3 * \widetilde{v_{3,t}}^{fd} + v_d^4 * \widetilde{v_{4,t}}^{fd}
\end{aligned}
$$

$$
\begin{aligned}
\widetilde{e_{dt} * M_{ft}}^{fd} &= (v_{1,d} + v_{1,t} + v_{2,d} + v_{2,t} + v_{3,d} * v_{3,t} + v_{4,d} * v_{4,t})(v_{1,f} + v_{1,t} + v_{2,f} + v_{2,t} + v_{5,f} * v_{5,t} + v_{4,f} * v_{4,t}) \\
&- \sum_{t\in\tau^{fd}} (v_{1,d} + v_{1,t} + v_{2,d} + v_{2,t} + v_{3,d} * v_{3,t} + v_{4,d} * v_{4,t})(v_{1,f} + v_{1,t} + v_{2,f} + v_{2,t} + v_{5,f} * v_{5,t} + v_{4,f} * v_{4,t})/n_T^{fd} \\
&= (v_{1,f} + v_{2,f})\widetilde{e_{dt}}^{fd} + \widetilde{e_{dt}v_{1,t}}^{fd} + \widetilde{e_{dt}v_{2,t}}^{fd} + v_{5,f} * \widetilde{e_{dt}v_{5,t}}^{fd} + v_{4,f} * \widetilde{e_{dt}v_{4,t}}^{fd} \\
&= (v_{1,f} + v_{2,f})(\widetilde{v_{1,t}}^{fd} + \widetilde{v_{2,t}}^{fd} + v_d^3 * \widetilde{v_{3,t}}^{fd} + v_d^4 * \widetilde{v_{4,t}}^{fd}) \\
&+ (v_{1,d}\widetilde{v_{1,t}}^{fd} + \widetilde{v_{1,t}v_{1,t}}^{fd} + v_{2,d}\widetilde{v_{1,t}}^{fd} + \widetilde{v_{2,t}v_{1,t}}^{fd} + v_{3,d}\widetilde{v_{1,t}v_{3,t}}^{fd} + v_{4,d}\widetilde{v_{1,t}v_{4,t}}^{fd}) \\
&+ (v_{1,d}\widetilde{v_{2,t}}^{fd} + \widetilde{v_{1,t}v_{2,t}}^{fd} + v_{2,d}\widetilde{v_{2,t}}^{fd} + \widetilde{v_{2,t}v_{2,t}}^{fd} + v_{3,d}\widetilde{v_{2,t}v_{3,t}}^{fd} + v_{4,d}\widetilde{v_{2,t}v_{4,t}}^{fd}) \\
&+ (v_{1,d}v_{3,d}\widetilde{v_{3,t}}^{fd} + v_{1,t}v_{3,d}\widetilde{v_{3,t}}^{fd} + v_{2,d}v_{3,d}\widetilde{v_{3,t}}^{fd} + \widetilde{v_{2,t}v_{3,t}}^{fd}v_{3,d} + v_{3,d}v_{3,d}\widetilde{v_{3,t}v_{3,t}}^{fd} + v_{3,d}v_{4,d}\widetilde{v_{3,t}v_{4,t}}^{fd}) \\
&+ (v_{1,d}v_{4,f}\widetilde{v_{4,t}}^{fd} + v_{4,f}\widetilde{v_{1,t}v_{4,t}}^{fd} + v_{2,d}v_{4,f}\widetilde{v_{4,t}}^{fd} + v_{4,f}\widetilde{v_{2,t}v_{4,t}}^{fd} + v_{3,d}v_{4,f}\widetilde{v_{3,t}v_{4,t}}^{fd} + v_{4,d}v_{4,f}\widetilde{v_{4,t}v_{4,t}}^{fd})
\end{aligned}
$$

Appendix 2.C Descriptive Statistics for China's Customs Dataset

Similar to recent contributions by Berman, Martin, and Mayer (2012) and Amiti, Itskhoki, and Konings (2014), this paper uses annual data on unit values at the HS08 level as the measure for export price. These papers differ in the time periods used for the analysis. See table 2.C.1.

Table 2.C.1 Comparison of the data with other research

	Crowley, Han and Song	Berman, Martin, and Mayer (2012)	Amiti, Itskhoki, and Konings (2014)
Home Country	China	France	Belgium
Frequency	annual	annual	annual
Periods	2000-2011	1995-2005	2000-2008
Price Measure	unit value	unit value	unit value
Import/Export	imports and exports	exports	imports and exports
Level of Disaggregation	8-digit code	8-digit code	8-digit code

China's dramatic ten-fold increase in export value over 2000-2011 includes extensive margin net entry on both the firm and firm-product dimensions. See table 2.C.2.

Table 2.C.2 Chinese exports: exporting firms, products and value, 2000-2011

	Products	Exporters	Product-Exporter Pairs	Obs.	Quantity (billions of units)	Value (billions US$)
2000	6,712	62,770	904,136	1,882,375	506	249
2001	6,722	68,487	991,015	2,121,515	585	291
2002	6,864	76,631	1,140,033	2,396,290	557	299
2003	6,989	93,582	1,429,506	3,051,436	692	421
2004	7,003	118,005	1,771,554	3,792,239	919	569
2005	7,112	140,834	2,182,173	4,689,924	802	717
2006	7,159	184,134	3,152,047	7,330,049	950	895
2007	7,166	193,133	3,285,334	7,311,976	1,164	1,219
2008	7,204	205,995	3,229,604	7,775,485	1,200	1,429
2009	7,316	215,647	3,344,899	8,043,300	1,037	1,199
2010	7,359	233,760	3,827,380	9,686,571	1,225	1,575
2011	7,399	253,893	4,128,888	10,577,429	1,268	2,554

In figures 2.C.1 and 2.C.2 we present the distribution of price changes by the duration with which the firm-destination-product survives and at different frequencies, respectively. For each firm-product-destination triplet in the dataset, we count the number of repeated trade records. k indicates that price observations on the firm-product-destination triplet appeared $k+1$ times in our annual dataset from 2000 to 2011. Hence, firm-product-destinations with a higher number k indicate either the item is traded more frequently or the item survived for more periods. Figure 2.C.1 shows that an item with more records is less spiked with smaller price dispersion. That is, items with more records are less likely to either keep their price fixed between t and $t+1$ or experience abnormally large price changes. In other words, they are changing their prices steadily. In figure 2.C.2 we plot the distribution of price changes conditional on observing a price change for a firm-product-destination period. For each firm-product-destination, we calculate the number of periods, s, between price changes. We then organize price changes into bins of $s = 1, s = 2, ..., s = 8+$. The five panels in figure 2.C.2 display the distribution of price changes given the number of periods between price changes. This figures shows the magnitude of price changes are larger and more dispersed if the number of years between the price changes is larger.

Fig. 2.C.1 Distribution of price changes for firm-product-destination triplets that survive for $1+k$ periods, $k = 1,2,...$

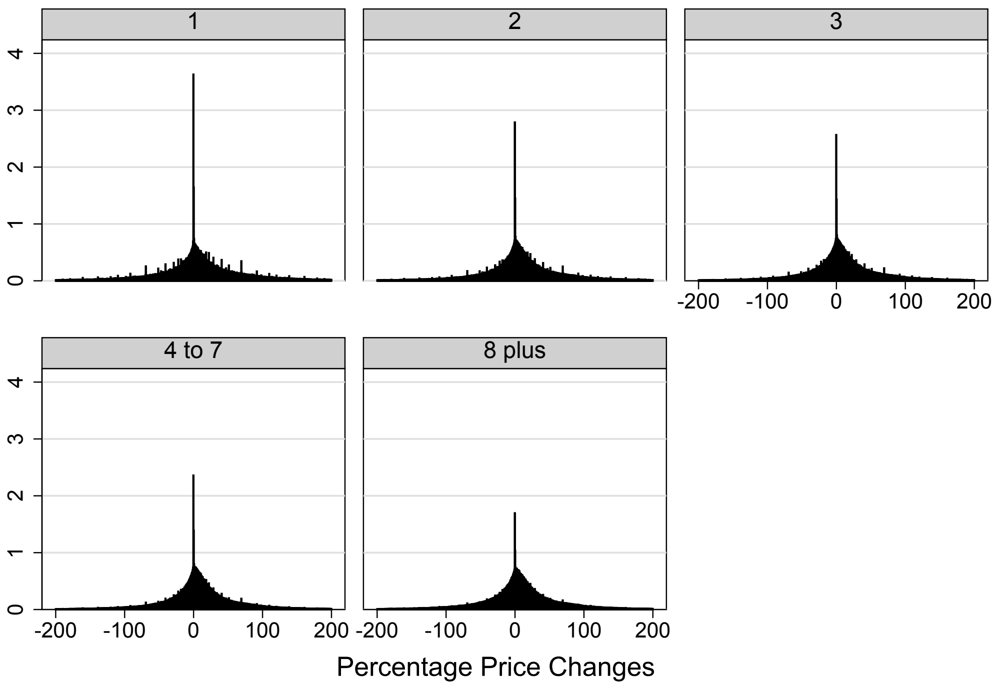
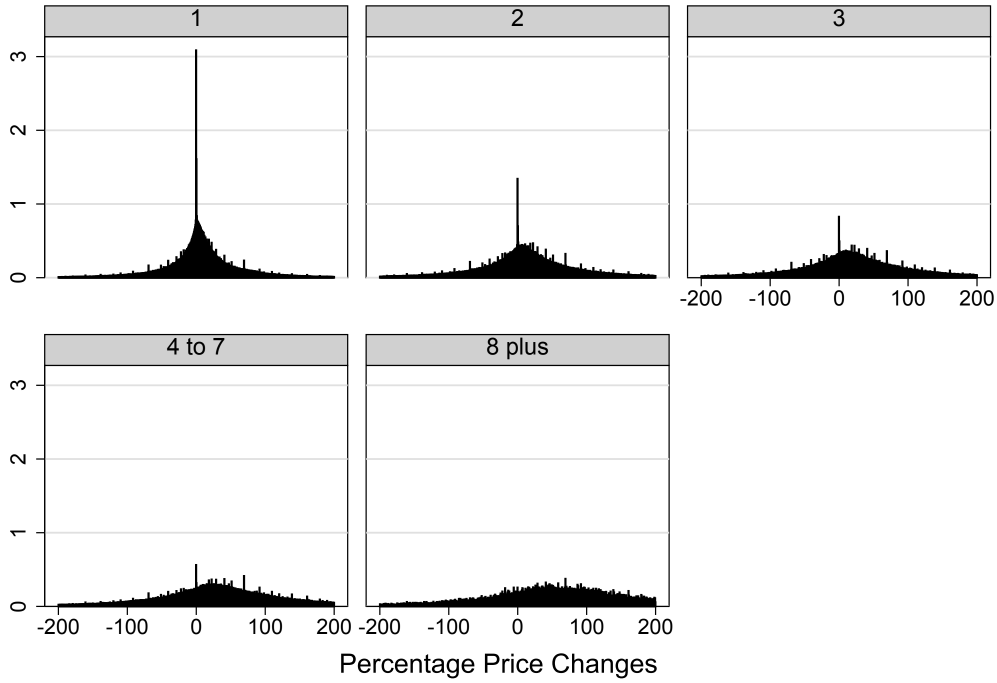


Fig. 2.C.2 Distribution of price changes between t and $t+s$, $s = 1, 2, ...$



2.C.1 Chinese count and mass classifiers and the construction of unit values

We construct unit values at the firm-product-destination level as the value of exports divided by the reported quantity of exports. The Chinese customs authority reports 30 distinct quantity measures which we use in construction of unit values. In the Chinese language, these quantity measures, or classifiers, are intrinsically informative about the nature of the good being traded. Of the 30 quantity measures in our estimation dataset, 22 are count classifiers and 8 are mass classifiers.[55] In Chinese, count classifiers are used to measure distinct items while mass classifiers are used to measure things that are naturally measured by weight, volume, length, etc. In our dataset for 2007, 42 percent of transactions are of goods recorded with count classifiers and 58 percent are of goods recorded with mass classifiers. We argue that the use of classifiers as a reporting unit implies that unit values in Chinese data are more closely aligned to transactions prices than they are in other national customs datasets. Firstly, the high share of transactions which use count classifiers combined with the wide variety of count classifiers used to measure similar products means that unit values for these HS08 products are closely aligned with the true transactions price. Secondly, those

[55] According to Cheng and Sybesma (1999), p. 515, "...the distinction between the two types of classifiers is made with explicit reference to two different types of nouns: nouns that come with a built-in semantic partitioning and nouns that do not – that is, count nouns and mass nouns."

goods whose measurement is reported in mass classifier units are items whose intrinsic nature implies that the most appropriate metric is the reported measure of kilograms, square meters, etc. Again, this implies that unit values for these commodities will be close to true transactions prices.

To illustrate the variety of count classifiers used for similar objects, note that "Women's or girls' suits of synthetic fibres, knitted or crocheted" (HS61042300) and "Women's or girls' jackets & blazers, of synthetic fibres, knitted or crocheted" (HS61043300) are measured with two distinct Chinese count classifiers, "套" and "件," respectively. Further, table 2.C.3 documents the intrinsic information content of the measurement units for HS04 product groups 8211 and 8212. The Chinese language descriptions of all of these HS08 products convey the similarity across products; each Chinese description contains the Chinese character 'dao' (刀), which means 'knife' and is a part of longer compound words including table knife and razor. Interestingly, three different Chinese count classifiers, "套," "把," and "片," are used to count sets of knives (HS82111000), knives and razors (HS82119100 - HS82121000), and razor blades (HS82122000), respectively.

Table 2.C.3 Examples of quantity classifiers in Chinese customs data

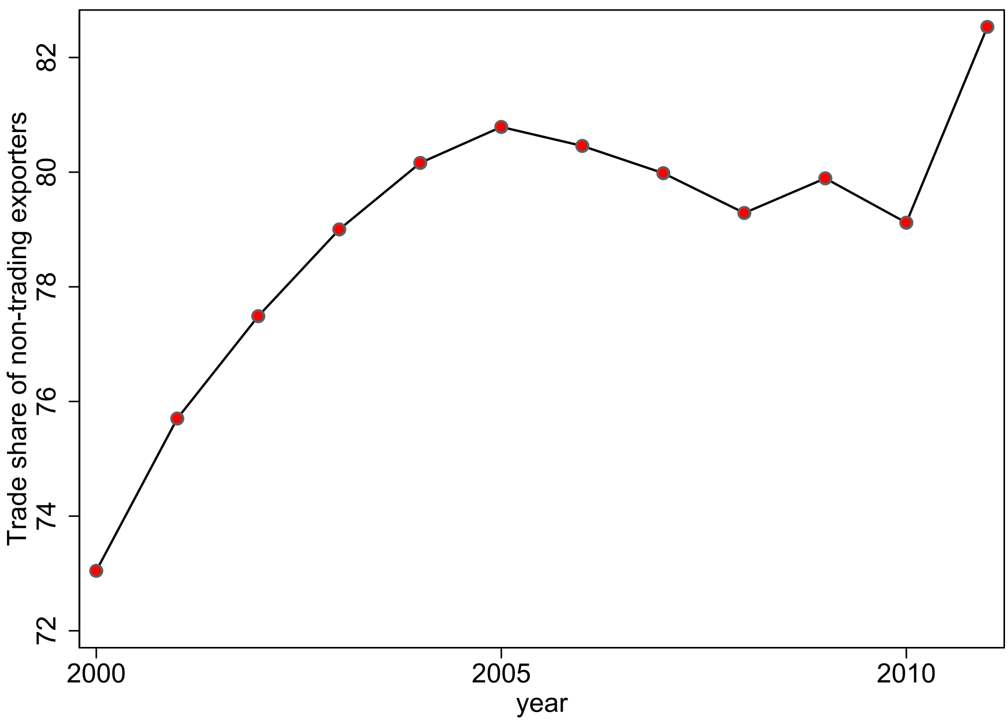
Quantity Measure	HS08 Code	English Description	Chinese Description
套	82111000	Sets of assorted knives	成套的刀
把	82119100	Table knives having fixed blades	刃面固定的餐刀
把	82119200	Other knives having fixed blades	其他刃面固定的刀
把	82119300	Pocket & pen knives & other knives with folding blades	可换刃面的刀
把	82121000	Razors	剃刀
片	82122000	Safety razor blades, incl razor blade blanks in strips	安全刀片, 包括未分开的刀片条

The most frequently used mass classifier is kilograms (56 percent of transactions). Examples of other mass classifiers include meters for "Knitted or crocheted fabric of cotton, width $\leq$ 30cm" (HS60032000), square meters for "Carpets & floor coverings of man-made textile fibres" (HS57019010), and liters for "Beer made from malt" (HS22030000).

2.C.2 Classification of firms as non-trading (manufacturing) or trading firms

We identify trading firms by the inclusion of certain Chinese characters in the firm's name, i.e., "jinchukou" (import-export), "maoyi" (trade), "shangmao" (business and trade), "jingmao" (economic and trade), "cangchu"(warehouse), "wuliu"(logistics), and "huowu" (cargo). As described by Lu et al. (2013), the tradition of using Chinese characters to identify a firm's activities in its self-revealing name has been a feature of China's international engagement since the reforms of 1978. This practice has continued to the present day. Ahn et al. (2011) document the use of specific characters in a name correlates with differences in trading volume, product categories, and export destinations. Figure 2.C.3 shows that the share of Chinese export value directly conducted by non-trading (manufacturing) firms rose steadily from 2000 to 2005 as the Chinese government liberalized export markets to allow more firms to directly export. Interestingly, after 2005, firms designated as trading firms continued to account for one-quarter of Chinese export value and dominate multi-destination trade.

Fig. 2.C.3 The share of Chinese exports by non-trading (manufacturing) firms



Chapter 3

Firm Level Exchange Rate Pass Through – A Machine Learning Approach

Understanding how exporters react to exchange rate shocks is important for evaluating international shock transmissions and setting optimal international monetary policy. Empirical studies have documented substantial heterogeneity in the degree to which different firms and products respond to exchange rate shocks. In addition, estimates of exchange rate pass through (ERPT) are time varying and depend on observed and unobserved variables in a nonlinear way. This paper proposes a machine learning algorithm that systematically detects the determinants of ERPT and estimates ERPT at the firm level in a large-scale custom dataset. The accuracy of the algorithm is tested on simulated data from an extended multi-country version of Atkeson and Burstein (2008). Applying the algorithm to China's custom data from 2000-2006, this paper estimates the ERPT of China's exporters and documents new evidence on the nonlinear relationships among market structures, unit value volatility and ERPT.

3.1 Introduction

In the last decade, the increasing availability of large scale firm level datasets has greatly enlarged the ability to understand firm level heterogeneity and its implications at the aggregate level. Especially in the literature of ERPT, understanding why firms have different pricing behaviour in response to a common exchange rate shock has important implications in setting the optimal monetary policy[1].

Unlike micro studies in other fields, international trade firm-level datasets recently made available contain a significant proportion of firms in an economy and almost all custom transactions at firm product (8-digit) level in a given period. As the richness and the scale of micro dataset available to researchers develop rapidly, conventional methods applied extensively by empirical researchers, fixed effects related methods for example, are either less flexible in their functional assumptions or not very effective in gathering all possible aspects of heterogeneity[2] in a large scale dataset. Therefore, these methods may not be the best option to understand firm-level heterogeneities. Conventional ERPT estimation methods generate large standard errors when applied at sector/firm level due to unobserved variables, e.g. marginal cost, heterogeneity in product characteristics and market structures. Empirically, researchers trade off controlling for unobserved variables against the flexibility of functional forms[3].

On the other hand, recent researches in machine learning focus especially on large datasets and heterogeneities. It seems to be the natural alternative for trade problems. In spite of successful applications of these algorithms in various subjects, economists often stay alarmed with the usage of these algorithms for two reasons. First, a machine learning algorithm often involves extracting maximum amount of information in a certain dataset and thus the results are often data driven and may not necessarily reflect the true relationship between variables being studied. Second, a machine learning algorithm may be good in making predictions but does not identify causal relationships, nor does it enhance our economic understanding.

For causality, the mainstream empirical papers working with firm-level data take a restrictive approach. Methods include restricting the dataset so that the subgroup being studied is no longer subject to omitted variable bias or adding multiple fixed effects such that "irrelevant" and possible confounding variations can be partitioned out. However, adding restrictions and layers means dropping observations and losing information[4]. These

[1] Seminal contributions include Dornbusch (1987), Corsetti and Pesenti (2005), Corsetti and Dedola (2005), Corsetti, Dedola, and Leduc (2007), Corsetti, Dedola, and Leduc (2008a), Corsetti, Dedola, and Leduc (2008b), Atkeson and Burstein (2008).

[2] The ability to search for all possible aspects of heterogeneities is important for predicting the firm-level response to policy shocks.

[3] For example, by dividing data into several bins according to destinations, quantiles of market shares, etc.

[4] Not a problem if obtaining one (aggregate) coefficient is the main concern but costly in understanding firm-level heterogeneities.

restrictions can help us build an "ideal" environment to study the hypothetical relationship but also limit our vision to a particular hypothetical situation.

Thanks to persistent advocates of applying machine learning methods to Economics[5], pioneer works on adapting machine learning methods to make casual inferences and solve policy problems have made a significant progress[6]. However, existing studies work under the condition of unconfoundedness. This condition will not be satisfied for international trade related problems. The marginal cost of the product being sold and the prices of competitors are unobserved and endogenous to exchange rate movements. Estimates ignoring these confounding variables will lead to biased point estimates of individual treatment effects.

This paper follows recent work on making casual inferences and proposes an algorithm specifically designed to estimate firm-level heterogeneities in response to macro shocks in a multi-dimensional panel. The proposed algorithm features in two aspects: (a) it uses the high predicting power of the gradient boosting regression tree algorithm (GBRT)[7] to construct counterfactual environments; (b) it uses orthogonal variations across dimensions to control for unobservables. The proposed algorithm contributes to the machine learning literature in its awareness that the monotonic property of tree based algorithms can be used together with orthogonal variations across dimensions to control for unobserved components[8].

The central idea behind the proposed algorithm is that machine learning algorithms making causal inferences should be assisted with structural information implied by economic models. Most machine learning algorithms take an agnostic data driven approach. As in most estimation techniques, adding correct structural assumptions will increase the precision of estimation. However, how to add economic assumptions into a machine learning algorithm is still not clear[9]. This proposed algorithm presents a novel approach to feed

[5]See Varian (2014) for an introduction of various machine learning algorithms and how they can be applied to study economic questions.

[6]See Athey and Imbens (2015), Bajari, Nekipelov, Ryan, and Yang (2015), Chernozhukov, Hansen, and Spindler (2015), Kleinberg, Ludwig, Mullainathan, and Obermeyer (2015). Pioneer works on adapting machine learning methods to make casual inferences include Wager and Athey (2015), Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, et al. (2016), Athey and Imbens (2016), Athey, Tibshirani, and Wager (2016), Athey, Imbens, Pham, and Wager (2017).

[7] GBRT is featured in its power of out-of-sample prediction accuracy due to its ability to capture nonlinearities. GBRT algorithm has been widely applied in frontier studies of a wide range of topics, e.g. global distribution and the risk of dengue [Bhatt et al. (2013)], effect of climate change [Cox et al. (2013), Randall and Van Woesik (2015)]. The algorithm is proved to be effective in solving practical classification and prediction problems and has been actively implemented in international computing and machine learning challenges [For details, click List of Winning Solutions]. An introduction to the GRBT algorithm can be found in the appendix.

[8]Most custom datasets have disaggregated and detailed firm level transaction records, but key elements such as the marginal costs are not observed and are difficult to be estimated even when one can complement the custom database with some industry surveys.

[9]In general, the range of structural assumptions is not limited to probabilistic assumptions and could include information on how variables interact in a structural model. This paper only works on a limited case and shows that adding the log-linearised version of the structural pricing equation will significantly improve firm level ERPT estimates of the proposed algorithm.

structural information into a tree based machine learning algorithm in a multi-dimensional panel framework[10].

The proposed algorithm is designed to work directly with large scale custom datasets and identify the ERPT parameter for each exporter in an economy. The algorithm learns from reading records of trade patterns. It not only predicts export prices and quantities at the firm level conditioning on values of future environments of aggregate variables but also generates the genetic rules governing the data generating process[11]. To assess the performance of the proposed algorithm, I build the following multi-country trade model.

In order to understand how firms optimally price their products under a multi-sector multi-country environment, I extend the two-country model of Atkeson and Burstein (2008) to a multi-country framework and introduce heterogeneity in productivity distributions across sectors and countries. The main features of the model are as follows. First, there are N countries in the world and each country owns S sectors with heterogeneous productivity distributions. Second, within each sector there are local firms as well as exporters from other countries competing under Cournot competition with a demand elasticity structure similar to Atkeson and Burstein (2008). The result of the competition is determined by the productivity distribution of the participating firms as well as aggregate variables such as bilateral exchange rate shocks. Due to the Cournot competition structure, there is no closed form solution for the model[12]. I construct counterfactual environments to understand how ERPT differs under different productivity distributions of local firms and foreign exporters and different compositions of exchange rate shocks.

In terms of macroeconomic modeling, although such a framework may be helpful in understanding the pricing behavior of an exporter facing competition from local competitors and other exporters from other countries, the need to add extra levels of heterogeneity seems to be less justified. The main drawback of a micro-founded multi-sector multi-country trade model is the lack of ability to map into the real world due to its demanding requirement in calibration. In practice, estimating the productivity distribution for a particular country/sector, provided the existence of good data, is already a challenge[13].

However, such highly detailed micro-founded models may have much to offer in an alternative modeling strategy. The procedure is given as follows. First, notice what data are available in reality. Second, simulate the highly micro founded model with arbitrary calibration. Third, subset the simulated data and construct a dataset similar to what is observed in the reality. Fourth, write an algorithm or econometric method to estimate key

[10] As I will illustrate in the model section, adding structural information has a marginal effect on the predictive power of the dependent variable but is critical in getting the correct causal inference.

[11] For example, how ERPT depends on observed firm level characteristics.

[12] A market of N firms would give N simultaneous equations.

[13] For this particular question of interest, one would need to estimate the productivity distributions for each country in the world, which is difficult.

parameters of interest[14] for each individual/firm of interest in the constructed dataset. Fifth, change the parameter value of the model, re-simulate and construct a new dataset. Test the algorithm's ability to estimate the key parameter of interest and revise the algorithm if not. Sixth, apply the algorithm to observed data. If the model is believed to be correct, the estimates from the algorithm are reliable.

The advantage of this approach is that we will have a structural model that enables us to understand how the mechanism works and figure out the key variable of interest in reality provided that the model is correctly specified[15]. This approach is useful in a scenario where we do not have full information to estimate the whole model but may have enough information to identify a subset of parameters implied by the structural model. The proposed approach is similar to the indirect inference approach but differs in that my proposed approach, particularly steps 4 and 5, puts emphasis on building an algorithm that provides the correct estimates of interest for all possible calibrations of the micro-founded model. The indirect inference approach emphasises on finding an auxiliary model such that estimates from the true model or data are as close as possible to the estimates from the auxiliary model.

The algorithm is proved to be successful in recovering the true ERPT parameter at firm level in the simulated model and is applied to the custom database of China from 2000-2006. Using this nonparametric approach, my finding confirms that ERPT is a nonlinear function of firm-level characteristics depending on various measures of market structure. Consistent with theoretical and empirical works, the relationship between ERPT and several market share measures resembles a U shape[16].

The rest of the paper is organised as follows. Section 2 formalises the empirical question. Section 3 introduces the proposed algorithm and explains the mechanism behind it. Section 4 presents the theoretical model of ERPT and various exercises for recovering the true ERPT estimates using the proposed algorithm. Section 5 presents empirical results on China's custom data. Section 6 concludes.

3.2 Problem

This section gives a formal presentation of the empirical question that this paper tries to address. In addition, I construct two-dimensional numerical examples to illustrate how

[14] Please note that these parameters are not necessarily the same as the parameters needed for calibration.

[15] If the model specification is in doubt, an additional loop between step 2 and 5 can be added to evaluate possible alternatives in model specification. I am still working on the proper way to evaluate and compare different model specifications under my proposed procedure.

[16] Krugman (1986), Dornbusch (1987), Atkeson and Burstein (2008), Melitz and Ottaviano (2008), Chen, Imbs, and Scott (2009), Berman, Martin, and Mayer (2012), Amiti, Itskhoki, and Konings (2014), Auer and Schoenle (2016)

conventional fixed effects related methods may fail to capture the nonlinear features of ERPT estimates.

The pricing equation of exporters can be presented as follows.

$$p_{\mathcal{I}} = g\left(e_{\mathcal{I}_1}, X_{\mathcal{I}}, M_{\mathcal{I}_2}, \epsilon_{\mathcal{I}}\right)$$

where $\mathcal{I} = \{i, f, d, t\}$ represents the dimensions along which a variable varies, with i, f, d, t standing for product, firm, destination and time respectively; the missing variables vary along dimensions that satisfy $\mathcal{I}_2 \subset \mathcal{I}$ and $\mathcal{I}_1 \neq \mathcal{I}_2$; g is an unknown function; p is a scalar dependent variable representing the exporter's price; e is the key variable of interest, the bilateral nominal exchange rates; X is a vector of observed feature variables; M is a vector of unobserved variables that correlate with e; ϵ is an error term that does not correlate with e. The objective is to understand how changes in exchange rates affect the exporter's price conditioning on the set of observed firm level characteristics X, such as various market share measures, and unobserved variables M that are not varying along all dimensions, such as firm specific marginal costs.

$$\frac{\partial g(.)}{\partial e_{d,t}}$$

3.2.1 Conventional approaches in the literature are not informative about the underlying structure of ERPT

In this subsection, I construct examples to explain why conventional fixed effect methods are not very informative. In my examples, I restrict my focus to the same exporter selling the same product to different destinations d over a certain time period t . For simplicity, I assume a linear process of export prices $p_{d,t}$ that depends on bilateral exchange rates $e_{d,t}$, market shares $ms_{d,t}$, and marginal cost of the product mc_t. $\beta_{d,t}$ represents the ERPT coefficient which is assumed to be a nonlinear function of market shares and marginal costs.[17]

$$p_{d,t} = \mu + \beta_{d,t} e_{d,t} + ms_{d,t} + mc_t + \epsilon_{d,t}$$

In constructing the rest of series, I assume simple linear relationships under which all explanatory variables, $e_{d,t}, ms_{d,t}$ and mc_t, are correlated with each other. Specifically, market shares are constructed to be linear in a destination time specific factor and nominal exchange

[17] A more realistic model is discussed in section 4.

rates. Marginal cost is constructed similarly.

$$ms_{d,t} = u_{d,t} + 0.1e_{d,t}$$
$$mc_t = u_t - 0.1\overline{e_t}; \quad \overline{e_t} := \frac{\sum_d e_{d,t}}{n_d}$$
$$e_{d,t} \sim N(0,1),\ u_{d,t} \sim uniform(0,1),\ \epsilon_{d,t} \sim N(0,0.01)$$

Next, I simulate the model for three different underlying ERPT functions $\beta_{d,t}$ and compare results of applying the standard fixed effect estimator with destination and time fixed effects.

$$Spec1: \quad \beta_{d,t} = (ms_{d,t} - 0.5)^2 + mc_t;\ u_t \sim uniform(0,1)$$
$$Spec2: \quad \beta_{d,t} = 2(ms_{d,t} - 0.5)^2 * mc_t;\ u_t \sim uniform(0,1)$$
$$Spec3: \quad \beta_{d,t} = 2(ms_{d,t} - 0.5)^2 * mc_t;\ u_t \sim N(0,1)$$

The objective is to read simulated data records of $d, t, p_{d,t}, e_{d,t}, ms_{d,t}$ and estimate the ERPT $\beta_{d,t}$. There are two difficulties in estimating ERPT in this simulated example: (a) the ERPT is not a constant parameter but an unknown function of firm characteristics; (b) the marginal cost mc_t is not observed.

Table (3.1) presents results with each specification being simulated for 2000 destinations[18] and 40 time periods. Columns (1) - (5) resemble the empirical discovery process of the relationship between ERPT and market shares. The estimated coefficients in column (1) represent a general response of prices to exchange rates. Column (2) adds market share in levels and finds significant coefficients for both variables. Columns (3) and (4) try different interaction terms between exchange rates and market share but no significant result is found. This reflects the main drawback of fully specified structural equations compared to nonparametric approaches. The rejection of one specification is not informative about the alternative right specification. If the researcher stops at regression (4), the discovery that ERPT is U shaped in market share is likely to be delayed.

Even at the correct regression specification column (5)[19], results are not very informative about the underlying structure driving the heterogeneity of ERPT due to the existence of the unobserved marginal cost. The estimated coefficients of the interaction terms from column (2) - (5) are very similar to the results under specification 1. From regression results under specification 1 and 2, it is difficult to make an inference on how $\beta_{d,t}$ depends on firm-level

[18] In my original experiment, I set this number to be significantly bigger than the number of time periods n_T. A more realistic example where $n_D = 200$ can be found in the appendix.

[19] Regression (5) is the specification closest to the true data generating process among these three specifications. For example, specification 1 can be rewritten as follows: $p_{d,t} = 10 + ms_{d,t}^2 e_{d,t} - ms_{d,t}e_{d,t} + (mc_t + 0.25)e_{d,t} + ms_{d,t} + mc_t + \epsilon_{d,t}$ where $\overline{mc_t} = \frac{\sum_t mc_t}{n_T} = 0.5$. Coefficient on $ms_{d,t}$ is close to the theoretical value of 1. The exchange rate interacting with market share has a significant coefficient close to -1 and the interaction term with exchange rate squared has a coefficient close to 1. The coefficient on $e_{d,t}$ is slightly downward biased as the mean of mc_t equals 0.5, which gives the theoretical value of 0.815.

Table 3.1 Estimates from the fixed effect method

	(1)	(2)	(3)	(4)	(5)
Specification 1: $p_{d,t} = \mu + \left[(ms_{d,t} - 0.5)^2 + mc_t\right] e_{d,t} + ms_{d,t} + mc_t + \epsilon_{d,t}$					
$e_{d,t}$	0.757***	0.635***	0.759***	0.638***	0.796***
	(0.002)	(0.001)	(0.003)	(0.002)	(0.003)
$ms_{d,t}$		1.198***		1.198***	0.995***
		(0.004)		(0.004)	(0.005)
$e_{d,t} * ms_{d,t}$			-0.006	-0.006	-1.019***
			(0.006)	(0.004)	(0.011)
$e_{d,t} * ms^2_{d,t}$					1.016***
					(0.011)
Adjusted R^2	0.663	0.865	0.722	0.865	0.879
Specification 2: $p_{d,t} = \mu + \left[2(ms_{d,t} - 0.5)^2 * mc_t\right] e_{d,t} + ms_{d,t} + mc_t + \epsilon_{d,t}$					
$e_{d,t}$	0.212***	0.090***	0.217***	0.093***	0.242***
	(0.002)	(0.001)	(0.003)	(0.001)	(0.001)
$ms_{d,t}$		1.197***		1.196***	1.002***
		(0.002)		(0.002)	(0.002)
$e_{d,t} * ms_{d,t}$			-0.009**	-0.005***	-0.962***
			(0.004)	(0.002)	(0.004)
$e_{d,t} * ms^2_{d,t}$					0.957***
					(0.004)
Adjusted R^2	0.168	0.825	0.227	0.825	0.885
Specification 3: $p_{d,t} = \mu + \left[2(ms_{d,t} - 0.5)^2 * mc_t\right] e_{d,t} + ms_{d,t} + mc_t + \epsilon_{d,t}; \quad u_t \sim N(0,1)$					
$e_{d,t}$	-0.177***	-0.275***	-0.180***	-0.279***	-0.088***
	(0.017)	(0.016)	(0.016)	(0.016)	(0.019)
$ms_{d,t}$		1.005***		1.003***	1.038***
		(0.016)		(0.016)	(0.016)
$e_{d,t} * ms_{d,t}$			0.302***	0.296***	0.301***
			(0.016)	(0.016)	(0.016)
$e_{d,t} * ms^2_{d,t}$					-0.190***
					(0.011)
Adjusted R^2	0.001	0.049	0.006	0.053	0.056
Time FE	yes	yes	yes	yes	yes
Individual FE	yes	yes	yes	yes	yes
Observations	80,000	80,000	80,000	80,000	80,000

Note: This table presents estimation results after applying the conventional fixed effect estimator to Monte-Carlo simulated data from specification 1 to 3.

characteristics such as market share $ms_{d,t}$ and marginal cost mc_t. Specification 3 shows that the estimated coefficients can be very sensitive to the distribution of the unobserved variable mc_t where the random factor u_t is assumed to be standard normally distributed rather than uniformly distributed.

3.3 Algorithm

This section explains the proposed algorithm. The first part of this section introduces the general property which the proposed algorithm relies on under the framework of statistical learning theory. The second part explains how this property can be exploited to control for unobserved variables in tree based algorithms.

3.3.1 The proposed idea under statistical learning theory

A standard statistical learning problem can be formulated as follows. Consider an input space $\mathcal{X}$ and output space $\mathcal{Y}$. $(X, Y) \in \mathcal{X} \times \mathcal{Y}$ are random variables with an unknown joint distribution P. We observe a sequence of n i.i.d. pairs of (X_i, y_i) sampled according to P. The goal of the learning problem is to construct a function $g : \mathcal{X} \to \mathcal{Y}$ such that this function minimises the risk of all possible measurable functions:

$$R(g) := \int h(g(X), Y) dP$$

where $h(.)$ is a criterion function[20]. Empirically, the optimal g is given by

$$\widehat{g}_n := \arg\min_{g \in \mathcal{G}} \frac{1}{n} \sum_{i=1}^{n} h(g(X_i), p_i)$$

where the expectation is taken over the distribution of P_{XY}. $\mathcal{G}$ is a space of allowed functions depending on the classification algorithm. $\widehat{g}_n$ stands for the estimated function f from data. The main concern of the statistical learning theory is to establish bounds for $R(\widehat{g}_n) - \inf_g R(g)$ so that we know when empirical error $R(\widehat{g}_n)$ is a good representation of the true risk measure $\inf_g R(g)$. This measure can be further decomposed into two components, the estimation error and the approximation error.

$$R(\widehat{g}_n) - \inf_g R(g) = \underbrace{\left(R(\widehat{g}_n) - \inf_{g \in \mathcal{G}} R(g) \right)}_{\text{Estimation Error}} + \underbrace{\left(\inf_{g \in \mathcal{G}} R(g) - \inf_g R(g) \right)}_{\text{Approximation Error}} \tag{3.1}$$

[20]Empirically, $h(,)$ may take the form of $|y - g(X)|$ or $(y - g(X))^2$ depending on the assumptions of P.

Estimating the approximation error is usually hard since it requires knowledge about the target.[21] Important contribution on establishing the relationship between estimation error bounds and entropy measures of classifiers/algorithms has been made by Vladimir N. Vapnik[22].

The method introduced in this paper takes a different appraoch. Instead of assuming that the observed set of variables are given, I consider a parallel set of learning problems. The dependent variable will be the same and take the same value. Most feature variables will also be the same. But some feature variables may be different.

Suppose we have a set of parallel learning problems indexed by $1,...,j$. For each j, we have an input space $\mathcal{X} \times \mathcal{M}^{(j)}$ and an output space $\mathcal{Y}$. $(X, M^{(j)}, Y) \in \mathcal{X} \times \mathcal{M}^{(j)} \times \mathcal{Y}$ are random variables with joint distribution $P^{(j)}$ unknown to us. We observe n i.i.d. pairs of $(X_i, M_i^{(0)}, y_i)$. We know that $M_i^{(j)}$ is a function of $M_i^{(0)}$. For each j, we have the conventional learning problem of constructing a function $g^{(j)} : \mathcal{X} \times \mathcal{M}^{(j)} \to \mathcal{Y}$ such that this function minimises the risk:

$$R^{(j)}(g) := \int h(g(X, M^{(j)}), Y) dP^{(j)}$$
$$\widehat{g}_n^{(j)} := \arg\min_{g \in \mathcal{G}} \frac{1}{n} \sum_{i=1}^{n} h(g(X_i, M_i^{(j)}), p_i)$$

Define the numerical measure of the partial derivative as

$$h_2(g, x_1, X_{-x_1}, M^{(j)}, \epsilon) := \frac{g(x_1 + \epsilon, X_{-x_1}, M^{(j)}) - g(x_1 - \epsilon, X_{-x_1}, M^{(j)})}{2\epsilon}$$

Suppose we are originally interested in the case $g^{(0)} : \mathcal{X} \times \mathcal{M}^{(0)} \to \mathcal{Y}$, the question is to what extent we can infer the answer of (0) from results from $g^{(j)} : \mathcal{X} \times \mathcal{M}^{(j)} \to \mathcal{Y}$. In this case, we

[21]Most statistical learning algorithms take an agnostic approach and avoid making specific assumptions about the underlying distribution.

[22]See Vapnik (1999) for a literature review.

can write the problem as an expression similar to equation (3.1):

$$\int h_3[h_2(\widehat{g}_n^{(j)},x_1,X_{-x_1},M^{(j)},\epsilon) - h_2(\arg\inf_g R^{(0)}(g),x_1,X_{-x_1},M^{(0)},\epsilon)]dP^{(0)} =$$

$$\int h_3 \Bigg[\underbrace{\Big(h_2(\widehat{g}_n^{(j)},x_1,X_{-x_1},M^{(j)},\epsilon) - h_2(\arg\inf_{g\in\mathcal{G}} R^{(j)}(g),x_1,X_{-x_1},M^{(j)},\epsilon)\Big)}_{\text{Estimation Error}}$$
$$+ \underbrace{\Big(h_2(\arg\inf_{g\in\mathcal{G}} R^{(j)}(g),x_1,X_{-x_1},M^{(j)},\epsilon) - h_2(\arg\inf_{g\in\mathcal{G}} R^{(0)}(g),x_1,X_{-x_1},M^{(0)},\epsilon)\Big)}_{\text{Substitution Error}}$$
$$+ \underbrace{\Big(h_2(\arg\inf_{g\in\mathcal{G}} R^{(0)}(g),x_1,X_{-x_1},M^{(0)},\epsilon) - h_2(\arg\inf_{g} R^{(0)}(g),x_1,X_{-x_1},M^{(0)},\epsilon)\Big)}_{\text{Approximation Error}} \Bigg] dP^{(0)}$$

The first term, the estimation error, is where most frontier machine learning algorithms making casual inferences work on[23]. The third term is a conventional term but very difficult to measure without prior assumptions.

The second term is new from this paper. It reflects the effect of substituting the learning problem from (0) to (j). Note that this substitution error depends on three elements, i.e. the group of allowed functions $\mathcal{G}$, the relationship between the variable of interest x_1 and $\mathcal{M}^{(0)}$, and the relationship between input spaces being substituted $\mathcal{M}^{(0)}$ and $\mathcal{M}^{(j)}$.

The interesting part is that the linkage between the size of the substitution error and the set of allowed functional classes $\mathcal{G}$, and the relationship between input spaces being substituted $\mathcal{M}^{(0)}$ and $\mathcal{M}^{(j)}$ can be exploited to control for unobserved variables by adding the third channel of optimisation.

Consider two cases. If two input spaces $\mathcal{M}^{(0)}$ and $\mathcal{M}^{(j)}$ are very different from each other, the range of allowed functions $\mathcal{G}$ will have a considerable impact on the substitution error through affecting $\widehat{g}_n^{(j)}$ and $\widehat{g}_n^{(j)}$ being selected. On the contrary, if two input spaces $\mathcal{M}^{(0)}$ and $\mathcal{M}^{(j)}$ are identical, the set of functions $\mathcal{G}$ can be of any range and do not change the substitution error.

Alternatively, for any given set of functions $\mathcal{G}$, it is possible to figure out the maximum "distance" between $\mathcal{M}^{(0)}$ and $\mathcal{M}^{(j)}$ such that the substitution error is zero.

[23]e.g. Wager and Athey (2015), Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, et al. (2016), Athey and Imbens (2016), Athey, Tibshirani, and Wager (2016), Athey, Imbens, Pham, and Wager (2017).

3.3.2 Tree based algorithms

This paper exploits a special case where the set of functions $\mathcal{G}$ are tree based algorithms and designs a procedure that can be applied to control for variables that do not vary along all dimensions in a multi-dimensional panel.

The next two subsections explain how tree based algorithms can be exploited to control for unobserved components and facilitate the identification of casual inferences. Subsection 3.3.2 starts with a simple example and outlines the condition (compared with the conventional monotonic transformation) that needs to be satisfied for a variable to be a good proxy for the unobserved variable. I will refer to this condition as weak monotonic transformation. I also discuss additional simulation results for general cases. These simulation results suggest that a more general form of the monotonic transformation condition exists. I will try my best to describe this condition and refer to it as the conditional monotonic transformation property.

In practice, even a conditional (weak) monotonic transformation of the unobserved variable is difficult to find. Subsection 3.3.2 shows that, in a multi-dimensional panel, parameters from structural estimations in a set of restricted dimensions can be used as a proxy for the weak monotonic transformation of unobserved variables not varying along all dimensions.

One-dimensional example

In this subsection, I temporarily abstract away from my ERPT question and discuss this one-dimensional example that helps to understand and clarify the mechanism of the proposed algorithm. Consider the case of identifying the individual treatment effect.[24]

$$\begin{aligned} y_i &= \beta_i T_i + M_i \\ \beta_i(M_i) &:= M_i \\ T_i &\in \{0,1\},\ M_i \in \{0,1\} \end{aligned}$$

where T_i is a treatment indicator randomly drawn from $\{0,1\}$ with equal probability and β_i is the treatment effect for individual i. M_i is the unobserved variable. In this first example, I assume $\beta_i(M_i) = M_i$ for simplicity. More general cases are discussed in later sections. The objective is to find β_i given data of individual outcomes y_i and its treatment indicator T_i. The data generating process (the functional form of each variable) is unknown to economists. M_i is unobserved.

Suppose all explanatory variables are observed and the functional form is known, obtaining the causal inference is equivalent to estimating parameter values of the function

[24] You can also think of this setting in terms of the conventional framework characterising treatment effects: $y_i = [y_{1i}(M_i) - y_{0i}(M_i)]T_i + y_{0i}(M_i)$ with $y_{1i}(M_i) = 2M_i$ and $y_{0i}(M_i) = M_i$.

and taking the partial derivative. Similarly, if all explanatory variables are observed but the functional form is unknown, one could fit a nonparametric function and then perform a numerical partial differentiation with the estimated model. That is, suppose M_i is observed, we can estimate the individual treatment effect β_i using the following two-step procedure:

1. Use a nonparametric econometric method or a machine learning algorithm to recognise the pattern of y_i using T_i and M_i. Obtain

$$model_1 : (T_i, M_i) \to p_i$$

2. Use $model_1$ to construct counterfactual predictions conditioning on the value of M_i and calculate individual treatment effect[25].

$$\beta_i^{Est} = model_1(1, M_i) - model_1(0, M_i)$$

In this procedure, the ability to make predictions conditioning on the value of M_i is important. If the explanatory variable M_i is unobserved, the individual treatment effect β_i will not be identifiable in general.

In many cases, economists do not observe M_i. But it may be possible to have/create a variable $\mathfrak{M}_i$ that preserves some structural information in M_i. If we could construct counterfactuals conditioning on the structural information provided by $\mathfrak{M}_i$, we will be able to recover β_i using the above procedure. In general, the structural information contained in the alternative variable $\mathfrak{M}_i$ could be highly nonlinear. I find that the tree based algorithms have a unique advantage in addressing this type of problems.

Consider the following data generating process of 200 individuals:

Table 3.1 Values of y_i

y_i	β_i	M_i	T_i
0	0	0	0
0	0	0	1
1	1	1	0
2	1	1	1

Table 3.2 Assignment of M_i

M_i	i
0	1-100
1	101-200

In this constructed example, the assignment of M_i happens to be ordered and I want to utilise the information provided by the index i to estimate the individual treatment effect β_i. To capture the highly nonlinear information contained in i, I will run a tree based algorithm

[25] In the case where T_i is continuous, a similar numerical derivation can be obtained by $\beta_i^{Est} = \frac{model_1(T_i+c,i)-model_1(T_i-c,i)}{2c}$

with the supervisor/dependent variable y_i on explanatory variables T_i and i in step 1 and construct counterfactuals based on the estimated tree structure $model_1(T_i,i)$ in step 2.

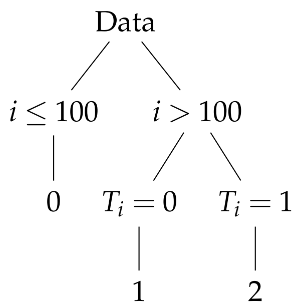
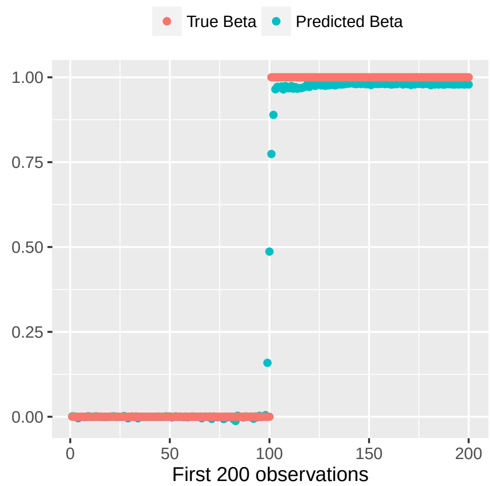


Fig. 3.1 Fitted $model_1(T_i,i)$

The result of applying a decision tree is presented in figure 3.1. In producing $model_1(T_i,i)$, the algorithm compares the resulted MPE[26] between the best split of T_i and the best split of i. Splitting the sample into $T_i = 0$ and $T_i = 1$ will result in $MPE = 0.5$, while splitting into $i \leq 100$ and $i > 100$ (the best split for i) will result in $MPE = 0.25$. Therefore, the algorithm will choose to split i for its first split. After this split, the MPE for the subgroup $i \leq 100$ is 0 and no further split is needed. The algorithm will try to find the best split for the subgroup $i > 100$. In this case, the best split will be $T_i = 0$ and $T_i = 1$.

	$\hat{y}_i$ Evaluated at			
i	$T_i = 1$	$T_i = 0$	Estimated β_i	T
1-50	0	0	0	
51-100	0	0	0	
101-150	2	1	1	
151-200	2	1	1	



The table on the left hand side illustrates the second step. The right hand side graph compares the true β_i with the estimated β_i in an environment closer to reality. First, a random noise is added to the data generating process, i.e. $y_i = \beta_i T_i + M_i + \epsilon_i$, where $\epsilon_i \sim N(0,0.01)$. Second, in estimating the $model_1(.)$, I add a random variable $\zeta_i \sim N(0,1)$ as an additional explanatory variable which is independently generated from the data generating process of

[26] I use $h(.) = |y_i - m_c|$ as the error criteria.

p_i. The idea is that the algorithm should be able to distinguish informative variables from uninformative ones by utilising additional structural information implied by the index i.

Figure 3.B.1 gives the estimated β_i under three different settings[27]. Sub-figure (a) represents the result when I use e_i and ζ_i as the explanatory variables. As the algorithm can no longer make predictions of β_i conditioning on useful information in M_i, the predicted β_i can be very different from the true β_i. Sub-figure (c) represents the estimates after adding $N-1$ dummies of the index i. Sub-figure (d) represents the estimates when the true $\beta_i = M_i$ is added as a feature variable.

In general, the assignment of M_i will not be ordered and adding index i will not provide relevant information. Figure 3.B.2 presents estimates of β_i where M_i is randomly drawn from $\{0,1\}$ with equal probability for each individual i.

To make the correct prediction of β_d, one needs to find a transformation of the unobserved variable M_i that satisfies a weak monotonic property $\daleth$ defined below:

Definition 1. Let $\{a_n\}$ be a sequence of real numbers. $\daleth : \{a_n\} \to \{b_n\} \in \mathbb{R}^N$ is a weak monotonic transformation if

$$\begin{aligned} a_j > a_i \Rightarrow b_j > b_i \quad & \forall i,j \in \{1,...,N\} \quad or \\ a_j > a_i \Rightarrow b_j < b_i \quad & \forall i,j \in \{1,...,N\} \end{aligned}$$

Proposition 1. *Let $\{a_n\}$ be a sequence of real numbers. Suppose that entering $\{a_n\}$ as an explanatory variable in a recursive binary splitting algorithm results in $k < n$ unique splitting points $a_{m_1},...,a_{m_k}$ with m_i indicating the index for the i th split, then entering any weak monotonic transformation of $\{a_n\}$ will result in the same splitting indices $m_1,...,m_k$.*

The proof of proposition 1 is given in appendix 3.B. Intuitively, since a binary splitting algorithm only uses ordinal information of an explanatory variable in its splitting criteria, any transformation that preserves the ordinal information of this explanatory variable should result in the same splitting points. Compared to proposition 1, a more interesting and useful property of tree based algorithms is the conditional monotonic transformation property as stated below:

Proposition 2. *Let X_n be a set of feature variables excluding a_n. If $var(a|X_n) \neq 0$ for some values of X_n and there is a large number of observations for these subsets of X_n, entering $\{a_n\}$ as a feature variable is equivalent to entering any $\daleth(\{a_n\}|X_n)$ in a recursive binary splitting algorithm.*

I have not proved proposition 2 yet. I illustrate my idea using the following numerical example. Consider the case where the treatment effect β_i depends on another observable explanatory variable X_i. To keep the story simple, I assume that X_i is independently drawn

[27] The additional variable ζ_i is included in all cases.

from $\{0,1\}$ with equal probability and β_i is linear in X_i and M_i.

$$\begin{aligned}
&y_i = \beta_i T_i + M_i \\
&\beta_i = X_i + M_i \\
&T_i \in \{0,1\},\ M_i \in \{0,1\},\ X_i \in \{0,1\}
\end{aligned}$$

I experiment on the following two formulations of M_t:

$$\mathfrak{M}_d^1 = \begin{cases} -|\epsilon_i| & if\ M_i = 0 \\ |\epsilon_i| & if\ M_i = 1 \end{cases} ; \qquad \mathfrak{M}_i^2 = \begin{cases} -|\epsilon_i| + X_i & if\ M_i = 0 \\ |\epsilon_i| + X_i & if\ M_i = 1 \end{cases} ; \quad where\ \epsilon_i \sim N(0,1)$$

These two formulations are (a) highly nonlinear but (b) satisfy the weak or conditional weak monotonic transformations[28]. Estimation results are given in Figure 3.B.3 and 3.B.4. Please note that the algorithm does not know the data generating process of $\mathfrak{M}_i^1$ and $\mathfrak{M}_i^2$ and thus cannot see the clear distinction between red and green points in sub-figures (a) and (b). It classifies by recognising patterns between p_i and $\mathfrak{M}_i^1, T_i$ (or $\mathfrak{M}_i^2, T_i, X_i$ in the second case).

To sum up, tree based algorithms only use the ordinal information in its classification process. Any transformation that contains the same ordinal information of the unobserved variable will produce the same tree structure. Therefore, if one can find a variable or a set of variables that contain approximately the same ordinal information of the unobserved variable, the casual inference can be made (approximately) as if we had observed the unobserved variable.

Utilising orthogonal dimensions

In general, it is difficult to apply the conditional weak monotonic transformation property in a one dimensional data framework. However, in a multi-dimensional panel, this property can be exploited together with orthogonal variations across dimensions to control for unobserved variables. The proposed approach exploits the fact that certain dimensions are less influenced by certain types of errors. Conditioning on a particular dimension, the structural estimates may be biased. However, as long as the bias is "well structured" in other dimensions, the weak monotonic transformation property will apply and the individual treatment effect is identifiable[29]. The idea of using all possible combinations of subsets of dimensional-limited structural estimations to control for unobserved variables is first proposed in this paper.

[28] Please note that $\mathfrak{M}_d^1$ and $\mathfrak{M}_i^2$ do not satisfy the conventional monotonic transformation definition which requires $a_j \leq a_i \Leftrightarrow b_j \leq b_i \quad \forall i,j \in \{1,...,N\}$

[29] In the context of the ERPT problem, "well structured" means that the covariance between the volatility (second moment) of bilateral exchange rates and the level (first moment) of marginal cost of the firm at the time dimension does not vary across destinations. This point is explained by analytical examples in the appendix.

In the context of the constructed two-dimensional examples in section 2, the procedure of the proposed algorithm can be applied as follows. First, run simple OLS regressions using the pricing equation implied by the structural model in all possible subsets of dimensions. Second, gather these estimates from regressions and enter them as variables in a tree based algorithm to predict the dependent variable[30]. In this process, only informative coefficients on predicting the dependent variable from the first step will be selected. Third, use the obtained non-parametric model to predict the dependent variable by changing the key variable of interest and keeping other explanatory variables and the obtained coefficients in step 1 fixed. Fourth, calculate the numerical partial derivative and perform a second algorithm mapping this numerical partial derivative on observed explanatory variables and estimated structural coefficients.

1. For $t = 1...n_t$, run OLS, and collect coefficients b_t^0, b_t^1

$$p_{d,t} = b_t^0 + b_t^1 e_{d,t}$$

   For $d = 1...n_d$, run OLS, and collect coefficients b_d^0, b_d^1

$$p_{d,t} = b_d^0 + b_d^1 e_{d,t}$$

2. Approximating p. Run GBRT entering coefficients $\{b_t^0, b_t^1, b_d^0, b_d^1\}$ as additional feature variables. Obtain

$$model_1 : (e_{d,t}, ms_{d,t}, b_t^0, b_t^1, b_d^0, b_d^1) \rightarrow p_{d,t}$$

3. Numerical differentiation. Use $model_1$ to construct counterfactual predictions conditioning on the values of $ms_{d,t}, b_t^0, b_t^1, b_d^0, b_d^1$ and calculate[31]:

$$\begin{aligned}
p_{d,t}^{Est1} &= model_1(e_{d,t} - \epsilon, b_t^0, b_t^1, b_d^0, b_d^1) \\
p_{d,t}^{Est2} &= model_1(e_{d,t} + \epsilon, ms_{d,t}, b_t^0, b_t^1, b_d^0, b_d^1) \\
\beta_{d,t}^{Est} &= \frac{p_{d,t}^{Est2} - p_{d,t}^{Est1}}{2\epsilon}
\end{aligned}$$

4. Approximating β^{Est}. Run GBRT with the dependent variable $\beta_{d,t}^{Est}$ on $e_{d,t}, ms_{d,t}, b_t^0, b_t^1, b_d^0, b_d^1$, and get

$$model_2 : (e_{d,t}, ms_{d,t}, b_t^0, b_t^1, b_d^0, b_d^1) \rightarrow \beta_{d,t}^{Est}$$

[30] Unlike the fixed effect related methods, instead of partitioning out information, the proposed approach adds back these estimates to the main estimation question.

[31] Throughout my analysis, I choose ϵ to be half standard deviation of the policy variable, i.e. $\epsilon = 0.5std(e_d)$

Table 3.3 presents the estimated ERPT from applying the proposed algorithm to three examples constructed in the section 2.[32]

The algorithm is evaluated in two aspects, the ability to recover the key parameter of interest, $\beta_{d,t}$ and the ability to discover the underlying structure $\beta_{d,t}$. In contrast to conventional regression methods, the algorithm estimates β for each d and t, which generates 80,000 estimates. I construct three measures to evaluate the algorithm's ability to recover the key parameter of interest, $\beta_{d,t}$: (a) the usual absolute measure of distances defined as the sum of squared residuals, SSR; (b) the measure of the number of outliers or extreme values defined as the number of estimated β^{Est} that lies outside one standard deviation of the true β over total number of estimated β^{Est},[33]; (c) visualisation plotting the first 50 observations.

$$SSR := \sum_d \sum_t (\beta^{Est}_{d,t} - \beta_{d,t})^2$$
$$Error\ Rate := \frac{|\{\beta^{Est}_{d,t} : |\beta^{Est}_{d,t} - \beta_{d,t}| > \sigma_\beta\}|}{|\{\beta^{Est}_{d,t}\}|}$$

For evaluating the ability to recover the underlying structure of the ERPT function, I construct the following three measures. Measure 1 and 2 will enable us to compare the true relationship between ERPT and market share with the algorithm estimated relationship. Measure 3 is helpful in understanding why the algorithm estimated relationship is different from the true relationship under some circumstances.

1. The true relationship between ERPT and market share evaluated at different quantiles of the marginal cost.
   - Calculate $mc^q := quantile(mc, q)$ from data; $q \in [0.3, 0.5, 0.7]$[34]
   - Plot $f(ms) = (ms - 0.5)^2 + mc^q$
2. The estimated relationship between ERPT and market share evaluated at different quantiles of feature variables excluding market share ms, i.e. $\mathcal{X}_{-ms} := e_{d,t}, b^0_t, b^1_t, b^0_d, b^1_d$.
   - Calculate the q-th quantile of each variable in $\mathcal{X}_{-ms}$;
   - Plot $f(ms) = model_2[ms, (\mathcal{X}_{-ms})^q]$ where $(\mathcal{X}_{-ms})^q := (e_{d,t})^q, (b^0_t)^q, (b^1_t)^q, (b^0_d)^q, (b^1_d)^q$
3. The estimated relationship between ERPT and market share evaluated at the reverse engineered quantiles of the marginal cost.

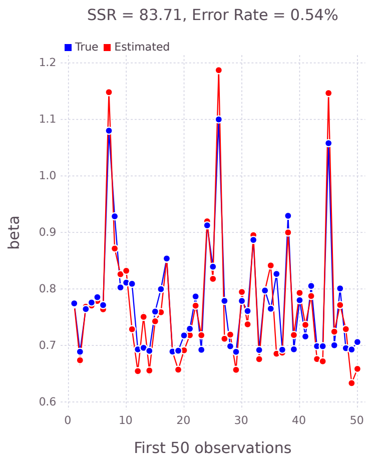
[32]In evaluating the algorithm, I construct two different datasets of the same size generated by the data generating process specified in section 2. The algorithm is first trained in one dataset. The fitted model is then tested in the second dataset.

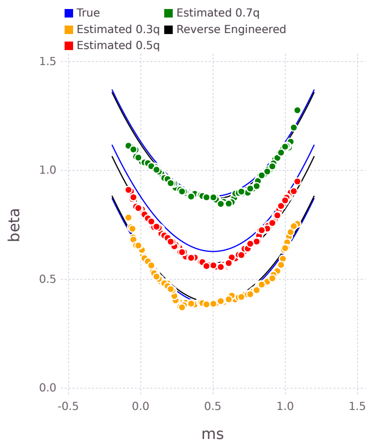
[33]In my examples, the panel is balanced, $|\{\beta^{Est}_{d,t}\}| = n_D n_T$

[34]In my initial experiments, I arbitrarily chose these three quantiles 0.3, 0.5, 0.7. A more common choice may be 0.25, 0.5, 0.75.

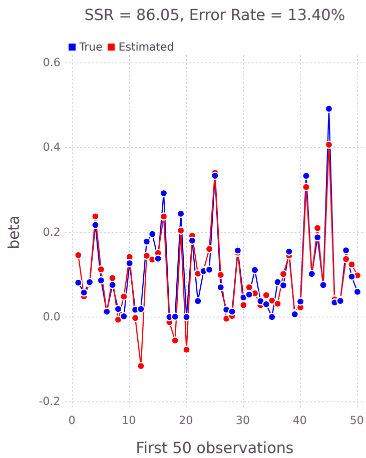
Table 3.3 Estimates of the proposed algorithm

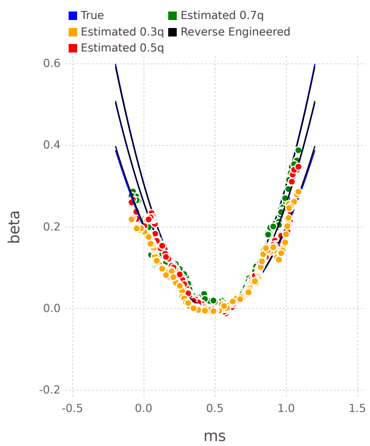
Specification 1:



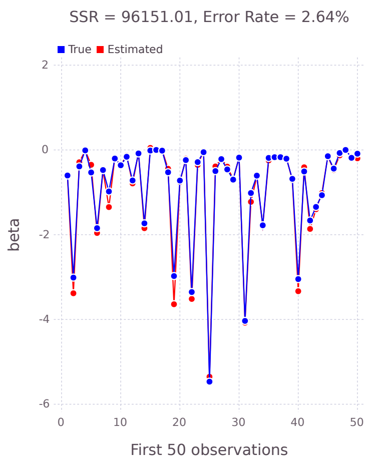


Specification 2:





Specification 3:



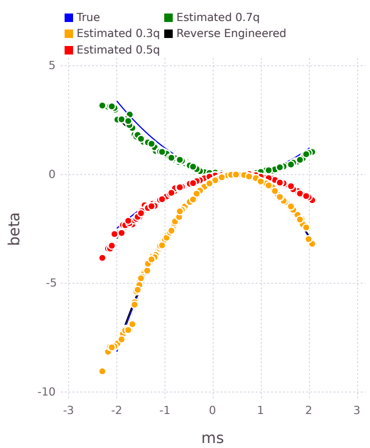
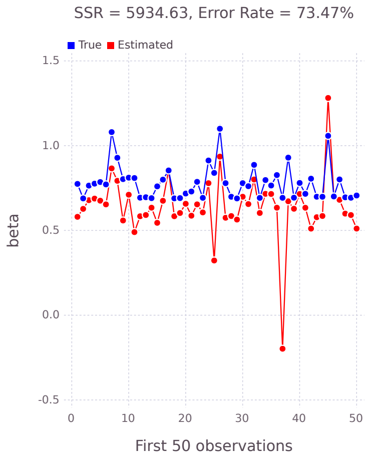
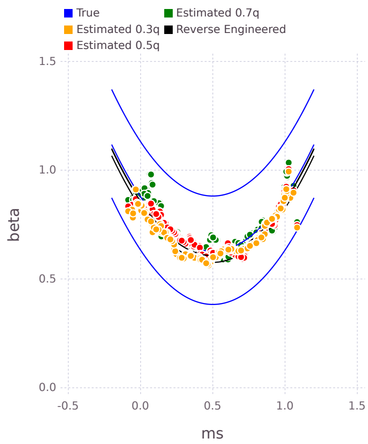


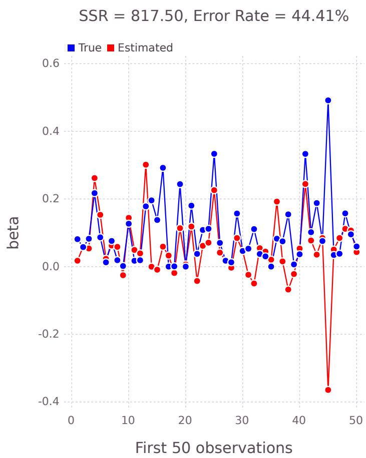
Table 3.4 No additional information

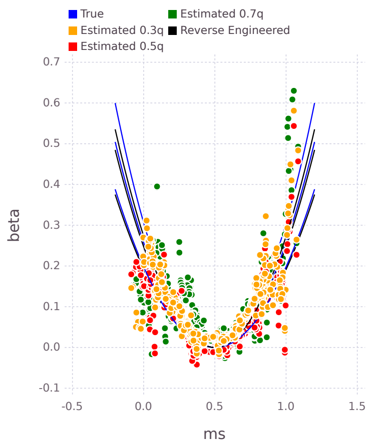
Specification 1:



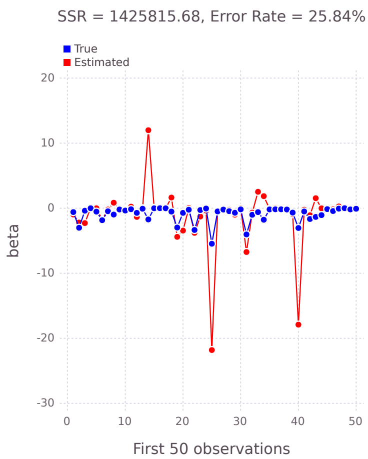


Specification 2:





Specification 3:



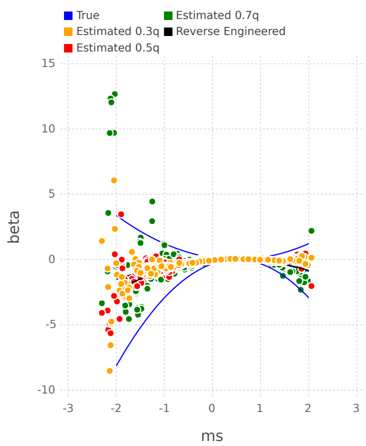
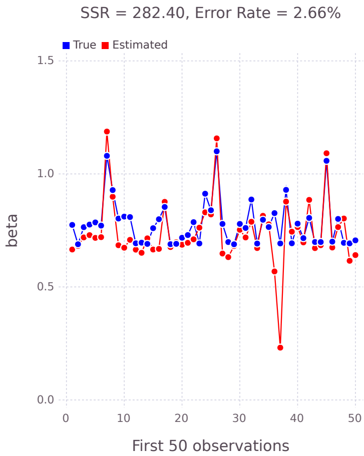
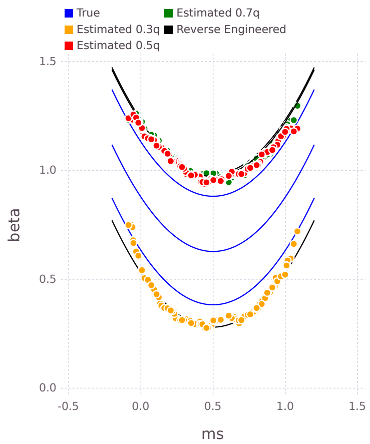


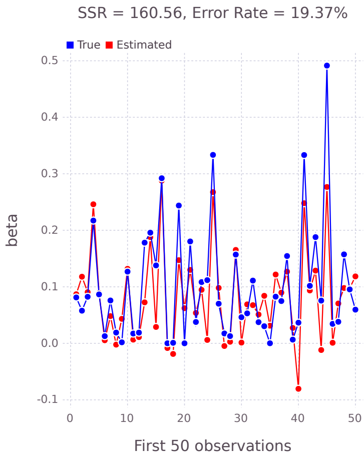
Table 3.5 Adding dimensional index

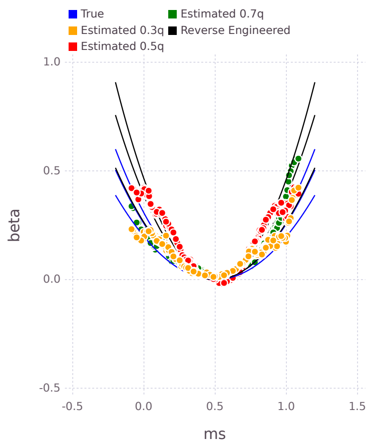
Specification 1:



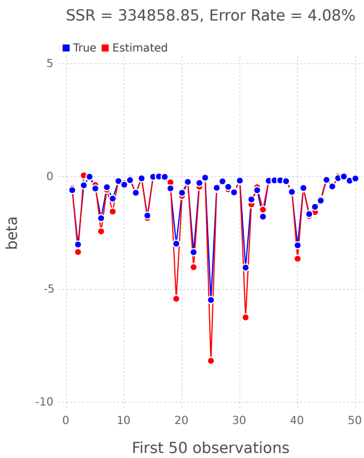


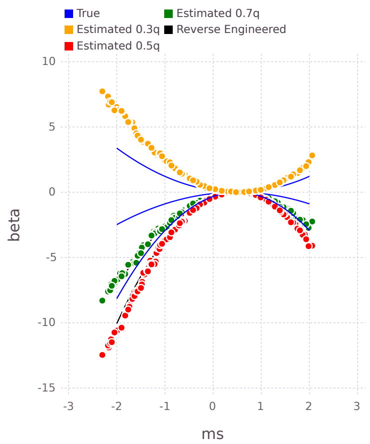
Specification 2:





Specification 3:





- Estimate mc^q implied by $(\mathcal{X}_{-ms})^q$; $q \in [0.3, 0.5, 0.7]$
  (a) Run GBRT with mc_t as the dependent variable on $\mathcal{X}_{-ms}$ and get $model_{mc}$
  (b) Estimate $mc^q = model_{mc}[(\mathcal{X}_{-ms})^q]$
- Plot $f(ms) = (ms - 0.5)^2 + mc^q$

I compare results for the proposed method with two alternative settings. Table 3.4 presents results when no additional information is added. $model_1$ will be a function mapping $(e_{d,t}, ms_{d,t}) \to p_{d,t}$ and $\mathcal{X}_{-ms} = e_{d,t}$. The estimation procedure includes step 2-4 only. Table 3.5 presents results using indices d and t as controls. $model_1$ will be a function mapping $(e_{d,t}, ms_{d,t}, d, t) \to p_{d,t}$ with $\mathcal{X}_{-ms} = (e_{d,t}, d, t)$.

Comparing results of three tables, the proposed method is significantly better at estimating $\beta_{d,t}$ and approximating the underlying structure of $\beta_{d,t}$ in all three specifications. The method without adding any additional information generates large errors in the point estimate of $\beta_{d,t}$ due to alignments of the unobserved variable mc_t. Given that, the graph on the right hand side shows that the estimated relationship represents the ERPT function evaluated at the median of the unobserved variable mc_t. Adding dimensional indices as additional feature variables will improve the accuracy of point estimates (by a smaller amount compared to the proposed approach) but does not provide additional information on the quantile of the unobserved variable mc_t. As a result, the resulting underlying structure of $\beta_{d,t}$ can be very different from the true structure.

The key to improve the estimates relies on feeding the correct additional structural information about the functional forms to the machine learning algorithm. This type of algorithm has not been explored by existing machine learning approaches because adding such structural information is not possible for prediction problems[35]. A formal presentation of the algorithm can be found in the appendix.

3.4 Model and Recovering ERPT from Simulated Exporters

The previous section tests the algorithm using simple numerical examples. This section tests the performance of the algorithm in a workhorse international macroeconomic model with heterogeneous firms.

3.4.1 Model

I take the seminal contribution of Atkeson and Burstein (2008) as the benchmark model. The model is designed to understand how strategic competition due to different market structures (productivity distributions) could reach different equilibria after an exchange rate shock. There are N countries in the world trading with each other. Within each country,

[35] They require information of the dependent variable to estimate structural coefficients.

there are a large but limited number of sectors S. As in Atkeson and Burstein (2008), these sectors can be interpreted as "the lowest level of disaggregation of commodities used in economic censuses and price index construction". Within each sector, there are a limited number of firms producing goods. Each firm produces a distinct product with the elasticity of substitution within the sector being ρ_s.

To make the model tractable, I will stick to the following two simplifications made in the original model. First, the model starts with an equilibrium and firms do not make entry and exit decisions[36]. Second, firms only use labour in their production and no imported inputs are needed.

To customise the model to fit the purpose of this paper, I extend the model in three aspects. First, I allow asymmetries in industry structures. To achieve this, I assume a large but limited number of sectors. Second, I extend the original two-country framework to an N-country trading system. This modification allows the model to study the effect of asymmetric exchange rate shocks on trade pattens. Third, to ensure a unique equilibrium in this multi-country world, I assume that only the best domestic firm in each sector exports.[37] This setting can be tough as there exists a hidden sector specific trading barrier such that only the best firm in each sector finds it profitable to export. Technically, this simplification makes this multi-firm multi-sector multi-country model stable and avoids multiple equilibria. In an N-country framework, it generates a very nice market structure with productive N-1 firms from trade partners and a bunch of domestic firms that may be less productive but large in numbers [see figure 3.D.2].

Firm's Problem

Variables in this model have five dimensions with f, s, o, d, t standing for firm, sector, origin, destination, time respectively. The final consumption $D_{d,t}$ in destination d is aggregated across sectors using the CES production function with the elasticity of substitution across sectors being equal to η. The price index for final consumption $P_{d,t}$ can be derived as follows.

$$D_{d,t} \equiv \left[\sum_s (D_{s,d,t})^{\frac{\eta-1}{\eta}}\right]^{\frac{\eta}{\eta-1}}, \quad P_{d,t} \equiv \left[\sum_s (p_{s,d,t})^{1-\eta}\right]^{\frac{1}{1-\eta}} \tag{3.2}$$

Within a sector, there are foreign firms in this sector $\mathbb{1}_{s,o}$ winning the exporting games I_E from each origin o and all domestic firms in this sector $\mathbb{1}_{s,d}$ competing together with the within-sector elasticity of substitution ρ_s. The sectoral demand $D_{s,d,t}$ and price $P_{s,d,t}$ are given by:

[36]The rationale is that firms' decisions depend on long-run sum of expectations of all future profits. As this model aims to study short-run effects of exchange rate fluctuations, this is a relatively safe condition.

[37]When the best firm is determined, it exports to all countries.

$$D_{s,d,t} \equiv \left[\sum_{o} \sum_{f \in \mathbb{1}_{s,o} \cap \mathbb{1}_E} (q_{f,s,o,d,t})^{\frac{\rho_s - 1}{\rho_s}} + \sum_{f \in \mathbb{1}_{s,d}} (q_{f,s,o,d,t})^{\frac{\rho_s - 1}{\rho_s}} \right]^{\frac{\rho_s}{\rho_s - 1}}$$

$$P_{s,d,t} \equiv \left[\sum_{o} \sum_{f \in \mathbb{1}_{s,o} \cap \mathbb{1}_E} (p_{f,s,o,d,t})^{1-\rho_s} + \sum_{f \in \mathbb{1}_{s,d}} (p_{f,s,o,d,t})^{1-\rho_s} \right]^{\frac{1}{1-\rho_s}}$$

Firms compete in quantities $q_{f,s,o,d,t}$ under Cournot competition within each sector s[38]:

$$\max_{q_{f,s,o,d,t}} q_{f,s,o,d,t}(p_{f,s,o,d,t} e_{o,d,t} - mc_{f,s,o,t})$$

subject to

$$q_{f,s,o,d,t} = \left(\frac{p_{f,s,o,d,t}}{P_{s,d,t}} \right)^{-\rho_s} \left(\frac{P_{s,d,t}}{P_{d,t}} \right)^{-\eta} D_{d,t} \tag{3.3}$$

where $mc_{f,s,o,t}$ is the marginal cost of firm f from sector s and origin o at time t.

Price, Market Share and Demand Elasticity

The optimal price $p_{f,s,o,d,t}$ for an exporter from origin o to destination d can be expressed as a function of price elasticity of demand $\varepsilon_{f,s,o,d,t}$, marginal cost $ms_{f,s,o,d,t}$ and bilateral exchange rate $e_{o,d}$, which is defined as units of currency o per unit of currency d at time t.

$$p_{f,s,o,d,t} = \frac{\varepsilon_{f,s,o,d,t}(ms_{f,s,o,d,t})}{\varepsilon_{f,s,o,d,t}(ms_{f,s,o,d,t}) - 1} \frac{mc_{f,s,o,t}}{e_{o,d,t}} \tag{3.4}$$

The price elasticity of demand $\varepsilon_{f,s,o,d,t}$ can be expressed as a function of the market share and the elasticity of substitution. Specifically, under the assumption that $\rho > \eta$, the price elasticity of demand is a strictly decreasing function of market share, i.e. bigger firms face a less elastic demand and charge a higher markup.

$$\varepsilon_{f,s,o,d,t} = \frac{1}{\frac{1}{\rho}(1 - ms_{f,s,o,d,t}) + \frac{1}{\eta} ms_{f,s,o,d,t}} \tag{3.5}$$

where market share is defined as

$$ms_{f,s,o,d,t} = \frac{p_{f,s,o,d,t} q_{f,s,o,d,t}}{\sum_f p_{f,s,o,d,t} q_{f,s,o,d,t}} = \frac{p_{f,s,o,d,t}^{1-\rho}}{\sum_f (p_{f,s,o,d,t})^{1-\rho}} \tag{3.6}$$

[38]In this nested CES structure, the main theoretical result is not sensitive to whether firms compete in prices or quantities. Atkeson and Burstein (2008) show that similar expressions can be derived if firms are competing in prices.

Substituting (9) into (8), we can express elasticity of demand as relative prices. ERPT is less than one as a decrease in $e_{d,t}$ leads to an increase in optimal price, which in turn leads to a lower market share and increases the optimal markup. A log-linearised version of the above description can be derived as follows:

Log-linearising equation (3.4), deviations of optimal price can be expressed as a function of deviations of its own market share, its own marginal cost and the bilateral exchange rate between the origin country and the destination country.

$$\widehat{p}_{k,s,o,d,t} = \kappa_{k,s,o,d,t}\widehat{ms}_{k,s,o,d,t} + \widehat{mc}_{k,s,o,t} - \widehat{e}_{o,d,t} \tag{3.7}$$

where $\kappa_{f,s,o,d,t}$ is the price elasticity with respect to a firm's own market shares, which equals the desired markup times a multiplier due to differences in elasticity of substitution across sectors and within sectors.

$$\kappa_{f,s,o,d,t} \equiv \left(\frac{\varepsilon_{f,s,o,d,t}}{\varepsilon_{f,s,o,d,t} - 1}\right)\left(-\frac{1}{\rho_s} + \frac{1}{\eta}\right) \tag{3.8}$$

Note that both $\widehat{mc}_{k,s,o,t}$ and $\widehat{e}_{o,d,t}$ are state variables and exogenous to firms. After a shock, firms reach the new equilibrium through Cournot competition. The deviation of market share $\widehat{ms}_{k,s,o,d,t}$ for firm k depends on ex ante market structure, i.e. market share distributions $\{ms_{k,s,o',d,t}\}_{k\in\mathbb{1}_f,o'\in\mathbb{1}_o}$, marginal cost shocks $\{\widehat{mc}_{k,s,o',t}\}_{k\in\mathbb{1}_f,o'\in\mathbb{1}_o}$, and the bilateral exchange rate movements of all trade partners from country d, $\{\widehat{e}_{o',d,t}\}_{o'\in\mathbb{1}_o}$.

$$\begin{aligned}
&\widehat{ms}_{k,s,o,d,t}\left[1 - (1 - ms_{k,s,o,d,t})(1-\rho_s)\kappa_{k,s,o,d,t}\right] \\
&= (1 - ms_{k,s,o,d,t})\left\{(1-\rho_s)\left[\widehat{mc}_{k,s,o,t} - \widehat{e}_{o,d,t}\right]\right\} \\
&- \sum_{o'}\sum_{f\neq k} ms_{f,s,o',d,t}\left\{(1-\rho_s)\left[\widehat{mc}_{f,s,o',t} - \widehat{e}_{o',d,t} - \kappa_{f,s,o',d,t}\widehat{ms}_{f,s,o',d,t}\right]\right\}
\end{aligned} \tag{3.9}$$

It is worth stressing that even under a firm specific shock, the equilibrium effect of changing market shares for other firms $\sum_{o'}\sum_{f\neq k} ms_{f,s,o',d,t}\kappa_{f,s,o',d,t}\widehat{ms}_{f,s,o',d,t}$ will not be zero in most cases[39]. The importance of competitors' market share reactions is weighed by the market share with its importance strictly increasing in the market share of the competitor[40]. Substituting (3.9) into (3.7), we can obtain a general equation for price deviations in a multi-country environment.

$$\widehat{p}_{k,s,o,d,t} = \lambda_{k,s,o,d,t}\left[\widehat{mc}_{k,s,o,t} - \widehat{e}_{o,d,t} - \kappa_{k,s,o,d,t}\widehat{CE}_{k,s,o,d,t}\right] \tag{3.10}$$

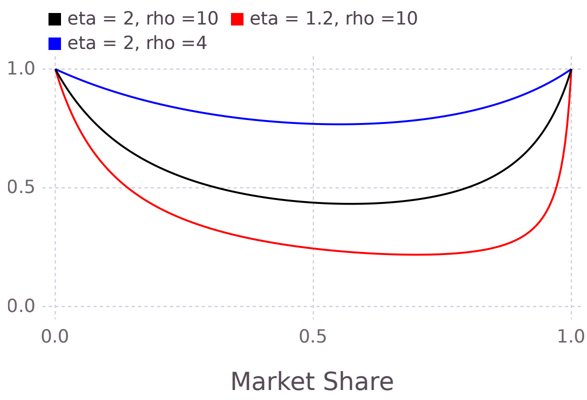
[39]In the presence of $\widehat{ms}_{f,s,o',d,t}$, there is no simple analytical solution for the optimal market share change after a shock even after log-linearisation. Given a set of realised shocks and prior market structure, market share conditions (3.9) will formulate a system of f nonlinear equations and can be solved numerically. As I will show in later simulations, reaction from other firms will make ERPT fail to present the U-shaped response in market share.

[40]Note that the expression $\kappa_{f,s,o',d,t}$ is strictly increasing in market share $ms_{f,s,o',d,t}$.

where $\lambda_{k,s,o,d,t}$ is the theoretical ERPT and it is U-shaped in market share as derived in most ERPT literature,

$$\lambda_{f,s,o,d,t} = \frac{1}{1-(1-ms_{f,s,o,d,t})(1-\rho_s)\kappa_{f,s,o,d,t}} \tag{3.11}$$

Fig. 3.1 Plot of $\lambda_{f,s,o,d,t}$



and $\widehat{CE}_{k,s,o,d,t}$ is the total effect of competitors' reactions.

$$\widehat{CE}_{k,s,o,d,t} = \sum_{o'} \sum_{f \neq k} ms_{f,s,o',d,t}(1-\rho_s) \left[\widehat{mc}_{f,s,o',t} - \widehat{e}_{o',d,t} - \kappa_{f,s,o',d,t}\widehat{ms}_{f,s,o',d,t}\right] \tag{3.12}$$

In a multi-country setting, the optimal price response of an exporter is a function of origin specific exchange rate shock minus bilateral exchange rate shocks of all other trade partners weighted by a nonlinear function of corresponding competitor's market share.

The household's problem follows closely with Atkeson and Burstein (2008). There is a representative household in each destination d maximising its expected utility by choosing optimal final consumption $C_{d,t}$ and optimal labour supply $L_{d,t}$. The representative consumer can trade a complete set of international assets from all trade partners.

$$\max_{C_{d,t}, L_{d,t}} E_0 \sum_{t=0}^{\infty} \beta^t U(C_{d,t}, L_{d,t})$$

subject to

$$U_{d,t} = log[C_{d,t}^{\mu}(1-L_{d,t})^{1-\mu}]$$

$$P_{d,t}C_{d,t} + \sum_{o} \left[\sum_{\nu} p_{o,t}^{B}(\nu)B_{o,t}(\nu) - (1+i_{o,t-1})B_{o,t-1} \right] * e_{o,d,t} = W_{d,t}L_{d,t} + \Pi_{d,t}$$

where holding $B_{o,t}(\nu)$ will earn $B_{o,t}$ unit of currency o at $t+1$ if state ν happens. $p_{o,t}^{B}(\nu)$ is the price of bond from origin o with state ν. $i_{o,t-1}$ represents the interest paid in the unit of currency o from $t-1$ to t. $\Pi_{d,t}$ is the lump-sum profit transfer from all domestic firms and exporters in country d.

The optimal solution of household's problem is given by

$$\frac{1-\mu}{\mu}\frac{C_{d,t}}{1-L_{d,t}} = \frac{W_{d,t}}{P_{d,t}} \tag{3.13}$$

$$\frac{C_{o,t}P_{o,t}}{e_{o,d,t}C_{d,t}P_{d,t}} = \frac{C_{o,t+1}(\nu)P_{o,t+1}(\nu)}{e_{o,d,t+1}(\nu)C_{d,t+1}(\nu)P_{d,t+1}(\nu)} \tag{3.14}$$

where (3.13) represents the optimal division of consumption and labor and (3.14) stands for the conventional international risk sharing condition.

Other equilibrium conditions

The production function is assumed to be linear in labour where the marginal cost $mc_{f,s,o,t}$ of a firm is calculated by dividing the nominal wage of the origin country $W_{o,t}$ by its productivity $\Omega_{f,s,o,t}$. For each firm, the total quantity of products sold $\sum_d q_{f,s,o,d,t}$ equals the quantity produced $\Omega_{f,s,o,t}l_{f,s,o,t}$. The last equation is the labor market clearing condition.

$$mc_{f,s,o,t} = \frac{W_{o,t}}{\Omega_{f,s,o,t}}$$

$$\Omega_{f,s,o,t}l_{f,s,o,t} = \sum_d q_{f,s,o,d,t}$$

$$\sum_{f,s} l_{f,s,o,t} = L_{o,t}$$

I select the nominal wage $W_{o,t}$ in each origin as the numeraire and set it equal to one. In this model, the productivity distribution can be asymmetric across sectors and countries. As a result, the bilateral nominal exchange rate is not necessarily equal to one. In my simulation the steady state bilateral exchange rate is determined by the bilateral balance of trade condition, i.e.

$$\sum_{f,s} p_{f,s,d,o,t}q_{f,s,d,o,t} = \sum_{f,s} p_{f,s,o,d,t}q_{f,s,o,d,t} * e_{o,d,t} \quad for\ o \neq d$$

3.4.2 Recovering ERPT from simulated exporters

In the following subsection, I use the model to test the proposed algorithm. Specifically, I simulate the model under different scenarios, calculate the model implied ERPT at firm level by constructing counterfactual environments, run the proposed algorithm using simulated data and compare estimated pass through with its theoretical value.

Model Simulation

I use the same calibration for the elasticity of substitution across sectors η and within sectors ρ as in Atkeson and Burstein (2008). In the benchmark case, I choose a model of three countries. The number of sectors is chosen to be 25, consistent with the classification of popular industry coding standards[41]. For a given prior productivity distribution, the number of domestic firms in each county determines the degree of home bias in the sector. For a model of three countries, I set the number of domestic firms to be 3. As a result, there will be 5 firms in each sector, including two relatively more competitive foreign firms and three domestic firms. This setting gives a reasonable median home market share around 50% depending on the productivity distribution of the sector in other countries[42]. Firm level productivity shocks are assumed to follow a simple AR(1) process with persistence equal to 0.95.

	Countries	S	N	ρ	η	$\Phi(\Omega)$
Benchmark	3	25	3+2	10	2	Uniform
Robustness	4,5	10-35	3 to 10 + Countries - 1	10	2	Uniform

To ensure the existence of a unique equilibrium in each period, I consider a financial autarky case and give exogenous exchange rate shocks to the model[43]. I further assume that the no financial market exchange rate arbitrage condition holds, i.e.

$$e_{1,2,t} = \frac{e_{1,3,t}}{e_{2,3,t}}$$

which implies a maximum of 2 exchange rate shocks in this three country world[44].

$$e_{1,2,t} = \xi_{1,t} e_{1,2,ss}, \quad e_{3,2,t} = \xi_{3,t} e_{3,2,ss}, \quad \xi_{i,t} \sim uniform(0.8, 1.2)$$

There are two sets of state variables influenced by two sets of shocks, i.e. the set of productivity shocks $\Omega_{f,s,o,t}$ for each firm, each sector and each country and the set of bilateral exchange rate shocks $e_{o,d,t}$. In each period, productivity shocks and exchange rate shocks are first realised and the corresponding values of state variables are then calculated. The most productive domestic firm in each sector wins the exporting game and exports to all

[41] Increasing the number of countries and sectors will exponentially increase the number of nonlinear equations needed to solve for each period.

[42] In this model, increasing the number of domestic firms will not always lead to a greater home bias as it will also make foreign firms surviving from the exporting game more competitive. Home firms increase in numbers but foreign competitors increase in quality. The equilibrium result depends on the assumption of productivity distributions.

[43] The international risk sharing condition (3.14) no longer applies.

[44] The third bilateral exchange rate is determined by the no arbitrage condition.

trade partners. Collecting all equilibrium conditions for all countries, solving the model is equivalent to solving a large-scale constrained system of nonlinear equations[45].

In the following exercise, I will take country 1 as the home country and try to recover ERPT of country 1's exporters. Counterfactual macro state is constructed as follows[46]

$$e^c_{1,2,t} = e_{1,2,t-1}, \quad e^c_{3,2,t} = \xi_{3,t} e_{3,2,ss}, \quad e^c_{1,3,t} = \frac{e^c_{1,2,t}}{e^c_{3,2,t}}$$

Estimation procedure

After the model is simulated, an artificial dataset is constructed to resemble those observable variables in China's custom dataset and test the proposed algorithm. The objective of the algorithm is to use only information from the constructed dataset to (a) learn from trade patterns, (b) estimate price changes under a bilateral exchange rate shock at period t given market conditions at $t-1$ and (c) recover the model implied ERPT for estimated firms.

The estimation procedure is given as follows:

1. Simulate the model for 240 periods (20 years). Record variables that are accessible from a common custom database including $f, s, o, d, t, p_{f,s,o,d,t}, q_{f,s,o,d,t}$, plus some observable macroeconomic indices including $D_{s,d,t}, P_{s,d,t}, e_{d,t}, P_{d,t}, L_{d,t}, C_{d,t}$.

2. Re-simulate the model to recover the model implied counterfactual ERPT. To calculate the model implied theoretical ERPT, I load all variables including the productivity shock from the simulated model. I then construct the counterfactual equilibrium using the productivity distribution at period t-1 and bilateral exchange rate at period t. The price difference between the counterfactual and the original equilibrium reflects the equilibrium effect of a pure exchange rate shock.

3. Identify a simple regression relationship between the dependent variable and the observable independent variables based on economic theory.

$$log(p_{f,s,d,t}) = a + b * log(e_{d,t}) + c * log(p_{f,s,d,t-1})$$

4. Identify dimensions to be fixed as $\{s, t, d, sd\}$. Run regressions and collect coefficients $a_s, b_s, c_s, a_t, b_t, c_t, ...$

5. Create market share measure $ms_{fsdt,sdt} = \frac{p_{f,s,d,t} q_{f,s,d,t}}{P_{s,d,t} D_{s,d,t}}$

[45] The setting that only the most productive firm in a sector exports avoids potential multiple equilibria and returns a unique solution in most calibrations. The model is built using Julia JUMP module and solved using Ipopt solver.

[46] The algorithm is tested on various settings of exchange rate shocks and a world with maximum 5 countries. Related results are available upon request.

6. Run GBRT with supervisor $log(p_{f,s,d,t})$ on the policy variable $log(e_{d,t})$ and the feature variables

$$\begin{aligned}\mathcal{X} = \; & log(e_{d,t-1}), log(e_{d',t}), log(e_{d',t-1}), log(ms_{fsdt-1,sdt-1}),\\ & log(D_{s,d,t-1}), log(P_{s,d,t-1}), log(P_{d,t-1}), log(L_{d,t-1}), log(C_{d,t-1}),\\ & a_s, b_s, c_s, a_t, b_t, c_t, ...\end{aligned}$$

and obtain $model_1$

7. Numerical differentiation using the predicted price at current exchange rate and the predicted price if the exchange rate was the same as in the previous period[47].

$$\begin{aligned}p^{Est1}_{f,s,d,t} &= model_1(e_{d,t}, \mathcal{X})\\ p^{Est2}_{f,s,d,t} &= model_1(e_{d,t-1}, \mathcal{X})\\ ERPT^{Est}_{f,s,d,t} &= \frac{log(p^{Est1}_{f,s,d,t}) - log(p^{Est2}_{f,s,d,t})}{e_{d,t} - e_{d,t-1}}\end{aligned}$$

8. Run GBRT again with supervisor $ERPT^{Est}_{f,s,d,t}$ on $log(e_{d,t})$, $\mathcal{X}$ and obtain $model_2$.

Results

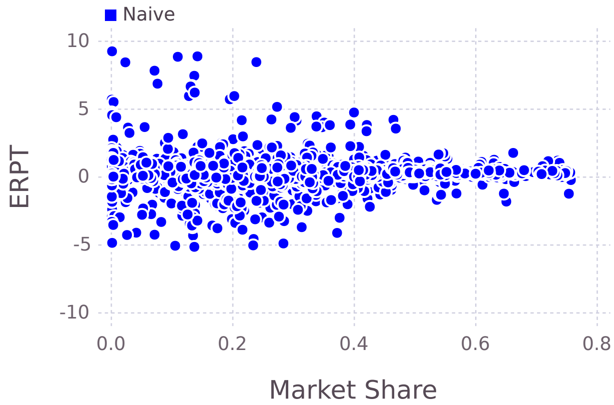
The performance of the algorithm is tested in two cases. Case 1 shuts down the idiosyncratic productivity shocks of firms in all countries[48], leaving only multilateral exchange rate shocks. Case 2 represents the world with both the idiosyncratic productivity shocks and multilateral exchange rate shocks.

In a multi-country world, multilateral rather than bilateral exchange rate movements matter. As derived in (3.9) and (3.12), the multilateral exchange rate shocks transmit into exporter's prices through the competition channel. However, controlling for the effect of multilateral exchange rate movements is not straightforward and most empirical works estimating the ERPT from the exporters' perspective only focus on bilateral movements.

[47] All other variables in $\mathcal{X}$ take their current value at time t.

[48] Firms are still different in their productivity drawn.

Fig. 3.2 Naive ERPT estimates



Note: The blue dots represent the model implied firm-level ERPT without accounting for the exchange rate movements of other trade partners. The model is simulated under case 1 where there is no productivity shock. If there is no exchange rate shock, the price $p_{f,s,1,2,t}$ will be the same across all time periods.

Ignoring this competition effect will potentially lead to seemingly unacceptable ERPT estimates. Figure 3.2 shows calculation of firm level ERPT for exporters in country 1 selling in country 2 without controlling for the exchange rate movement between country 2 and 3. In the simulated model, the price of exporters from country 1 at country 2, $p_{f,s,1,2,t}$, reacts to both $e_{1,2,t}$ and $e_{2,3,t}$. The bilateral exchange rate movements of other trade partners of the destination country could potentially magnify or mitigate the effect of the bilateral exchange rate movement from the origin country. If the ERPT is calculated without accounting for this effect, calculated results can be significantly greater than 1 or smaller than 0.

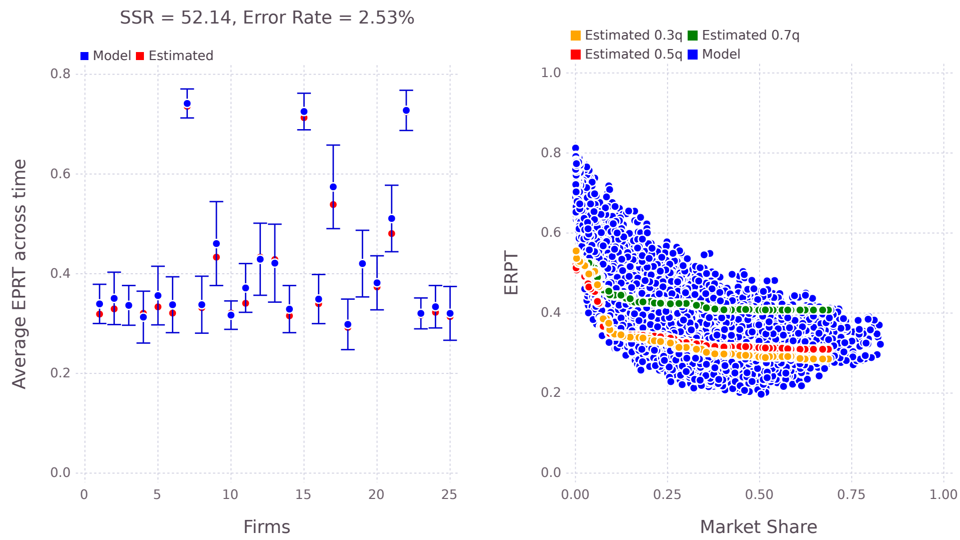
Figure 3.3 shows the results of the time-averaged estimated ERPT[49] from the proposed algorithm. The proposed algorithm performs extremely well under case 1. All time-averaged estimates lie within one standard deviation of the model implied estimate and are very close to the mean value of the model implied estimates. The error rate on point estimates is only 2.53%. The right graph shows that the estimated relationship between ERPT and market share is well aligned with the true relationship of the model implied estimates[50].

For the second case, adding productivity shocks increases the error rate. However, out of 25 firms, only two firms' time-average ERPT lie outside one standard deviation of the true

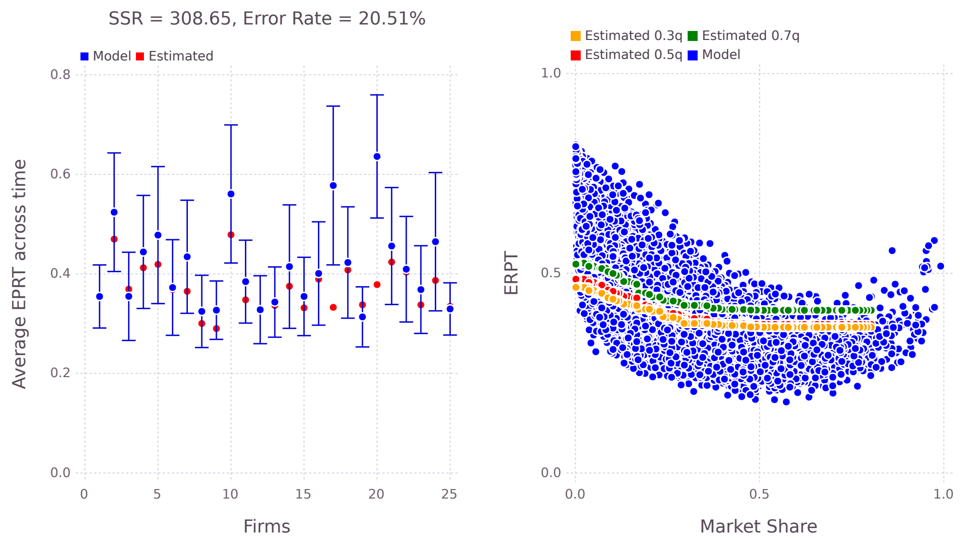
[49] i.e. $\frac{1}{n_t}\sum_t \beta^{Est}_{f,s,1,2,t}$. As only the best firm exports, the 25 firms in the figure stand for 25 sectors in the model. Graphs for detailed point estimates and the comparison with alternative methods can be found in the appendix.

[50] Note that for a given market share, the same exchange rate shock may have different impacts on each sector of an economy, which depends on two factors: (a) the underlying distribution of productivity for this particular sector of exporters from all countries and (b) the general equilibrium effect due to the change in aggregate environments of a local destination and countries exporting to this destination.

Fig. 3.3 ERPT estimates of the proposed algorithm versus model implied counterfactuals



Case 1: only exchange rate shocks



Case 2: add productivity shocks

Note: The left graph represents the time-average of ERPT for exporters originating from country 1 exporting to country 2. The x-axis of the left graph represents the index of exporters. The red dots represent estimates of the proposed algorithm. The blue dots represent the time-average of model implied firm-level ERPT backed up through counterfactual analysis. The blue bars reflect the time fluctuation of model implied ERPT for each firm. Error rate and SSR are calculated based on point estimates of ERPT for each firm-time combination (i.e. not time-averages).
The right graph represents the estimated relationship between market share and ERPT. The blue dots plot the model implied ERPT. The coloured lines provide the algorithm estimated relationship evaluated at different quantiles of the feature variables excluding the market share, $\mathcal{X}_{-ms}$.

value. The right figure is relatively weak in identifying the correct quantiles but still well aligned with the true relationship implied by the model.

3.5 Empirical Results

This section presents three empirical contributions on understanding firms' pricing behaviour. First, with the proposed algorithm, this paper presents estimates of ERPT for each firm-product-destination combination of China's exporters during the sampling period 2000-2006.[51] These estimates can be later used to construct effective exchange rate measures using a bottom-up approach based on firm-level ERPT; to identify the most and the least influenced commodity, industry and trade partner by exchange rate shocks; or to plot distributions of ERPT for different types of firms, industries and destinations, etc.

Second, this paper takes an agnostic approach to study the relationship between ERPT and various market share measures. With a four-dimensional panel (firm-product-destination-time), 12 market share measures can be constructed. Among these 12 market share measures, 9 measures are economically meaningful. Although there has been increasing attention in the trade and international literature on how different market share measures capture different aspects of firms' pricing decision and international shock transmissions, most studies work on a subset of the market share framework presented in table 3.1. My estimates contribute to the literature by assessing the relative statistical importance of market shares in explaining variations of ERPT and the unit value volatility. In addition, this paper empirically documents the nonlinear relationship between various market share measures and confirms various theoretical predictions.

[51] I use China Import and Export Custom Database funded by Cambridge Endowment for Research in Finance. Data are available at the monthly frequency from 2000 to 2006. I aggregate these monthly series into quarterly frequency to accommodate the availability of macro series such as CPI index. Details of the database and its related descriptive statistics can be found in the second chapter.

Table 3.1 Dimensions of market shares

Measure	Construction	Abbreviation	Notation
Classical Market Share	$\frac{V_{fdi}}{\sum_f V_{fdi}}$	`fdi_di`	Firm Share (DI)
Local Core Product Measure	$\frac{V_{fdi}}{\sum_i V_{fdi}}$	`fdi_fd`	Product Share (FD)
Destination Importance at Firm-product Level	$\frac{V_{fdi}}{\sum_d V_{fdi}}$	`fdi_fi`	Destination Share (FI)
Global Firm Competitiveness	$\frac{\sum_d V_{fdi}}{\sum_f \sum_d V_{fdi}}$	`fi_i`	Firm Share (I)
Global Core Product Measure	$\frac{\sum_d V_{fdi}}{\sum_d \sum_i V_{fdi}}$	`fi_f`	Product Share (F)
Destination Importance at Product Level	$\frac{\sum_f V_{fdi}}{\sum_f \sum_d V_{fdi}}$	`di_i`	Destination Share (I)
Local Firm Competitiveness	$\frac{\sum_i V_{fdi}}{\sum_f \sum_d V_{fdi}}$	`fd_d`	Firm Share (D)
Local Taste Preference	$\frac{\sum_f V_{fdi}}{\sum_f \sum_i V_{fdi}}$	`di_d`	Product Share (D)
Destination Importance within Firm	$\frac{\sum_i V_{fdi}}{\sum_d \sum_i V_{fdi}}$	`fd_f`	Destination Share (F)

Third, this paper provides the first evidence that the underlying factors explaining unit value volatility and ERPT may be different. The price volatility is strictly decreasing in all market share measures, while the relationship between ERPT and the market structure is nonlinear and varies depending on the specific share measure. In addition, increasing the volatility of bilateral and multilateral exchange rates, the volatility of destination CPI and the frequency of trade have ambiguous positive effects on unit value volatility. The effects of these variables on ERPT are heterogeneous and highly nonlinear. Interestingly, I find that both EPRT and unit value volatility are hump-shaped[52] in a number of observed trading periods.

3.5.1 Estimation procedure

The first two stages of the empirical procedure follow closely with the one introduced in section 3.4.1. In addition, I explore and estimate the factors explaining the volatility of unit values and compare them to the results on ERPT obtained in stage 2.

1. **Stage 1:**

[52] U-shaped from the importers' perspective.

(a) Identify a simple regression relationship between the dependent variable and observable independent variables based on economic theory.

$$log(p_{i,f,d,t}) = a + b * log(p_{i,f,d,t-1}) + c * log(e_{d,t})$$

(b) Identify dimensions to be fixed as $\{i, f, d, t, if, id, fd, ft\}$. Run regressions and collect coefficients $\{a_i, b_i, c_i, a_f, b_f, c_f, ...\}$

(c) Create and select combinations of market share measures.[53]

Table 3.2 Classification of market share measures

	Destination Specific	Global Counterparts
Firm	`fdi_di`	`fd_d` & `fi_i`
Product	`fdi_fd`	`fi_f` & `di_d`
Destination	`fdi_fi`	`di_i` & `fd_f`

(d) Estimate GBRT model with supervisor $log(p_{f,s,d,t})$ on the policy variable $log(e_{d,t})$ and feature variables[54]

$$\begin{aligned}\mathcal{X}_1 := \; & log(oneer_{d,t}), log(cpi_{d,t}), log(p_{i,f,d,t-1}), \\ & \texttt{fdi\_di, fdi\_fd, fdi\_fi, fd\_d, fi\_f, di\_i,} \\ & a_i, b_i, c_i, a_f, b_f, c_f, ...\end{aligned}$$

and obtain fitted $model_1$

(e) Numerical differentiation on predicted counterfactual bilateral exchange rates

$$\begin{aligned} p^{Est1}_{f,s,d,t} &= model_1(e_{d,t} + 0.5\sigma_{e_d}, \mathcal{X}_1) \\ p^{Est2}_{f,s,d,t} &= model_1(e_{d,t} - 0.5\sigma_{e_d}, \mathcal{X}_1) \\ ERPT^{Est}_{f,s,d,t} &= \frac{log(p^{Est1}_{f,s,d,t}) - log(p^{Est2}_{f,s,d,t})}{\sigma_{e_d}} \end{aligned}$$

2. **Stage 2:**
Estimate GBRT model with supervisor $ERPT^{Est}_{i,f,d,t}$ on feature variables including volatil-

[53] Table 3.2 reclassifies the 9 economically meaningful measures. As these market share measures are interdependent, one can find the minimal set of variables to represent the information of these 9 statistics. It can be shown that it is sufficient to include three destination specific measures (`fdi_di, fdi_fd, fdi_fi`) and the first column of global measures (`fd_d, fi_f, di_i`).

[54] $oneer_{d,t}$ indicates the orthogonal destination NEER which is constructed using quarterly data by the same method introduced in the second chapter.

ities of unit values and three macro price indicators (bilateral nominal exchange rates, ONEER, Destination CPI), two measures of firm-product-destination level characteristics (frequency of trade and observed trading periods), 6 market share measures and controlling coefficients.

$$\begin{aligned}\mathcal{X}_2 := \; & \sigma_{p_{i,f,d}}, \sigma_{e_d}, \sigma_{oneer_d}, \sigma_{cpi_d}, \text{Frequency of Trades}_{i,f,d}, \text{ Observed Trading Periods}_{i,f,d} \\ & \texttt{fdi\_di, fdi\_fd, fdi\_fi, fd\_d, fi\_f, di\_i,} \\ & a_i, b_i, c_i, a_f, b_f, c_f, \ldots.\end{aligned}$$

and obtain fitted $model_2$.

3. **Stage Volatility:**
Estimate GBRT model with supervisor volatility of unit values, $\sigma_{p_{i,f,d}}$, on the same set of feature variables as in stage 2 excluding $\sigma_{p_{i,f,d}}$

$$\begin{aligned}\mathcal{X}_{Vola} := \; & \sigma_{e_d}, \sigma_{oneer_d}, \sigma_{cpi_d}, \text{Frequency of Trades}_{i,f,d}, \text{ Observed Trading Periods}_{i,f,d} \\ & \texttt{fdi\_di, fdi\_fd, fdi\_fi, fd\_d, fi\_f, di\_i,} \\ & a_i, b_i, c_i, a_f, b_f, c_f, \ldots.\end{aligned}$$

and obtain fitted $model_{Vola}$.

3.5.2 Main results

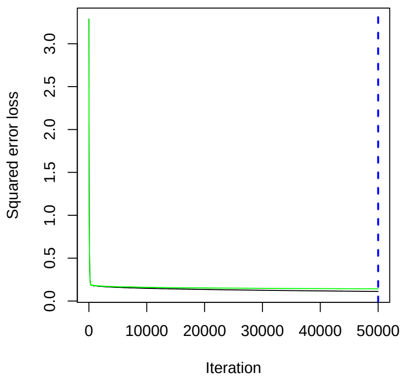
Deploying the algorithm on a real custom dataset is computationally demanding.[55] At this stage, the graphs are still sensitive to the economic equation relationship being assumed in step (a) of the first stage and the feature variables entering the first and the second stages of the algorithm. The following graphs summarise my preliminary findings.

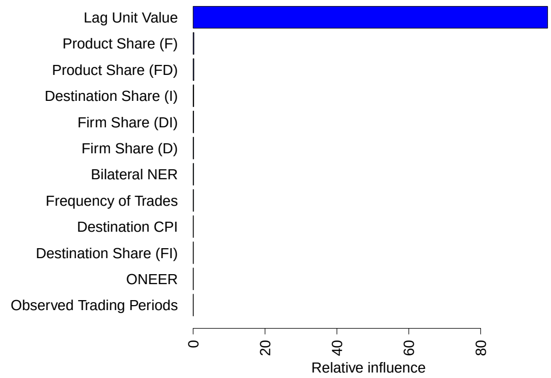
[55]The cleaned dataset has a size around 5 Gigabytes. In the proposed method, a large number of estimated structural parameters need to be stored in the memory. As a result, it currently requires around 100 times the memory of the original dataset. My codes are running on a computational cluster CamGrid [see http://help.uis.cam.ac.uk/supporting-research/research-support/camgrid/camgrid] which allows me to have maximum 128 Gigabytes memory. The following result is based on a sample of 5% randomly selected firms in the China's Custom dataset.

The second practical issue is that the amount of computational resources needed increases exponentially with the size of the dataset and the number of iterations to run. The computing time is mainly consumed in running cross validation simulations. Ideally, the optimal number of iterations needs to be determined by the cross validation simulations. By increasing the number of iterations, the within sample prediction error will always decrease but the cross validation error may or may not decrease depending on whether the additional iteration improves the fit for all parallel sub-samples. The optimal number of iteration is defined at the iteration where cross validation errors stop decreasing. However, due to computational time limits, I force the program to stop at 50,000 iterations before the optimal iteration is reached. With a 5% sample and 50,000 iterations, the program takes around 1 week to complete. As can be seen in table 3.3, the rate of the decreasing squared error loss is sufficiently low at the 50,000th iteration.

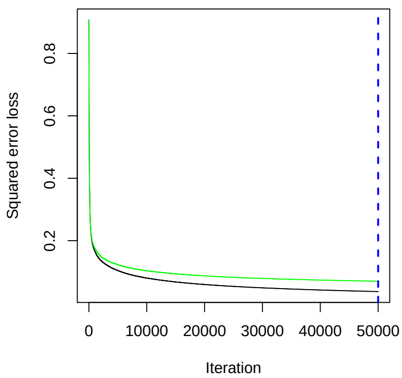
Table 3.3 Cross validation and relative importance

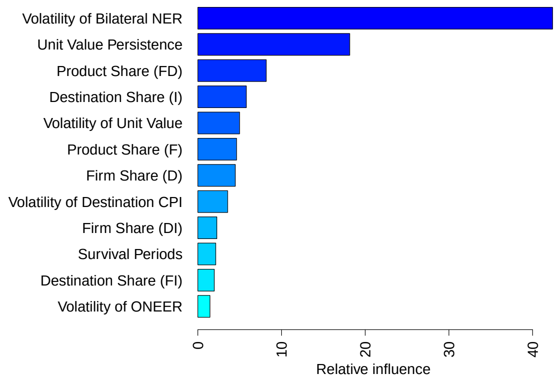
Stage 1:



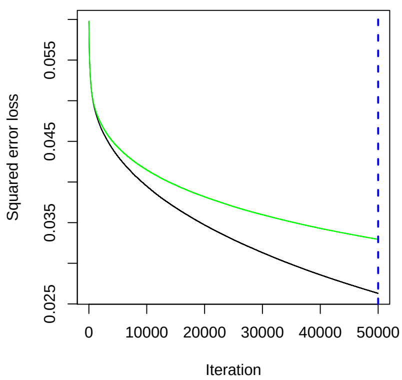


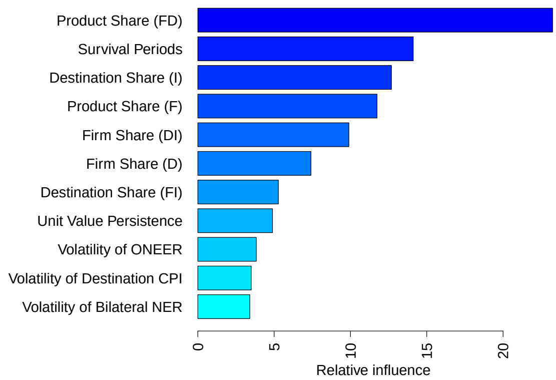
Stage 2:





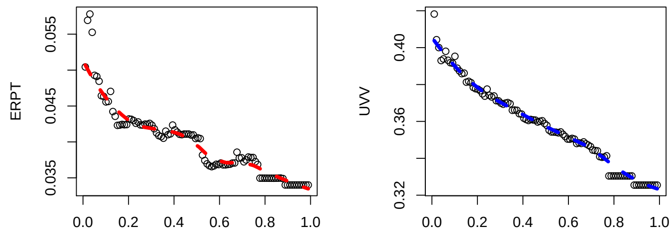
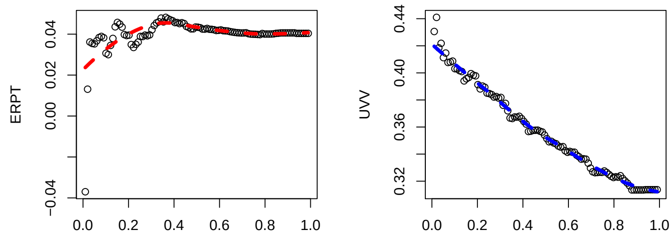
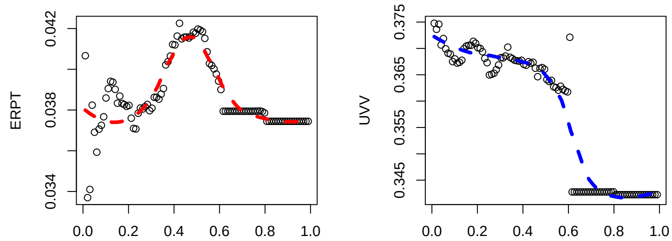
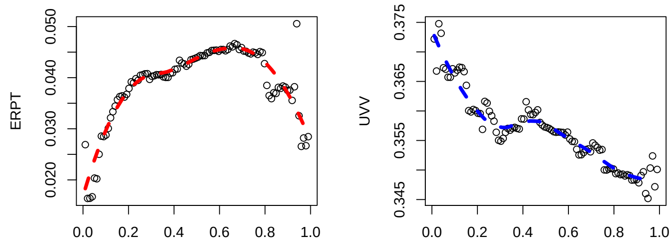
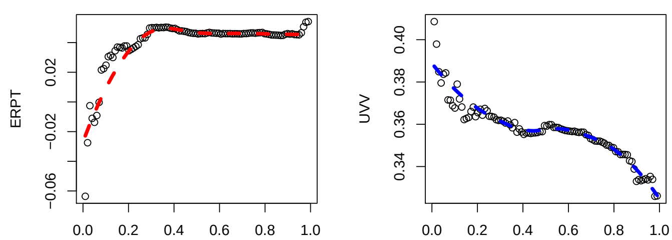
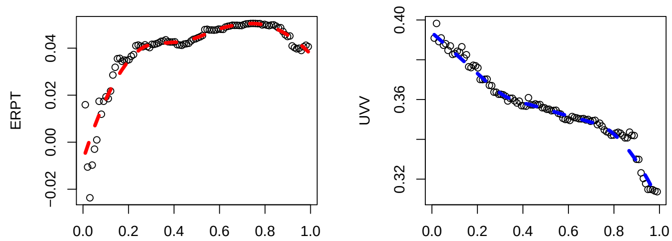
Volatility:





Note: The left panel presents cross validations of 3 models. The green and black line represent the cross-validation prediction error and within-sample prediction error respectively. The blue dashed line shows the optimal iteration indicated by cross validation errors. The right panel presents feature variables' contribution in error reduction. The supervisors in these three models are logged unit value, point estimate of ERPT and unit value volatility respectively. Unit value persistence and survival periods are measured at firm-product-destination level proxied by the frequency of trades and the number of observed trading periods respectively.

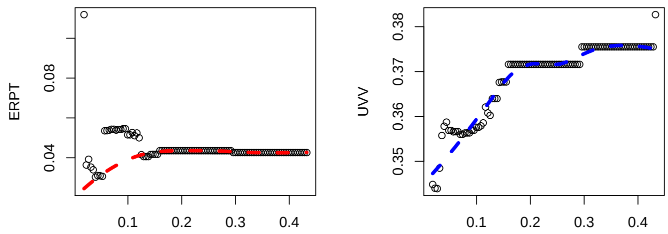
Table 3.4 Mapping firm-product-destination characteristics to ERPT (red) and unit value volatility (blue)

(fdi_di) **Classical Market Share**	
Local Core Product Measure (fdi_fd)	
Destination Importance at Firm-Product Level (fdi_fi)	
Local Firm Competitiveness (fd_d)	
Global Core Product Measure (fi_f)	
Destination Importance at Product Level (di_i)	

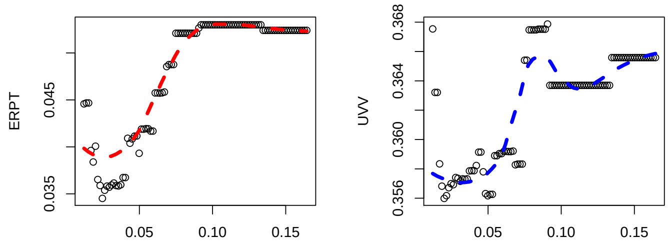
Note: The x-axis of each graph represents the percentile of market share measures, e.g. 1.0 equals 100th percentile. The circled-dots represent the estimated ERPT and unite value volatility respectively. The dashed coloured line represents the smoothed version using second order polynomials. A pass through value of 0.05 means that the RMB price goes up by 0.05% in reaction to a 1% bilateral exchange rate shock, i.e. a 95% destination country pass through. The median of the standard deviation of logged unit values at firm-product-destination level is around 0.36.

Table 3.5 Mapping firm-product-destination characteristics to ERPT (red) and unit value volatility (blue) (cont.)

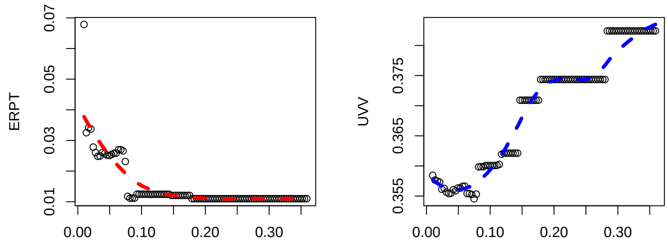
Volatility of Nominal Bilateral Exchange Rates



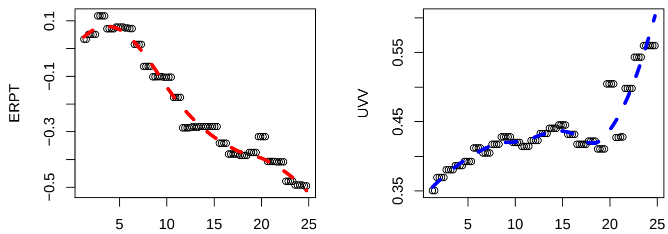
Volatility of Orthogonal Destination NEER



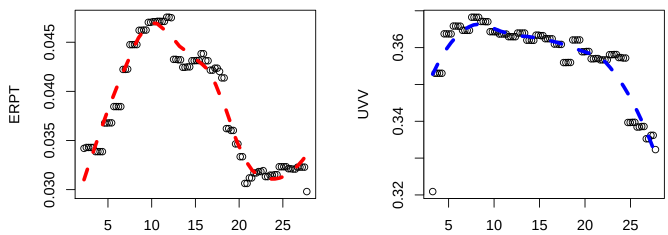
Volatility of Destination CPI



Frequency of Trades



Observed Trading Periods



Note: The x-axis of the graphs for the three volatility measures represents the standard deviation of logged macro price indicators. The x-axis of the group for the frequency of trade represents the period gap (in quarters) between two observations at the firm-product-destination level. The x-axis of the bottom two graphs represents the total number trade records observed in the sampling period at the firm-product-destination level. The circled-dots represent the estimated ERPT and unit value volatility respectively. The dashed coloured line represents the smoothed version using second order polynomials. A pass through value of 0.05 means that the RMB price goes up by 0.05% in reaction to a 1% bilateral exchange rate shock, i.e. a 95% destination country pass through. The median of the standard deviation of logged unit values at firm-product-destination level is around 0.36.

3.6 Conclusion

This paper differs from existing methodologies in emphasizing a holistic approach to estimating ERPT and proposes a machine learning algorithm to study the heterogeneity in ERPT at firm-level.

The core of the proposed algorithm consists of two elements. First, I find that the fact that tree based algorithms are robust to monotonic transformations of its feature variables can be exploited to control for unobserved components. Second, in a multi-dimensional panel, estimates from structural estimations in a range of limited-dimensional spaces can help to restrain the behaviour of unobserved components.

This paper extends Atkeson and Burstein (2008) and builds a multi-sector multi-country model to study how markets reach equilibrium under Cournot competition. From the simulated model, I construct a dataset that resembles available information in the real custom database to test the performance of the algorithm under complicated scenarios. The proposed method shows an extremely high accuracy rate on estimating firm-level ERPT and approximating the relationship between ERPT and the destination market share.

Applying the algorithm to China's custom data from 2000-2006, this paper documents new evidence on the relationships among various market share measures, firm-product-destination characteristics, unit value volatility and ERPT.

Appendix

Appendix 3.A Introduction to Classification And Regression Tree and Gradient Boosting Models

Classification And Regression Tree (CART)[56], a method of supervised learning, is a recursive binary splitting algorithm producing nonparametric mapping functions from independent variables (feature variables) to the dependent variable. Depending on the type of the dependent variables, tree based models are divided into classification trees (discrete dependent variable) and regressions trees (continuous dependent variable). Tree based methods are excellent at accommodating interactions between variables and complex nonlinear structures as well as handling outliers and missing observations. Modern decision tree algorithms are introduced by Breiman et al. (1984) and Friedman, Hastie, and Tibshirani (2001).

In a decision tree algorithm, the dataset is binary partitioned sequentially until certain stop criterion has been met. At each partition, the algorithm will search all possible splits for all feature variables and select the split that minimises the prediction error. The procedure of a basic decision tree algorithm is given as follows:

$$MPE = \sum_{\tau \in leaves(T)} \sum_{i \in \tau} h(y - m_c)$$
$$m_c = \frac{1}{n_c} \sum_{i \in \tau} p_i$$

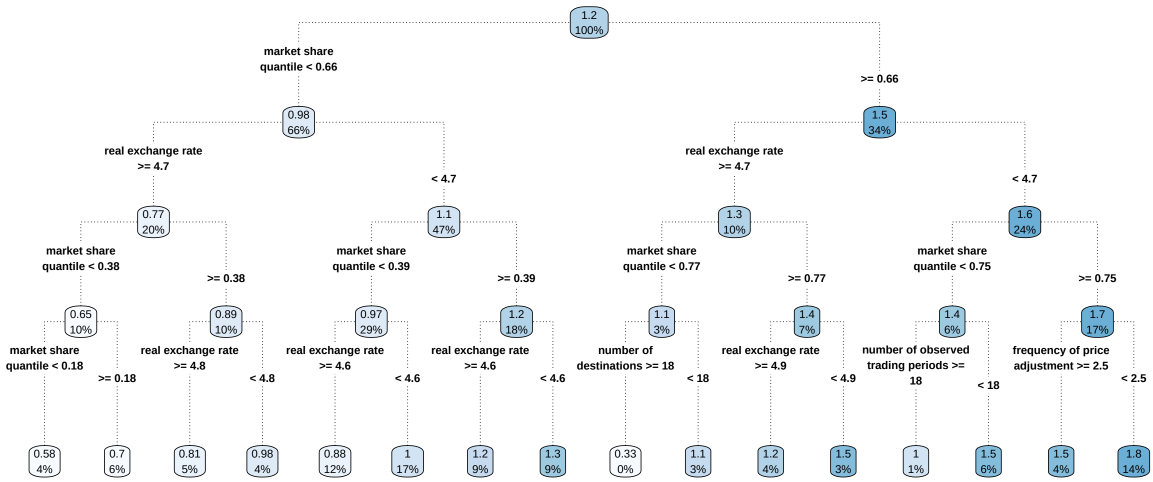
A decision tree algorithm recursively binary splits/partitions data at the point which minimises the mean prediction error (MPE) measured by criteria $h(.)$[57]. (1) The algorithm starts a tree of single node containing all points. If all the points in the node have the same value for all the input variables, stop. (2) Search over all binary splits of all variables for the one which reduces MPE as much as possible. If the largest decrease in MPE is below some

[56]A commonly used alias is "decision tree" algorithm. In my following discussions, I will refer to this algorithm as "decision tree" algorithm or tree based algorithm.

[57]Commonly used functions include $h(.) = |p_i - m_c|$ and $h(.) = (p_i - m_c)^2$.

threshold, or one of the resulting nodes contains fewer than q points, stop. Otherwise, take that split, creating two new nodes. (3) In each new node, go back to step 1.

Fig. 3.A.1 Predicting export unit values



Note: Calculation is based on quarterly data of China's import and export database 2000-2006. Unit values are measured in US dollars. The number in the circled note represents the average unit value of the classified group. The percentage below the number shows the proportion of data (counted by number of observations) located in this classification. Light (dark) blue indicates low (high) average unit values.

Figure 3.A.1 shows the results from applying CART to analyse the factors explaining the variation in export prices (unit values at firm-product level) of China's exporters. Entered feature/explanatory variables include the quantiles of market shares[58], logged real bilateral exchange rate, frequency of unit value adjustment at firm-product-destination level, number of observed trading periods and number of exporting destinations (during the period 2000-2006) at firm level. As can be seen from figure 3.A.1, the first split is made at the quantile of market shares. The algorithm predicts a higher average unit value for firms with high market shares among Chinese exporters. After the first split, several more splits are made sequentially in each subgroup based on other feature/explanatory variables. There is an interesting pattern for the last set of splits made for the left branch (market share quantile < 0.66) and the right branch (market share quantile ≥ 0.66). The left branch suggests that the unit value variation is mostly explained by market share and real exchange rates variations for those firm-product combinations with small market shares, whereas the right branch suggests that other firm-product characteristics, such as the number of exported destinations

[58] I use the classical market share measure at firm-product-destination level $\equiv \frac{\sum_t V_{i,f,d,t}}{\sum_f \sum_t V_{i,f,d,t}}$ calculated among China's exporters.

and the frequency of price adjustments, start to play a role after the first few splits for those firm-product combinations with large market shares[59].

There are three advantages of tree based algorithms.

First, the binary splitting rule represents a natural decision-making process and the resulting tree structure is easy to understand and interpret.

Second, the recursive binary splitting feature makes decision tree methods a natural nonlinear estimator. Interactions between variables are accounted from the sequential feature of the partition process as the next partition depends on the previous partitions being made. "Trees tend to work well for problems where there are important nonlinearities and interactions." Tree based algorithms can discover nonlinear patterns that conventional econometric methods may fail to detect. More discussions can be found in Varian (2014).

Third, tree based models are robust to certain types of outliers and irregularities of data. Due to the binary splitting structure, only ordinal information of explanatory variables is used. Therefore, the resulting tree structure is robust to monotonic transformation of the explanatory variables. As discussed in subsection 3.3.2, this property can be exploited to control for unobserved variables.

Given these advantages, empirically applying a decision tree algorithm also has various problems. First, finding the optimal decision tree in a large dataset is computationally difficult[60]. Second, practical decision tree solutions often lead to a local rather than global optimisation. Third, the algorithm is sensitive to small changes of the ordinal structure of explanatory variables. The resulted tree structure is often sensitive to its initial splits. More details and discussions of tree based algorithms can be found in Rokach and Maimon (2005).

The above problems can be overcome by various machine learning techniques including bagging, stacking, model averaging, random forest and boosting, etc.[61] The gradient boosting model, introduced by Friedman (2002)[62], is one of the most effective algorithms.

The boosting algorithm is based on the idea that adaptively integrating many small models can achieve and even outperform the predictive power of a single big model. Gradient boosting regression tree (GBRT) algorithm combines elements of gradient boosting and decision tree algorithms. In GBRT, trees are grown sequentially: each tree is grown conditional on the classification from previously grown trees. Adding the boosting procedure makes tree based models more robust, less path dependent and easy to work with large datasets.

The procedure of a workhorse GBRT algorithm is given as follows. A GBRT algorithm is a numerical optimisation technique with the objective to find the mapping $f(\mathbf{x})$ to minimise

[59] Please note that these results represent statistical relationships between variables only. As these classifications are not conditional on the characteristics of firms, products, destination competition environments, no further economic inference should be made based on these results.

[60] There have been papers proving that finding an optimal decision tree from a given data is NP-hard or NP-complete under different scenarios. See Hyafil and Rivest (1976) and Hancock et al. (1996)

[61] See Breiman (1996) and Breiman (2001)

[62] Freund and Schapire (1996) developed the first two-class boosting classification algorithm called AdaBoost.

the expected loss function Ψ by sequentially adding a new tree that best reduces the gradient of the loss function:

$$\widehat{f}(\mathbf{x}) = arg \min_{f(\mathbf{x})} E_{y,\mathbf{x}} \Psi(y, f(\mathbf{x})) \tag{3.15}$$

The algorithm starts by initialising $\widehat{f}(\mathbf{x})$ to be a constant and iterating the following steps until reaching the specified $Iter_{max}$.

1. Compute the negative gradient as the working response

$$h_i = -\frac{\partial}{\partial f(\mathbf{x_i})} \Psi(p_i, f(\mathbf{x_i})) \Big|_{f(\mathbf{x_i}) = \widehat{f}(\mathbf{x_i})} \tag{3.16}$$

2. Randomly select a fraction `bf` from the dataset (Random Forest/Bagging)

3. Fit a regression tree with `inter.depth` splits , $g(\mathbf{x})$, predicting h_i from the covariates $\mathbf{x_i}$.

4. Update the estimate of $f(\mathbf{x})$ as

$$\widehat{f}(\mathbf{x}) \rightarrow \widehat{f}(\mathbf{x}) + \mathtt{lr} * g(\mathbf{x}) \tag{3.17}$$

5. Repeat step 1-4 until $Iter_{max}$

	Φ	`iter`	`inter.depth`	`lr`	`bf`
Benchmark	Normal	Cross Validation	8	0.01	0.5
Robustness	-	50,000	1-10	0.005, 0.001	0.3, 1

A GBRT model is calibrated with four parameters.[63] First, the bagging fraction, `bf`.[64] Second, a parameter controls the depth of interactions between variables, `inter.depth`. Third, the shrinkage or the learning rate, `lr` controls the weight of each iteration and a higher value means a quicker convergence rate. Fourth, the distribution Φ of the error term defines the loss function Ψ.[65]

The optimal number of iterations is often selected by the cross validation process. The model is first run with large number of iterations and the best iteration `iter` is then selected with k-fold cross validations.

[63] Elith et al. (2008) provide a good introduction for modeling tuning practices. Ridgeway (2007) provides a good guidance on modeling tuning for the gbm package in R.

[64] This parameter helps to ensure the robustness of the model and prevents overfitting. The conventional value is $0.3 - 0.5$.

[65] For example, Gaussian implies a squared error loss function.

It is worth noticing that the difference in ideology of modeling between machine learning and economic models. Economic and most econometric models start with structural assumptions reflecting economists understanding of how the world operates. Machine learning models, on the other side, assume that the true data-generating process is infinitely complex and all variables in the model are correlated in a nonlinear manner. Machine learning approaches try to maximally recover the ground truth by appointing a learning algorithm (a classier) to learn the relationship between variables. More specifically, the objective of a learning algorithm is to recover patterns among variables with the performance evaluated by prediction/classification errors in a given dataset. The cost of such non-parametric ideology of machine learning approaches is data driven, i.e. the ability of an algorithm to describe the ground truth of the world depends critically on the quality of data being supplied. If an important variable is not observed in the dataset, conventional machine learning approaches fail to capture the information in this variable and the resulting model is less satisfactory. In this aspect, it is worth designing an approach to integrate structural economic models and machine learning algorithms.

Appendix 3.B Proof of Proposition 1 and Simulations of Example 1

Proof of Proposition 1. The first split is made at the point m_1. This implies

$$\begin{aligned}
&\frac{1}{|\{i : a_i \leq a_{m_1}\}|} \sum_{i:a_i \leq a_{m_1}} g[f(a_i) - \frac{\sum_{i:a_i \leq a_{m_1}} f(a_i)}{|\{i : a_i \leq a_{m_1}\}|}] + \frac{1}{|\{i : a_i > a_{m_1}\}|} \sum_{i:a_i > a_{m_1}} g[f(a_i) - \frac{\sum_{i:a_i > a_{m_1}} f(a_i)}{|\{i : a_i > a_{m_1}\}|}] \\
&< \frac{1}{|\{i : a_i \leq a_q\}|} \sum_{i:a_i \leq a_q} g[f(a_i) - \frac{\sum_{i:a_i \leq a_q} f(a_i)}{|\{i : a_i \leq a_q\}|}] + \frac{1}{|\{i : a_i > a_q\}|} \sum_{i:a_i > a_q} g[f(a_i) - \frac{\sum_{i:a_i > a_q} f(a_i)}{|\{i : a_i > a_q\}|}] \quad \forall q \neq m_1
\end{aligned} \tag{3.18}$$

where $g(.)$ is a loss function. Let $\{b_n\} = \urcorner(\{a_n\})$.

$a_j > a_i \Rightarrow b_j > b_i \quad \forall i,j \in \{1,...,N\}$ implies

$$\{i : a_i < a_j\} \subseteq \{i : b_i < b_j\} \quad and \quad \{i : a_i > a_j\} \subseteq \{i : b_i > b_j\} \quad \forall j \in \{1,...,N\} \tag{3.19}$$

Suppose that the transformation $\{b_n\}$ falls into the first category that $a_j > a_i \Rightarrow b_j > b_i \quad \forall i,j \in \{1,...,N\}$ and the first optimal splitting point of $\{b_n\}$ is $b_{m_1^*}$. As the splitting criterion only uses the ordinal information, it can be written as

$$\begin{aligned}
&\frac{1}{|\{i : b_i \leq b_{m_1^*}\}|} \sum_{i:b_i \leq b_{m_1^*}} g[f(a_i) - \frac{\sum_{i:b_i \leq b_{m_1^*}} f(a_i)}{|\{i : b_i \leq b_{m_1^*}\}|}] + \frac{1}{|\{i : b_i > b_{m_1^*}\}|} \sum_{i:b_i > b_{m_1^*}} g[f(a_i) - \frac{\sum_{i:b_i > b_{m_1^*}} f(a_i)}{|\{i : b_i > b_{m_1^*}\}|}] \\
&< \frac{1}{|\{i : b_i \leq b_q\}|} \sum_{i:b_i \leq b_q} g[f(a_i) - \frac{\sum_{i:b_i \leq b_q} f(a_i)}{|\{i : b_i \leq b_q\}|}] + \frac{1}{|\{i : b_i > b_q\}|} \sum_{i:b_i > b_q} g[f(a_i) - \frac{\sum_{i:b_i > b_q} f(a_i)}{|\{i : b_i > b_q\}|}] \quad \forall q \neq m_1^*
\end{aligned} \tag{3.20}$$

I want to prove $m_1 = m_1^*$.
(3.18) and (3.20) imply

$$\{i : a_i < a_{m_1}\} = \{i : b_i < b_{m_1^*}\} \tag{3.21}$$

$$\{i : a_i > a_{m_1}\} = \{i : b_i > b_{m_1^*}\} \tag{3.22}$$

(3.19) implies

$$\{i : a_i < a_{m_1}\} \subseteq \{i : b_i < b_{m_1}\} \quad and \quad \{i : a_i < a_{m_1^*}\} \subseteq \{i : b_i < b_{m_1^*}\} \tag{3.23}$$

$$\{i : a_i > a_{m_1}\} \subseteq \{i : b_i > b_{m_1}\} \quad and \quad \{i : a_i > a_{m_1^*}\} \subseteq \{i : b_i > b_{m_1^*}\} \tag{3.24}$$

(3.21) and (3.23), and (3.22) and (3.24) imply

$$\{i : a_i < a_{m_1^*}\} \subseteq \{i : b_i < b_{m_1^*}\} = \{i : a_i < a_{m_1}\} \subseteq \{i : b_i < b_{m_1}\} \tag{3.25}$$
$$\{i : a_i > a_{m_1^*}\} \subseteq \{i : b_i > b_{m_1^*}\} = \{i : a_i > a_{m_1}\} \subseteq \{i : b_i > b_{m_1}\} \tag{3.26}$$

From which, it can be derived that $a_{m_1} = a_{m_1^*}$ and $b_{m_1} = b_{m_1^*}$. Because

$$\{i : a_i < a_{m_1^*}\} \subseteq \{i : a_i < a_{m_1}\} \Rightarrow a_{m_1^*} \leq a_{m_1} \tag{3.27}$$
$$\{i : a_i > a_{m_1^*}\} \subseteq \{i : a_i > a_{m_1}\} \Rightarrow a_{m_1^*} \geq a_{m_1} \tag{3.28}$$

(3.25) and (3.26) can be simplified as

$$\{i : a_i < a_{m_1^*}\} = \{i : b_i < b_{m_1^*}\} = \{i : a_i < a_{m_1}\} = \{i : b_i < b_{m_1}\} \tag{3.29}$$
$$\{i : a_i > a_{m_1^*}\} = \{i : b_i > b_{m_1^*}\} = \{i : a_i > a_{m_1}\} = \{i : b_i < b_{m_1}\} \tag{3.30}$$

which implies

$$\{i : a_i = a_{m_1}\} = \{i : a_i = a_{m_1^*}\} \tag{3.31}$$

By the uniqueness of m_1, we have (from 3.18)

$$\{i : a_i > a_{m_1}\} \neq \{i : a_i > a_q\} \quad \forall q \in \{1, ..., N\} \neq m_1 \tag{3.32}$$

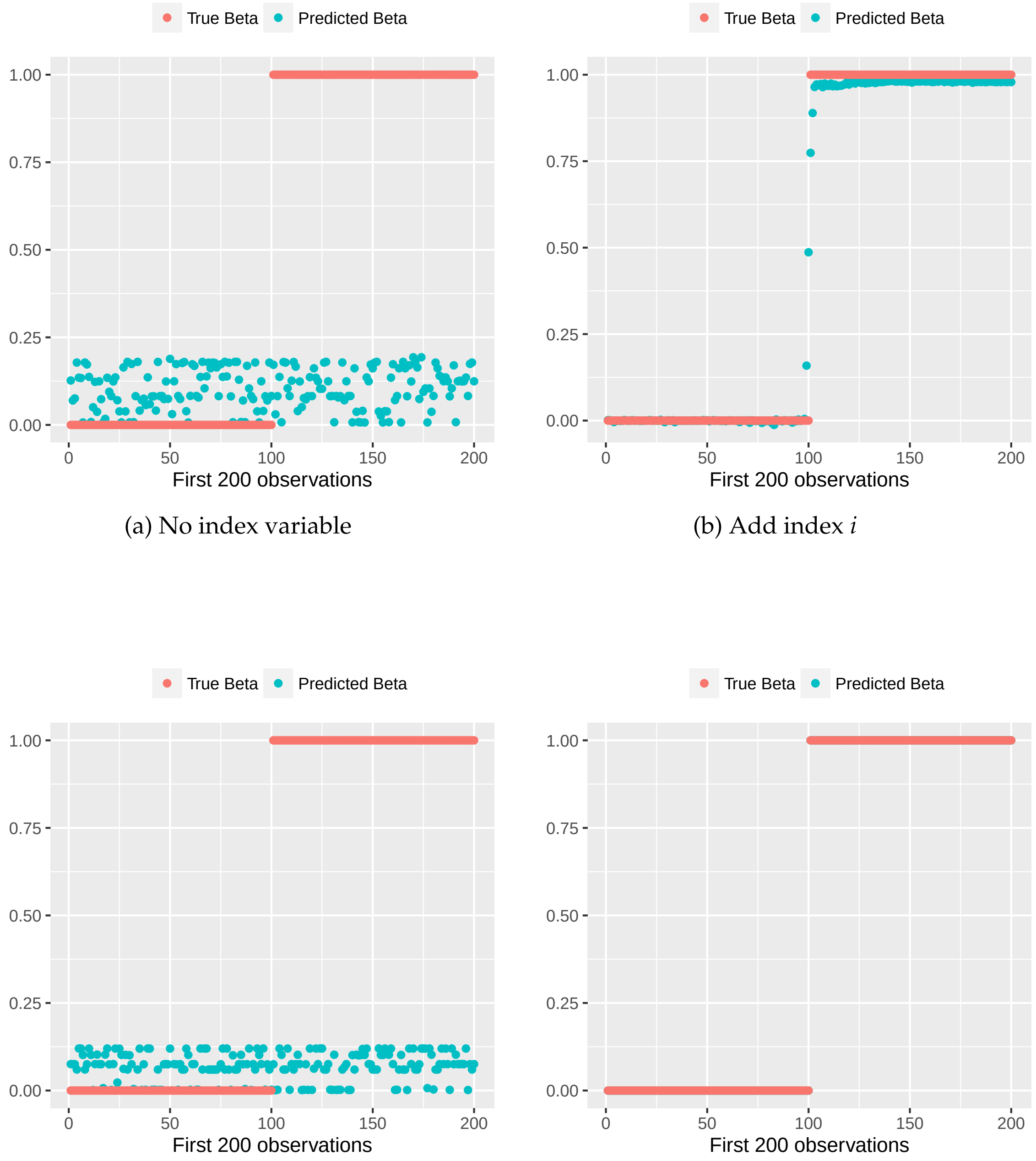
Therefore,

$$m_1 = m_1^* \tag{3.33}$$

The case where the transformation $\{b_n\}$ falls into the second category that $a_j > a_i \Rightarrow b_j < b_i \quad \forall i, j \in \{1, ..., N\}$ can be proved following a similar procedure. Recursively applying this argument to the rest of splitting points $a_{m_2}, ..., a_{m_k}$ completes the proof.

Remark. Intuitively, since a binary splitting algorithm only uses ordinal information of an explanatory variable in its splitting criteria, any transformation that preserves the ordinal information of this explanatory variable should result in the same splitting points.

Fig. 3.B.1 Simulation of example 1: the ordered case



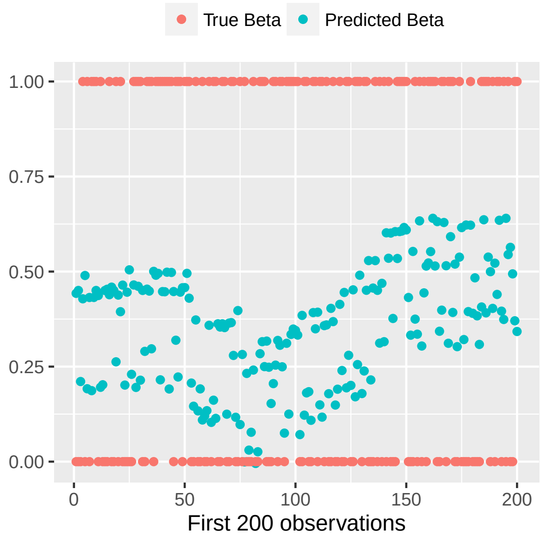
(a) No index variable

(b) Add index i

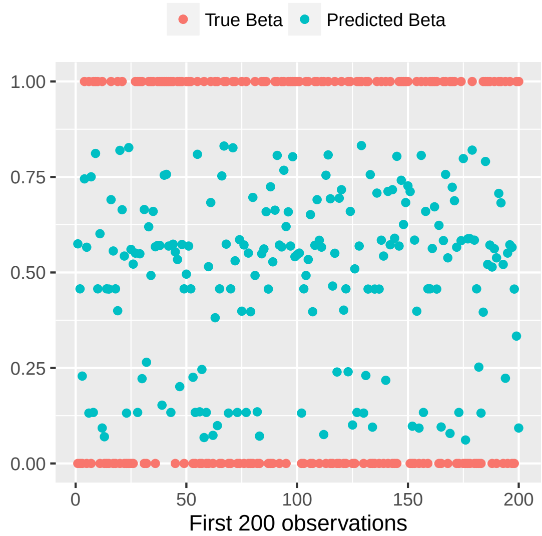
(c) Add dummy variables of i

(d) Add true β_d

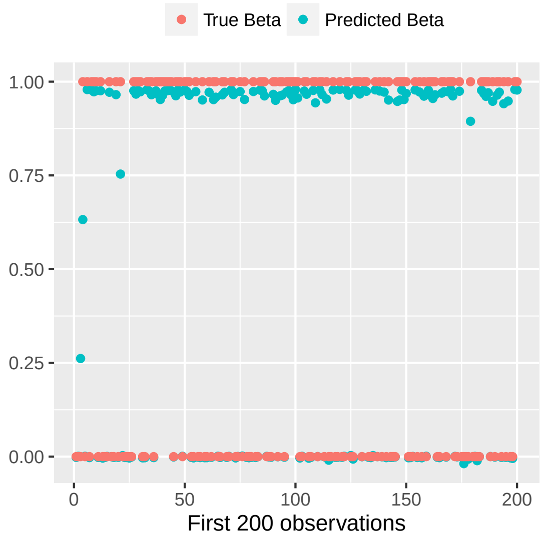
Fig. 3.B.2 Simulation of example 1: the randomly assigned case



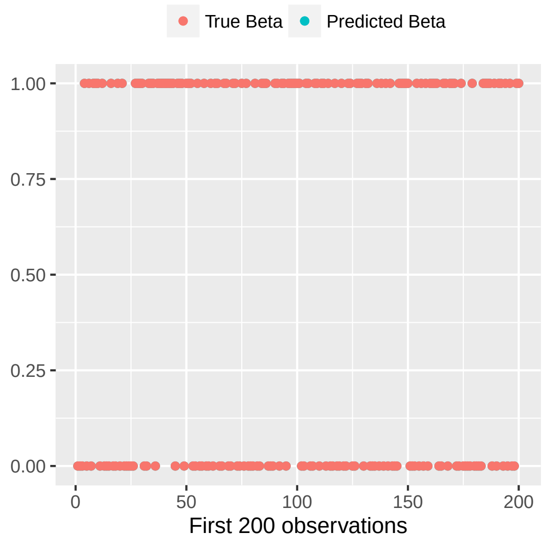
(a) Add index variable i



(b) Add dummy variables of i



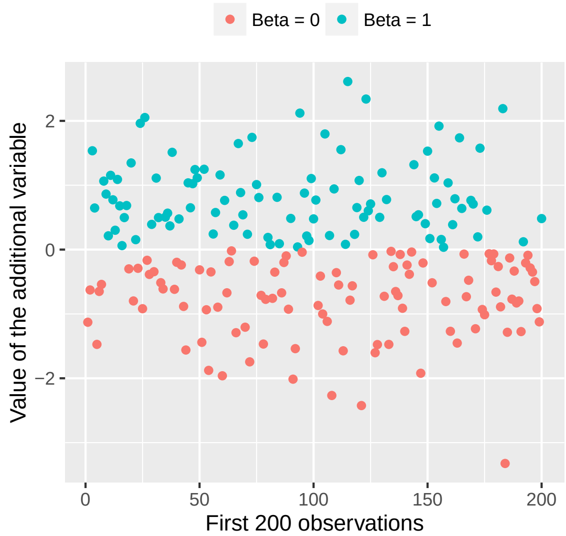
(c) Add $\mathfrak{M}_i$



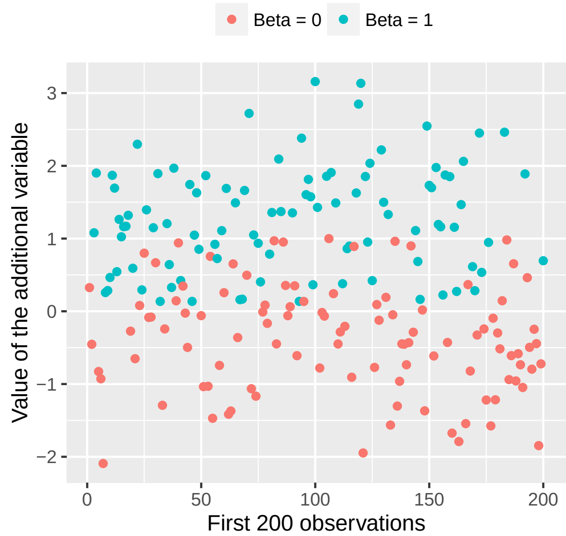
(d) Add true β_d

where $\mathfrak{M}_i = -|\epsilon|$ if $\beta_d = 0$ and $\mathfrak{M}_i = |\epsilon|$ if $\beta_d = 1$ $\epsilon; \sim N(0,1)$

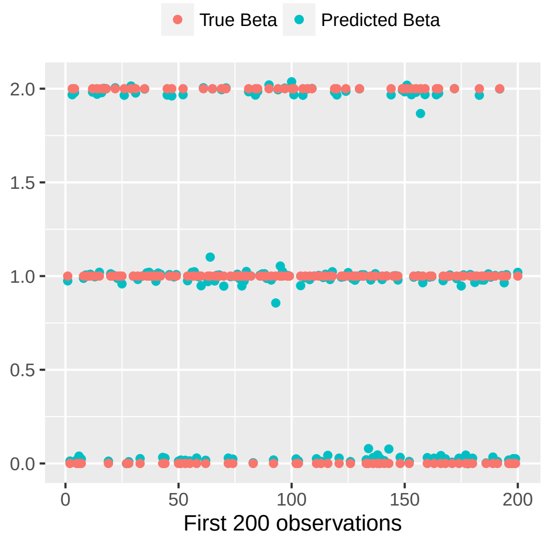
Fig. 3.B.3 Simulation of example 1: the property of weak monotonic transformation



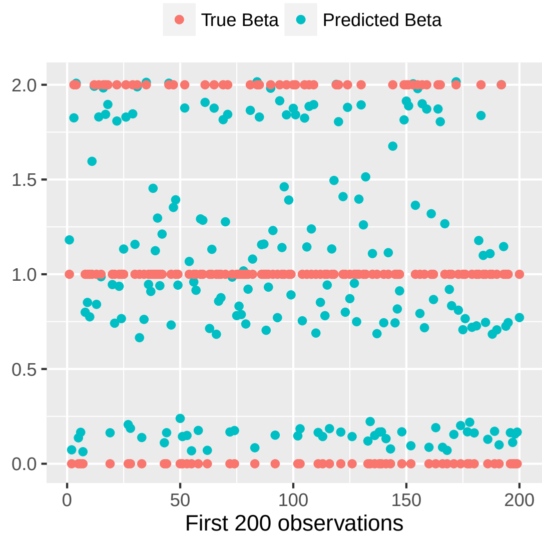
(a) Value of $\mathfrak{M}_d^1$



(b) Value of $\mathfrak{M}_i^2$



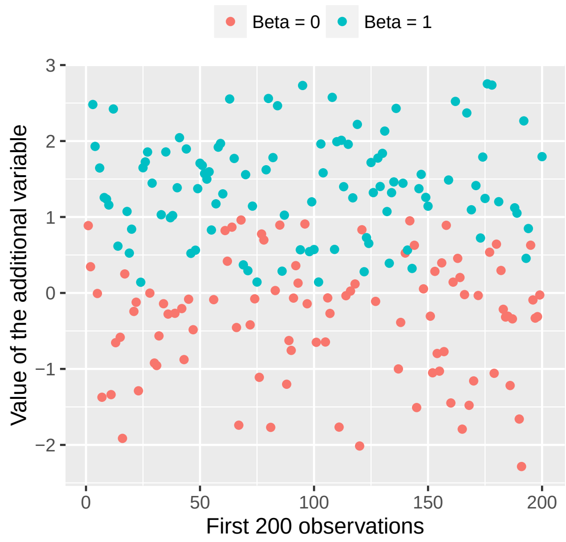
(c) Add $\mathfrak{M}_d^1$ as control



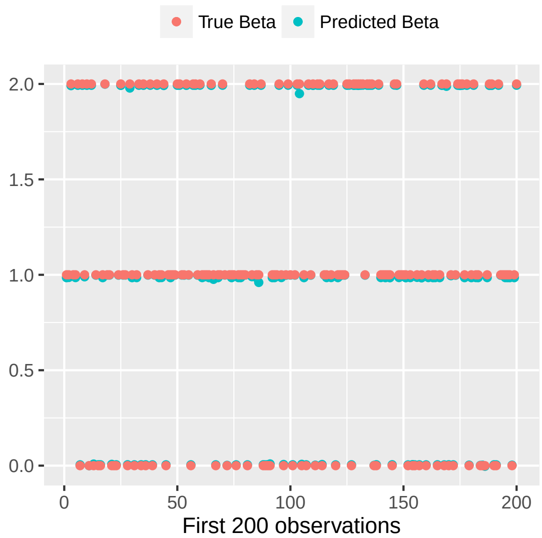
(d) Add $\mathfrak{M}_i^2$ as control

Fig. 3.B.4 Simulation of example 1: the property of weak monotonic transformation; increase n to 2,000

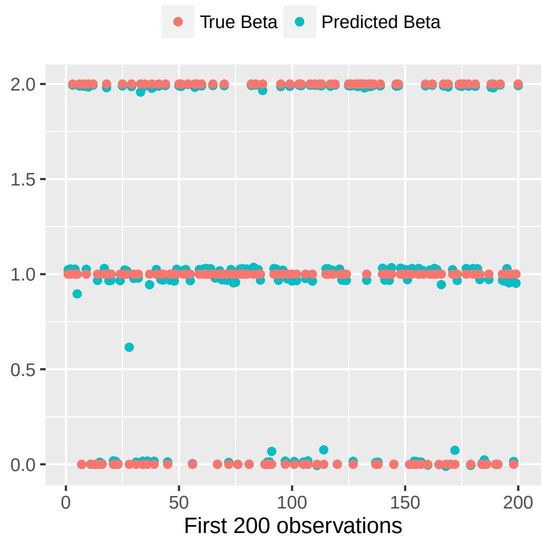
(a) Value of $\mathfrak{M}_d^1$



(b) Value of $\mathfrak{M}_i^2$



(c) Add $\mathfrak{M}_d^1$ as control



(d) Add $\mathfrak{M}_i^2$ as control

Appendix 3.C An Analytical Discussion of Simple Cases of the Two Dimensional Example

The key issue here is to understand how and when adding estimated parameters from structural estimations can help to control for the unobserved variables and estimate $\beta_{d,t}$. In the following subsection, I will discuss a simple case where the unobserved variable does not appear in the outer part of the equation of $p_{d,t}$.[66] For the data generating process given in equation (3.34) and (3.35), I want to estimate how the price $p_{d,t}$ changes under an exchange rate $e_{d,t}$ shock conditioning on values of M_t and $X_{d,t}$. However, M_t is unobserved. The objective is to understand how and when conditioning on values of coefficients $b_t^0, b_t^1, b_d^0, b_d^1$ and $X_{d,t}$ could achieve the same result as conditioning on values of M_t and $X_{d,t}$.

$$p_{d,t} = \beta_{d,t} e_{d,t} \tag{3.34}$$

$$\beta_{d,t} = f(X_{d,t}, M_t) \tag{3.35}$$

In the following discussion, I will try to express the true $\beta_{d,t}$ as a function of regression estimated coefficients.[67] In order to be clear about the information contained in estimated regression coefficients, I take first order approximations to decompose the data generating process into factors that only vary in one dimension. With this approximation, I will be able to obtain analytical solutions expressing values of b_t^1 and b_d^1 as $\beta_{d,t}$.[68] The gain in efficiency by adding coefficients versus using the indices will depend on the complexity of the hidden function implied by estimated coefficients versus the complexity of the hidden function implied by indices.

The coefficients of obtained from step 1 can be written as:

$$b_t^1 = \sum_i \beta_{d,t} w_{d,t} \quad b_d^1 = \sum_t \beta_{d,t} \omega_{d,t} \tag{3.36}$$

$$w_{d,t} := \frac{e_{d,t}(e_{d,t} - \overline{e_t})}{\sum_i (e_{d,t} - \overline{e_t})^2} \quad \omega_{d,t} := \frac{e_{d,t}(e_{d,t} - \overline{e_i})}{\sum_t (e_{d,t} - \overline{e_i})^2} \tag{3.37}$$

where $\overline{e_t} = \sum_i e_{d,t}/n_I$ with n_I being the number of observations at dimension i.[69] As I will illustrate below, by conditioning on b_t^1 and b_d^1, I am essentially conditioning on values of $\beta_{d,t}$ with a particular weight.

[66] In this simple case, I only need to make counterfactual predictions of $p_{d,t}$ changing $e_{d,t}$ conditioning on the value of $\beta_{d,t}$. When the unobserved variable M_t^{outer} does appear in the outer part of the equation of $p_{d,t}$, I will need to make predictions conditioning on the values of both $\beta_{d,t}$ and M_t^{outer}. In general, the missing component in the inner part of $\beta_{d,t}$ and the outer part M_t^{outer} need not take the same functional form nor the same value.

[67] More formally, I should define a multi-dimensional monotonic transformation measure. I leave this task for my future work.

[68] In this simple case, I only need to use information provided by b_t^1 and b_d^1. b_t^0 and b_d^0 will be used in more complicated cases as in subsection 3.C.1.

[69] I assume that the panel is balanced. I use notations n_I and n_T rather than the conventional N and T.

Case A: If weights $w_{d,t} = 1/n_I$ and $\omega_{d,t} = 1/n_T$ and $\beta_{d,t}$ can be approximated by $\beta_{d,t} = v_d + v_t + v_d * v_t$, then $\overline{\beta_t} = \overline{v^d}(1 + v_t)$ and $\overline{\beta_d} = (v_d + 1)\overline{v^t}$, where $\overline{v^t} := \sum_t v_t / n_T$.

$$\frac{b_t^1}{\overline{v^d}} * \frac{b_d^1}{\overline{v^t}} - 1 = v_d + v_t + v_d * v_t = \beta_{d,t} \tag{3.38}$$

Case B: If $\beta_{d,t} = v_d + v_t + v_d * v_t$ but $w_{d,t} \neq 1/n_I$ and $\omega_{d,t} \neq 1/n_T$, then

$$b_t^1 = \sum_i (v_d + v_t + v_d * v_t) w_{d,t} = \sum_i v_d w_{d,t} + v_t (\sum_i w_{d,t} + \sum_i v_d w_{d,t}) \tag{3.39}$$

$$b_d^1 = \sum_t (v_t + v_d + v_d * v_t) \omega_{d,t} = \sum_t v_t \omega_{d,t} + v_d (\sum_t \omega_{d,t} + \sum_t v_t \omega_{d,t}) \tag{3.40}$$

Notice $\sum_t v_t \omega_{d,t} = 0$ and $\sum_i v_d w_{d,t} = 0$

This is still too abstract. Let $e_{d,t} = k_t + k_d + k_d * k_t$. Then $\overline{e_t} = \sum_i (k_t + k_d + k_d * k_t)/n_I$ and $(e_{d,t} - \overline{e_t})^2 = [k_d - \overline{k^I} + (k_d - \overline{k^I}) * k_t]^2$.

$$\begin{aligned} b_t^1 &= \frac{\sum_i (v_d + v_t + v_d * v_t)(k_t + k_d + k_d * k_t)[k_d - \overline{k^I} + (k_d - \overline{k^I}) * k_t]}{\sum_i [k_d - \overline{k^I} + (k_d - \overline{k^I}) * k_t]^2} \\ &= v_t + (1 + v_t) \frac{\sum_i v_d (k_t + k_d + k_d * k_t)[k_d - \overline{k^I} + (k_d - \overline{k^I}) * k_t]}{\sum_i [k_d - \overline{k^I} + (k_d - \overline{k^I}) * k_t]^2} \end{aligned} \tag{3.41}$$

$$\begin{aligned} b_d^1 &= \frac{\sum_t (v_d + v_t + v_d * v_t)(k_t + k_d + k_d * k_t)[k_t - \overline{k^T} + (k_t - \overline{k^T}) * k_d]}{\sum_t [k_t - \overline{k^T} + (k_t - \overline{k^T}) * k_d]^2} \\ &= v_d + (1 + v_d) \frac{\sum_t v_t (k_t + k_d + k_d * k_t)[k_t - \overline{k^T} + (k_t - \overline{k^T}) * k_d]}{\sum_t [k_t - \overline{k^T} + (k_t - \overline{k^T}) * k_d]^2} \end{aligned} \tag{3.42}$$

Define sample conditional covariance measures as[70]:

$$Cov_d(x_{d,t}, z_{d,t}) := \sum_i (x_{d,t} - \overline{x_t})(z_{d,t} - \overline{z_t})/n_I \tag{3.43}$$

$$Var_i(x_{d,t}) := Cov_d(x_{d,t}, x_{d,t}) \tag{3.44}$$

[70]These definitions are used to simplify my notation only and may not be a consistent measure of conditional covariance.

From the relationship $\sum_i x_{d,t} z_{d,t}/n_I = Cov_d(x_{d,t}, z_{d,t}) + \sum_i x_{d,t} \sum_i z_{d,t}/n_I^2$, b_t^1 and b_d^1 can be rewritten as:

$$b_t^1 = [v_t + (1+v_t)\overline{v^d}] + (1+v_t)\frac{Cov_d\{v_d, (k_t + k_d + k_d * k_t)[k_d - \overline{k^I} + (k_d - \overline{k^I}) * k_t]\}}{Var_i(k_d - \overline{k^I} + (k_d - \overline{k^I}) * k_t)} \tag{3.45}$$

$$b_d^1 = [v_d + (1+v_d)\overline{v^t}] + (1+v_d)\frac{Cov_t\{v_t, (k_t + k_d + k_d * k_t)[k_t - \overline{k^T} + (k_t - \overline{k^T}) * k_d]\}}{Var_t[k_t - \overline{k^T} + (k_t - \overline{k^T}) * k_d]} \tag{3.46}$$

Equations (3.45) and (3.46) can be expressed as[71]:

$$b_t^1 = [v_t + (1+v_t)\overline{v^d}] + (1+v_t)\frac{Cov_d[v_d, (e_{d,t} - \overline{e_t})^2]}{Var_i(e_{d,t} - \overline{e_t})^2} \tag{3.47}$$

$$b_d^1 = [v_d + (1+v_d)\overline{v^t}] + (1+v_d)\frac{Cov_t[v_t, (e_{d,t} - \overline{e_i})^2]}{Var_t(e_{d,t} - \overline{e_i})^2} \tag{3.48}$$

The last part can be further simplified as

$$\frac{Cov_d[v_d, (e_{d,t} - \overline{e_t})^2]}{Var_i(e_{d,t} - \overline{e_t})^2} = \frac{Cov_d[v_d, (k_d - \overline{k^I})^2]}{Var_i(k_d)} =: c_1 \tag{3.49}$$

$$\frac{Cov_t[v_t, (e_{d,t} - \overline{e_i})^2]}{Var_t(e_{d,t} - \overline{e_i})^2} = \frac{Cov_t[v_t, (k_t - \overline{k^T})^2]}{Var_t(k_t)} =: c_2 \tag{3.50}$$

In this example, due to the simple factorisation of $e_{d,t}$, I assume that the last part is a constant and does not vary along the other dimensions. This property no longer holds in a more general factorisation process, e.g. $e_{d,t} = k_t^1 + k_d^1 + k_d^1 k_t^2$.

In general, I discuss three possibilities here:

Case B.1: If $Cov_d[v_d, (e_{d,t} - \overline{e_t})^2] = 0$ and $Cov_t[v_t, (e_{d,t} - \overline{e_i})^2] = 0$, the true $\beta_{d,t}$ can be expressed as a simple nonlinear equation of b_t^1 and b_d^1 as in Case A.

[71]Hint: rewrite $k_t + k_d + k_d * k_t = [k_d - \overline{k^I} + (k_d - \overline{k^I}) * k_t] + [k_t + \overline{k^I} + \overline{k^I} k_t]$

Case B.2: If $Cov_d[v_d, (e_{d,t} - \overline{e_t})^2] \neq 0$ and $Cov_t[v_t, (e_{d,t} - \overline{e_i})^2] \neq 0$, but $Var_t \left\{ \frac{Cov_d[v_d,(e_{d,t}-\overline{e_t})^2]}{Var_i(e_{d,t}-\overline{e_t})^2} \right\} = 0$ and $Var_i \left\{ \frac{Cov_t[v_t,(e_{d,t}-\overline{e_i})^2]}{Var_t(e_{d,t}-\overline{e_i})^2} \right\} = 0$, v_d and v_t can be written as follows:

$$b_t^1 = v_t + (1 + v_t)(\overline{v^d} + c_1) \tag{3.51}$$

$$b_d^1 = v_d + (1 + v_d)(\overline{v^t} + c_2) \tag{3.52}$$

$$v_t = \frac{b_t^1 - (\overline{v^d} + c_1)}{1 + (\overline{v^d} + c_1)} \tag{3.53}$$

$$v_d = \frac{b_d^1 - (\overline{v^t} + c_2)}{1 + (\overline{v^t} + c_2)} \tag{3.54}$$

It is clear that $\beta_{d,t}$ can be expressed as a nonlinear function of b_t^1 and b_d^1.

Case B.3: If $\frac{Cov_d[v_d,(e_{d,t}-\overline{e_t})^2]}{Var_i(e_{d,t}-\overline{e_t})^2} \neq c_1$ and $\frac{Cov_t[v_t,(e_{d,t}-\overline{e_i})^2]}{Var_t(e_{d,t}-\overline{e_i})^2} \neq c_2$ but $Var_t \left\{ \frac{Cov_d[v_d,(e_{d,t}-\overline{e_t})^2]}{Var_i(e_{d,t}-\overline{e_t})^2} \right\}$ and $Var_i \left\{ \frac{Cov_t[v_t,(e_{d,t}-\overline{e_i})^2]}{Var_t(e_{d,t}-\overline{e_i})^2} \right\}$ are very small, the weak monotonic property will make it work.

This exercise gives me two interesting insights. First, variances of the bias at the other dimension matter. As long as the bias of estimated parameters at one dimension is "well-structured" in the other dimension, adding these estimated parameters will help to estimate the desired $\beta_{d,t}$.

Second, it is the covariance between elements driving $\beta_{d,t}$ and a local measure of second moments of the policy variable $(e_{d,t} - \overline{e_t})^2$ and $(e_{d,t} - \overline{e_i})^2$ that matters. Due to the linear regression structure, first order terms are filtered out and only second order terms will influence the bias. This is a very useful property for economics studies. Taking this property into the context of my empirical ERPT question, exchange rates may be correlated with the marginal cost of the exporter[72], but it is less likely for the volatility of exchange rates to be correlated with the level movement of the marginal cost of the exporter. Even if these two terms are correlated, as long as the correlations (as a function of d) do not change systematically across destinations, estimated parameters b_t^1 and b_d^1 will provide useful information which can be analysed through a tree based machine learning algorithm.

In general, adding estimated parameters from regressions and/or other structural estimations from a range of dimension-limited partition spaces should always be more efficient than adding indices i and t or dummies related to the indices provided that the assumed structural equation is not very far from the true specification.

3.C.1 The case where M_t appears in the outer part of the linear form

In this subsection, I discuss the case where M_t appears in the outer part of the linear form. As not only $\beta_{d,t}$ but also M_t need to be backed up from the estimated parameters, I also need

[72]For example, the exporter may use imported inputs.

to use information from b_t^0 and b_d^0.

$$p_{d,t} = \beta_{d,t} e_{d,t} + M_t$$
$$\beta_{d,t} = f(X_{d,t}, M_t)$$

Regression estimated parameters can be written as

$$\begin{aligned}
b_t^1 &= \frac{\sum_i [f(X_{d,t}, M_t) e_{d,t} + M_t](e_{d,t} - \overline{e_t})}{\sum_i (e_{d,t} - \overline{e_t})^2} \\
&= \sum_i \beta_{d,t} w_{d,t} + M_t \\
b_t^0 &= \sum_i (p_{d,t} - b_t^1 * e_{d,t}) / n_I \\
&= \sum_i (\beta_{d,t} - \sum_i \beta_{d,t} w_{d,t}) e_{d,t} / n_I + M_t (1 - \sum_i e_{d,t} / n_I)
\end{aligned}$$

$$\begin{aligned}
b_d^1 &= \frac{\sum_t [f(X_{d,t}, M_t) e_{d,t} + M_t](e_{d,t} - \overline{e_i})}{\sum_t (e_{d,t} - \overline{e_i})^2} \\
&= \sum_t \beta_{d,t} \omega_{d,t} + \sum_t M_t \frac{\omega_{d,t}}{e_{d,t}} \\
b_d^0 &= \sum_t (p_{d,t} - b_d^1 * e_{d,t}) / n_T \\
&= \sum_t (\beta_{d,t} - \sum_t \beta_{d,t} \omega_{d,t}) e_{d,t} / n_T + \sum_t M_t (1 - \omega_{d,t}) / n_T
\end{aligned}$$

To visualise the underlying structure of these estimated parameters, I take the following first order factorisation. Let

$$e_{d,t} = k_t + k_d + k_d * k_t$$
$$\beta_{d,t} = v_d + v_t + v_d * v_t$$
$$M_t = m_t$$

Therefore,

$$b_t^1 = [v_t + (1 + v_t)\overline{v^d}] + (1 + v_t) \frac{Cov_d[v_d, (e_{d,t} - \overline{e_t})^2]}{Var_i(e_{d,t} - \overline{e_t})^2} + m_t \tag{3.55}$$

$$b_d^1 = [v_d + (1 + v_d)\overline{v^t}] + (1 + v_d) \frac{Cov_t[v_t, (e_{d,t} - \overline{e_i})^2]}{Var_t(e_{d,t} - \overline{e_i})^2} + \frac{Cov_t(m_t, e_{d,t})}{Var_t(e_{d,t} - \overline{e_i})^2} \tag{3.56}$$

$$b_t^0 = Cov_d[\beta_{d,t}, e_{d,t}] - (\frac{Cov_d[\beta_{d,t}, (e_{d,t} - \overline{e_t}) e_{d,t}]}{Var_i(e_{d,t} - \overline{e_t})} + M_t)\overline{e_t} + M_t \tag{3.57}$$

where I have used the following relationship in deriving the expression of b_t^0.

$$\sum_i [\beta_{d,t}(e_{d,t} - \overline{e_t})e_{d,t}]/n_I = Cov_d[\beta_{d,t}, (e_{d,t} - \overline{e_t})e_{d,t}] + \sum_i \beta_{d,t} \sum_i [(e_{d,t} - \overline{e_t})^2]/n_I^2 \tag{3.58}$$

With the assumed factorisation,

$$Cov_d[\beta_{d,t}, e_{d,t}] = (1 + v_t)(1 + k_t)Cov_d(v_d, k_d) \tag{3.59}$$

$$Cov_d[\beta_{d,t}, (e_{d,t} - \overline{e_t})e_{d,t}] = (1 + v_t)(1 + k_t)^2 Cov_d[v_d, (k_d - \overline{k^I})^2] \tag{3.60}$$

$$\frac{Cov_d[\beta_{d,t}, (e_{d,t} - \overline{e_t})e_{d,t}]}{Var_i(e_{d,t} - \overline{e_t})} = (1 + v_t)\frac{Cov_d[v_d, (k_d - \overline{k^I})^2]}{Var_i(k_d)} =: (1 + v_t)c_1 \tag{3.61}$$

This is where it becomes complicated. The expression of b_t^0 now involves a time-varying factor k_t of the observed policy variable $e_{d,t}$.

$$b_t^0 = (1 + v_t)(1 + k_t)Cov_d(v_d, k_d) - [(1 + v_t)c_1 + m_t](k_t + k_t\overline{k^I} + \overline{k^I}) + m_t \tag{3.62}$$

$Cov_d(v_d, k_d)$, c_1, $\overline{k^I}, \overline{v^d}, \overline{v^t}$ are constants. This leaves v_t, v_d, m_t, k_t to be solved in 3 equations (3.55), (3.56) and (3.62). The tricky part to figure out is how and when the conditional weak monotonic transformation property in proportion 2 works in this case.

If I derive the expression of b_d^0, it would involve v_d, k_d. Together with $e_{d,t}$, (3.55), (3.56) and (3.62), there are 5 equations with 5 unknowns v_t, v_d, m_t, k_d, k_t. If this problem can be solved approximately, the efficiency gain in adding these estimated parameters should depend on the complexity of these equations compared to the complexity of hidden functions of using indices or related dummies.

Note that I discussed a general case here where the unobserved variable M_t needs not necessarily be correlated with $\beta_{d,t}$.[73] If $\beta_{d,t}$ can be expressed as an explicit function of M_t and some observed variables, the derivation will be easier.

[73] I did not impose any restrictions on m_t and v_t.

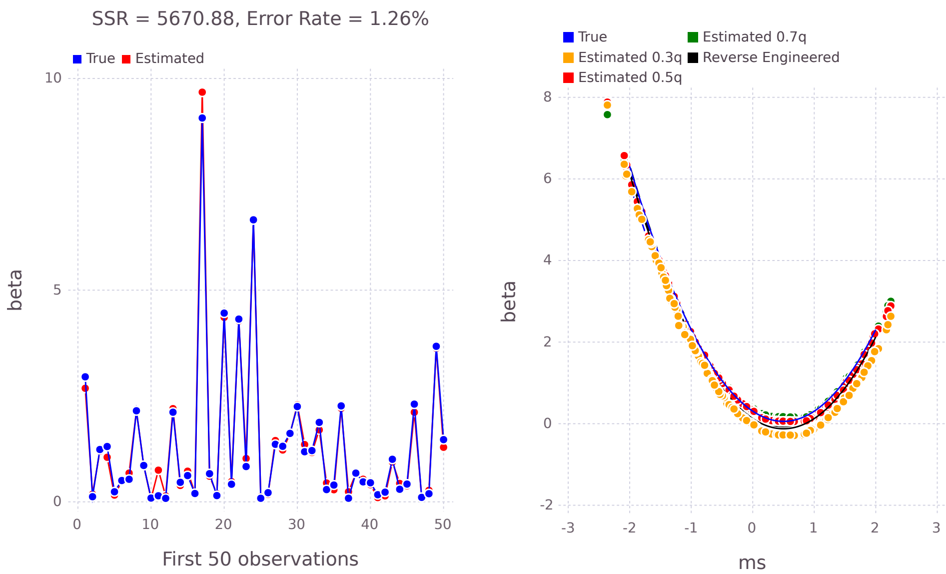
3.C.2 Tests on alternative specifications

High Nonlinearity

Setting:

$$
\begin{aligned}
p_{d,t} &= 10 + \beta_{d,t} e_{d,t} + ms_{d,t} - mc_t + \epsilon_{d,t} \\
\beta_{d,t} &= (ms_{d,t} - 0.5)^2 + sin(1000mc_t)mc_t \\
ms_{d,t} &= u_{d,t} + 0.1e_{d,t} \\
mc_t &= u_t - 0.1\overline{e_t} \\
\overline{e_t} &= \frac{\sum_d e_{d,t}}{n_d} \\
n_d &= 2000;\ n_t = 40 \\
u_{d,t} &\sim N(0,1),\ u_t \sim N(0,1),\ \epsilon_{d,t} \sim N(0,0.01)
\end{aligned}
$$

Fig. 3.C.1 High nonlinearity: the proposed algorithm

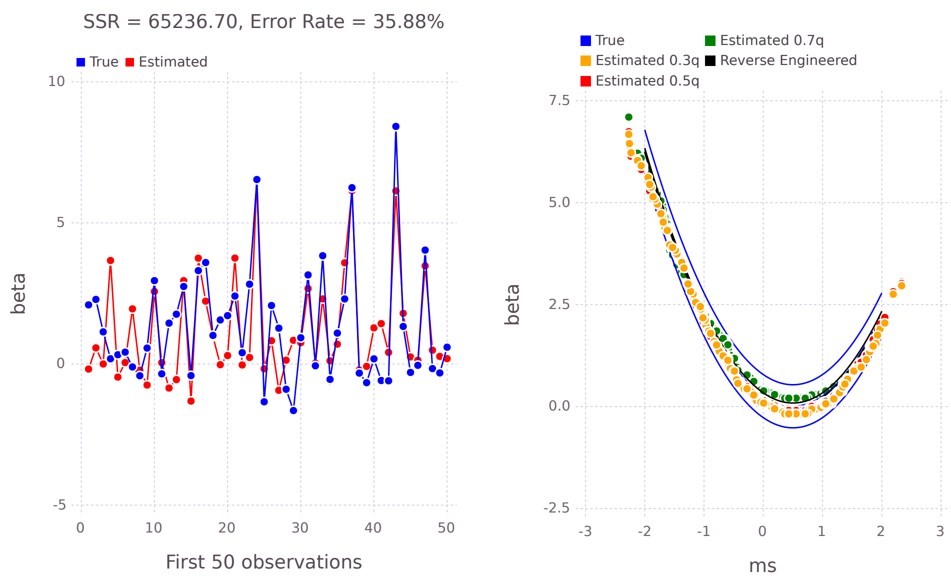


Not Identifiable

Setting:

$$
\begin{aligned}
p_{d,t} &= 10 + \beta_{d,t} e_{d,t} + ms_{d,t} - mc_{d,t} + \epsilon_{d,t} \\
\beta_{d,t} &= (ms_{d,t} - 0.5)^2 + mc_{d,t} \\
ms_{d,t} &= u_{d,t} + 0.1 e_{d,t} \\
mc_{d,t} &= u_{d,t} - 0.1 e_{d,t} \\
n_d &= 2000;\ n_t = 40 \\
u_{d,t} &\sim N(0,1),\ u_t \sim N(0,1),\ \epsilon_{d,t} \sim N(0,0.01)
\end{aligned}
$$

Fig. 3.C.2 Not identifiable: the proposed algorithm

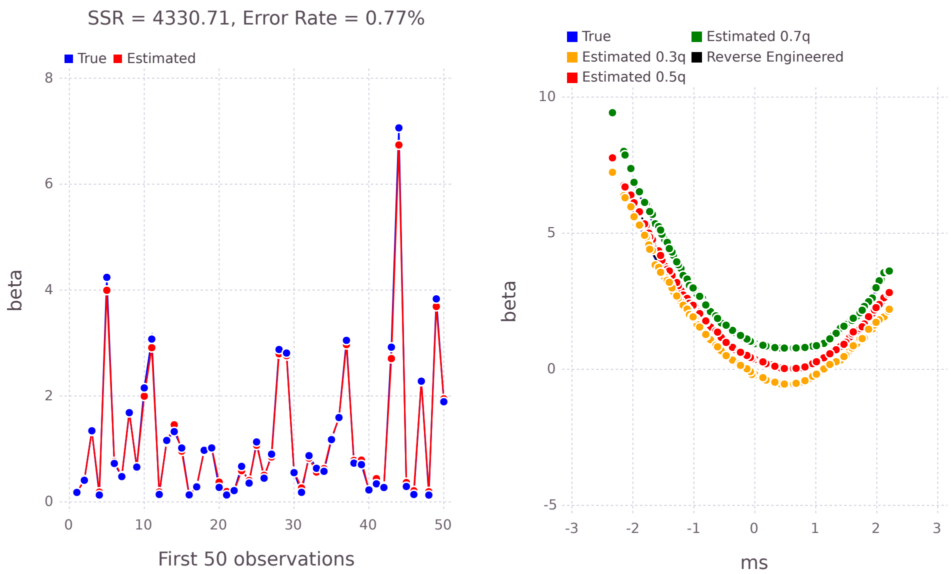


Larger Correlation

Setting:

$$
\begin{aligned}
p_{d,t} &= 10 + \beta_{d,t} e_{d,t} + ms_{d,t} - mc_t + \epsilon_{d,t} \\
\beta_{d,t} &= (ms_{d,t} - 0.5)^2 + mc_t \\
ms_{d,t} &= u_{d,t} + 0.1 e_{d,t} \\
mc_t &= u_t - 1\overline{e_t} \\
\overline{e_t} &= \frac{\sum_d e_{d,t}}{n_d} \\
n_d &= 2000;\ n_t = 40 \\
u_{d,t} \sim & N(0,1),\ u_t \sim N(0,1),\ \epsilon_{d,t} \sim N(0,0.01)
\end{aligned}
$$

Fig. 3.C.3 Larger correlation: the proposed algorithm

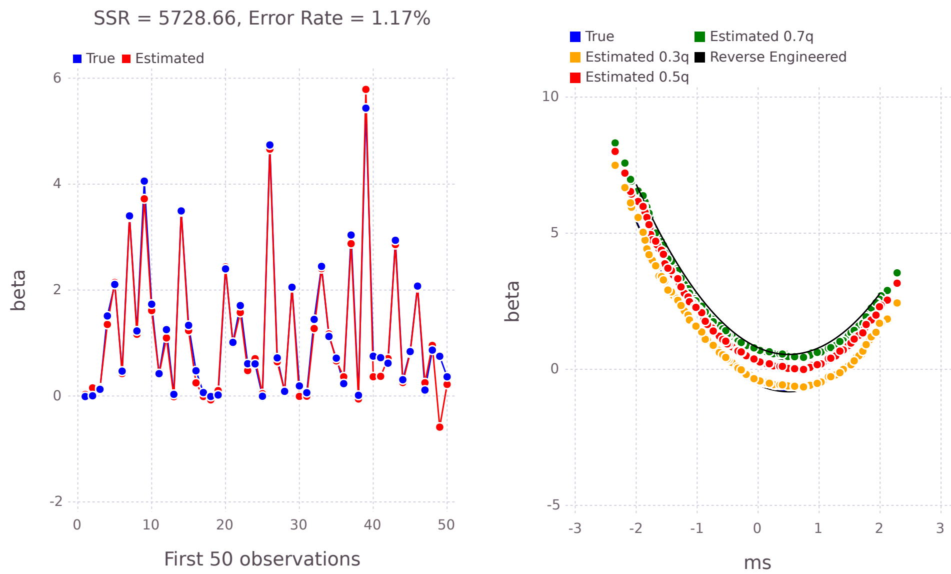


Different Function of the Outer Part

Setting:

$$
\begin{aligned}
p_{d,t} &= 10 + \beta_{d,t} e_{d,t} + ms_{d,t} - mc_t^2 + \epsilon_{d,t} \\
\beta_{d,t} &= (ms_{d,t} - 0.5)^2 + mc_t \\
ms_{d,t} &= u_{d,t} + 0.1 e_{d,t} \\
mc_t &= u_t - 0.1 \overline{e_t} \\
\overline{e_t} &= \frac{\sum_d e_{d,t}}{n_d} \\
n_d &= 2000;\ n_t = 40 \\
u_{d,t} &\sim N(0,1),\ u_t \sim N(0,1),\ \epsilon_{d,t} \sim N(0,0.01)
\end{aligned}
$$

Fig. 3.C.4 Different function of the outer part: the proposed algorithm

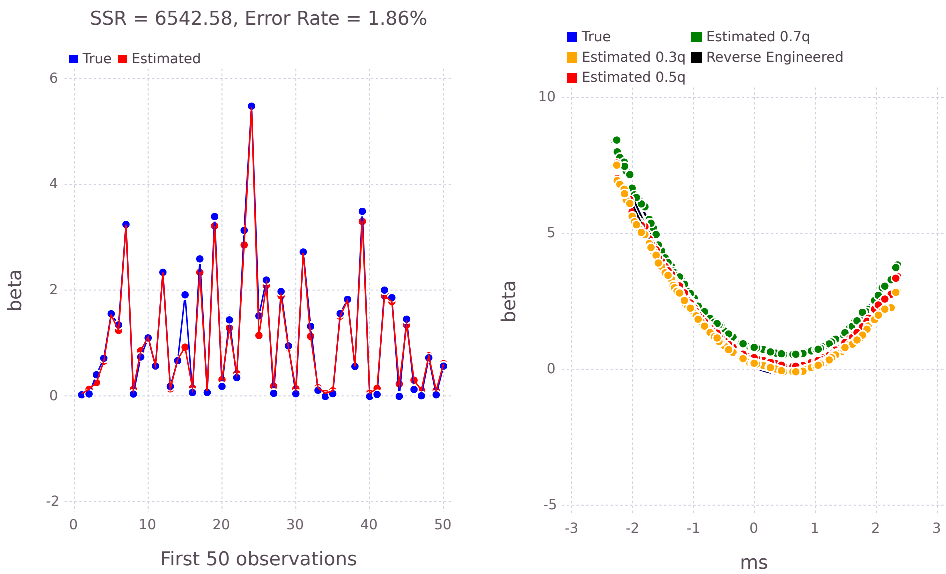


Arellano and Bond

Setting:

$$
\begin{aligned}
p_{d,t} &= 0.95 p_{d,t-1} + \beta_{d,t} e_{d,t} + ms_{d,t} - mc_t + \epsilon_{d,t} \\
\beta_{d,t} &= (ms_{d,t} - 0.5)^2 + mc_t \\
ms_{d,t} &= u_{d,t} + 0.1 e_{d,t} \\
mc_t &= u_t - 0.1 \overline{e_t} \\
\overline{e_t} &= \frac{\sum_d e_{d,t}}{n_d} \\
n_d &= 2000;\ n_t = 40 \\
u_{d,t} \sim & N(0,1),\ u_t \sim N(0,1),\ \epsilon_{d,t} \sim N(0,0.01)
\end{aligned}
$$

Fig. 3.C.5 Arellano and Bond: the proposed algorithm



Reduce Sample Size

Setting:

$$
\begin{aligned}
p_{d,t} &= 10 + \beta_{d,t} e_{d,t} + ms_{d,t} - mc_t + \epsilon_{d,t} \\
\beta_{d,t} &= (ms_{d,t} - 0.5)^2 + mc_t \\
ms_{d,t} &= u_{d,t} + 0.1 e_{d,t} \\
mc_t &= u_t - 0.1 \overline{e_t} \\
\overline{e_t} &= \frac{\sum_d e_{d,t}}{n_d} \\
n_d &= 200;\ n_t = 40 \\
u_{d,t} &\sim uniform,\ u_t \sim uniform,\ \epsilon_{d,t} \sim N(0, 0.01)
\end{aligned}
$$

Fig. 3.C.6 Reduce sample size: dummies

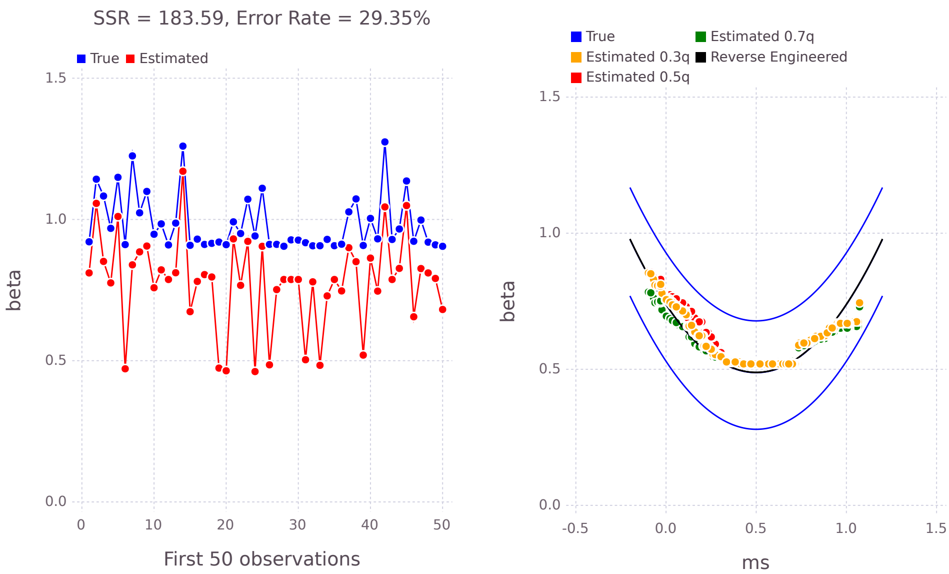
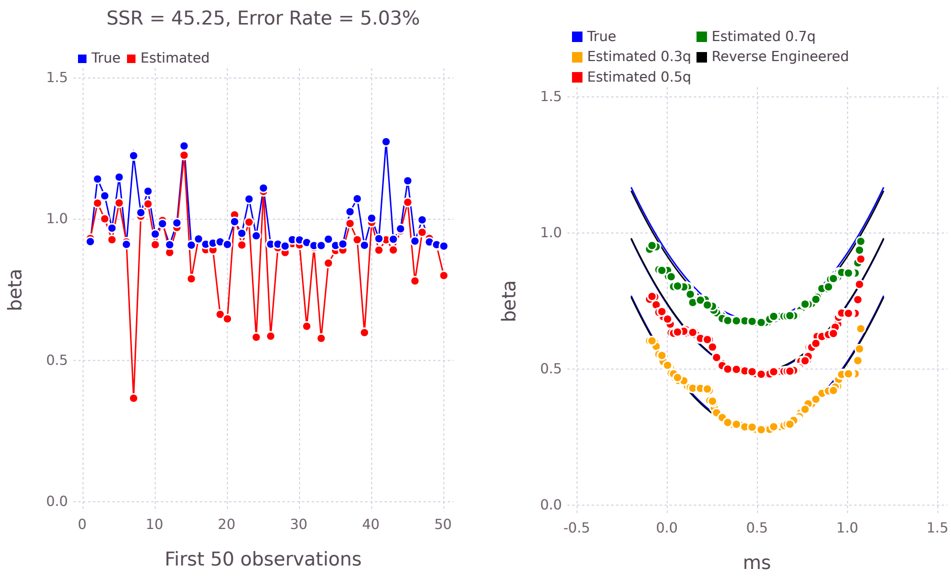
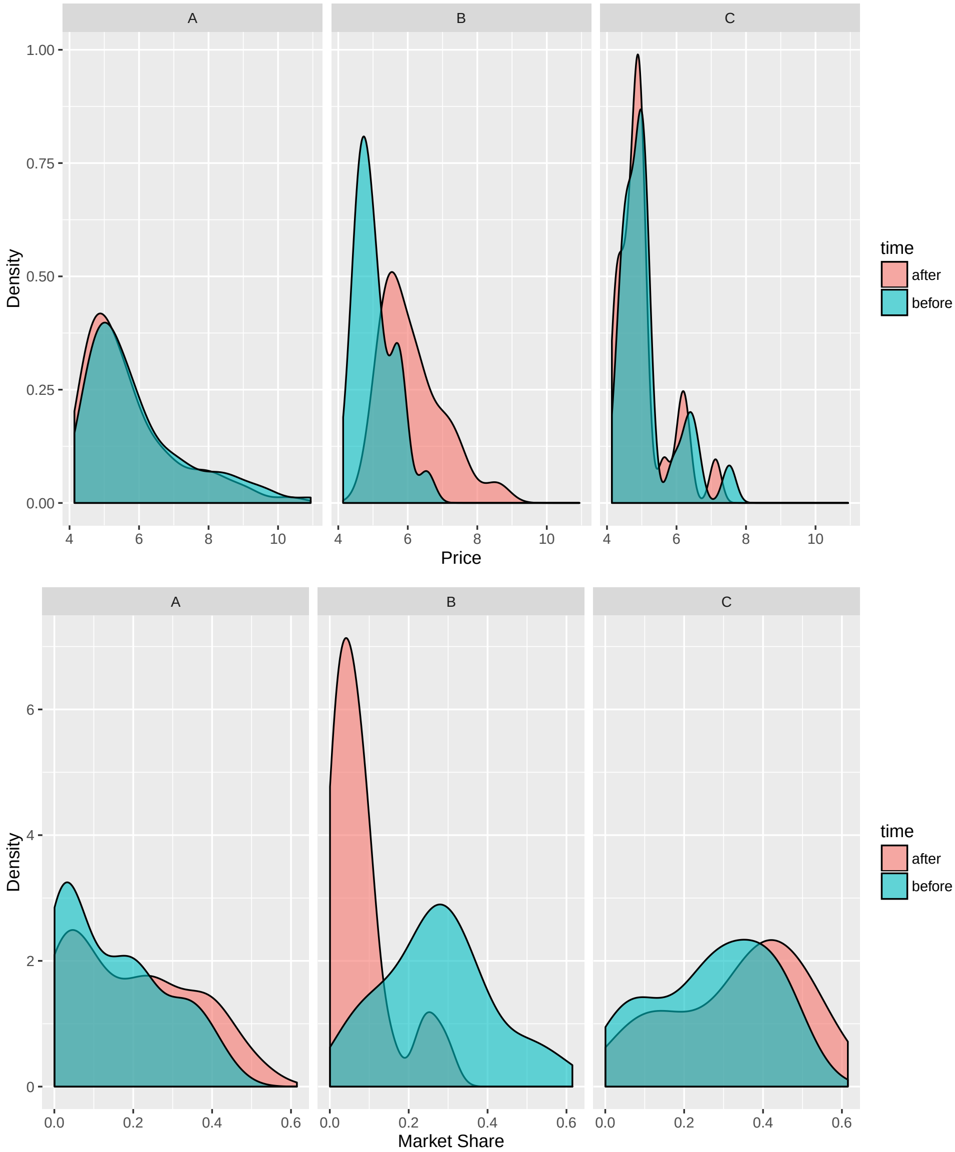


Fig. 3.C.7 Reduce sample size: the proposed algorithm



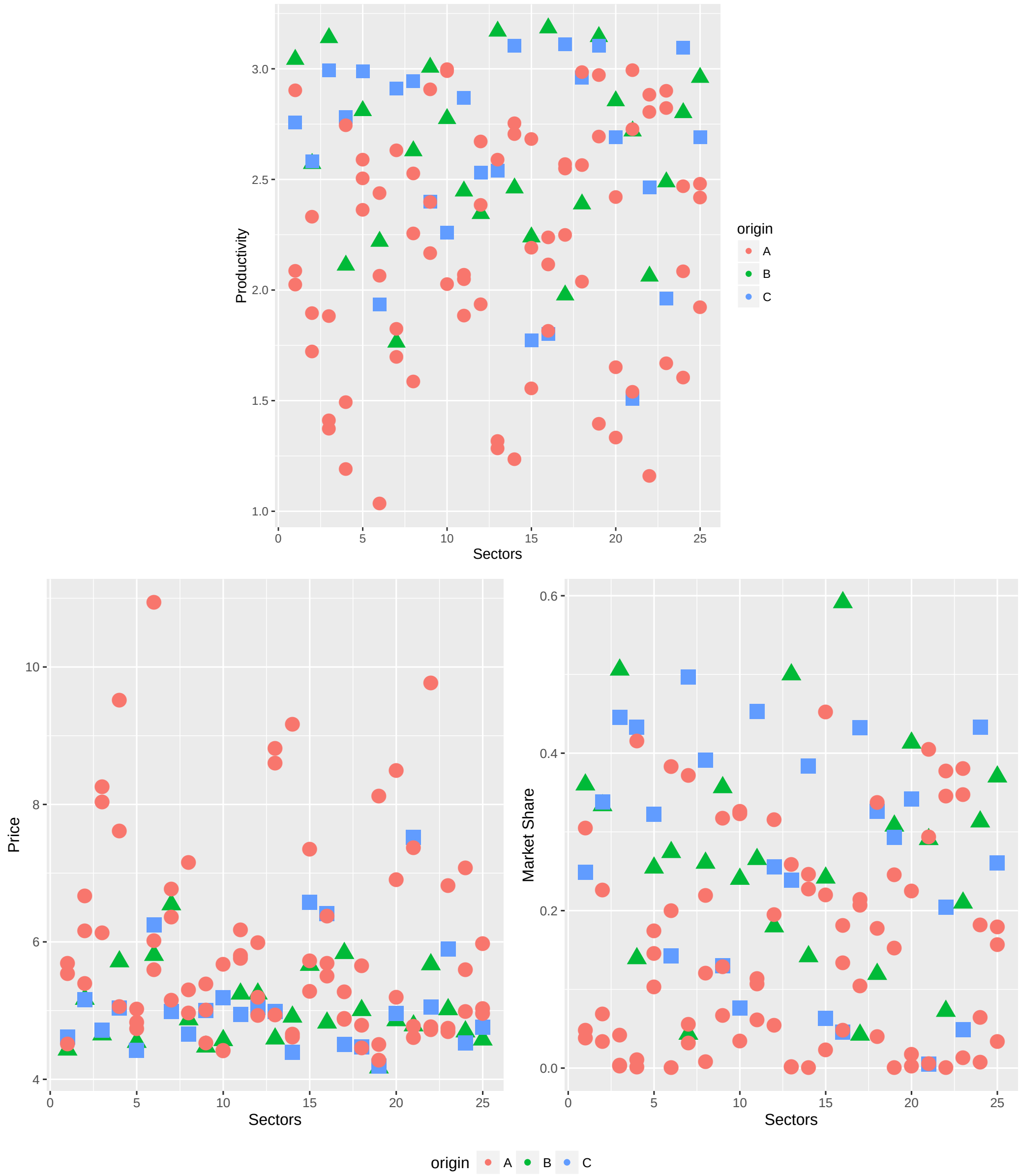
Appendix 3.D Details of the Simulated Model

Fig. 3.D.1 Responses of firms in country A to an appreciation in country B



Note: The left column presents the change of prices and market shares for domestic firms in country A. The middle and right columns present the reactions of exporters from country B and C respectively.

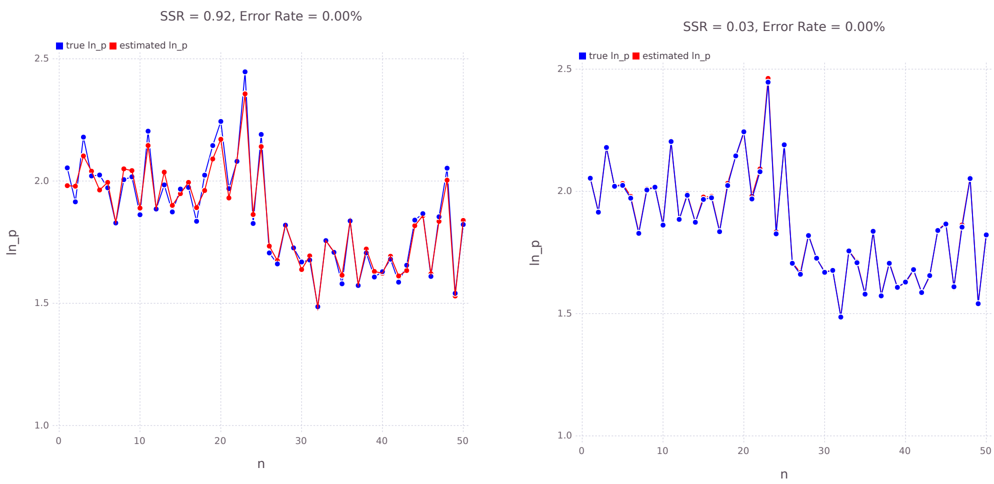
Fig. 3.D.2 Visualisation of simulated firms in country A



Note: The top graph depicts the realised productivity of firms in country A. In each sector, there are three domestic firms and two foreign firms from country B and C respectively (only the best firm in each sector exports). The bottom two graphs depict the price and market shares of firms in country A. Exporters are firms with relatively high productivity and charge relatively low prices and own larger market shares. The assumption that only the best firms export gives a realistic market structure in this multi-country world.

3.D.1 Case 1: only exchange rate shocks

Fig. 3.D.3 Case 1: Precision on price predictions



(a) GBRT without adding regression coefficients

(b) the proposed algorithm

Fig. 3.D.4 Case 1: Without adding regression coefficients

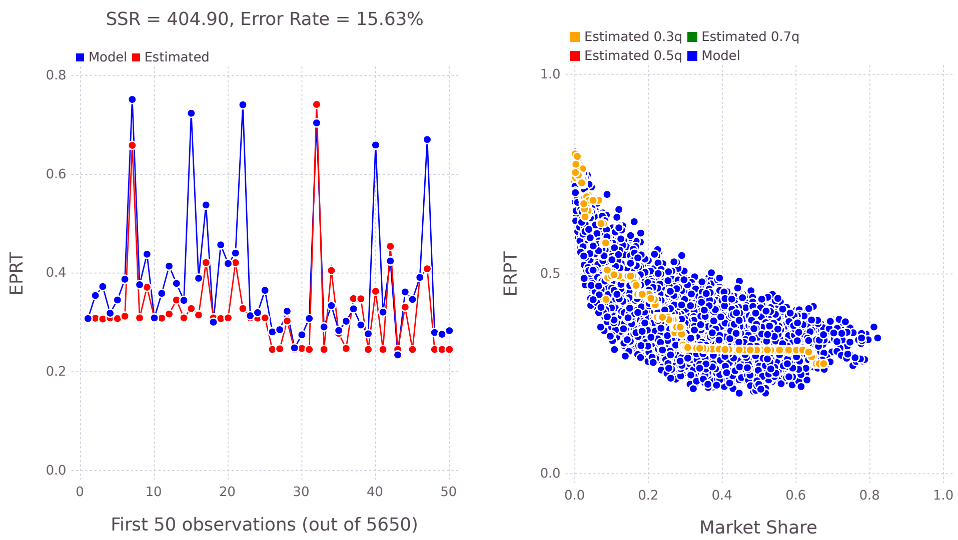
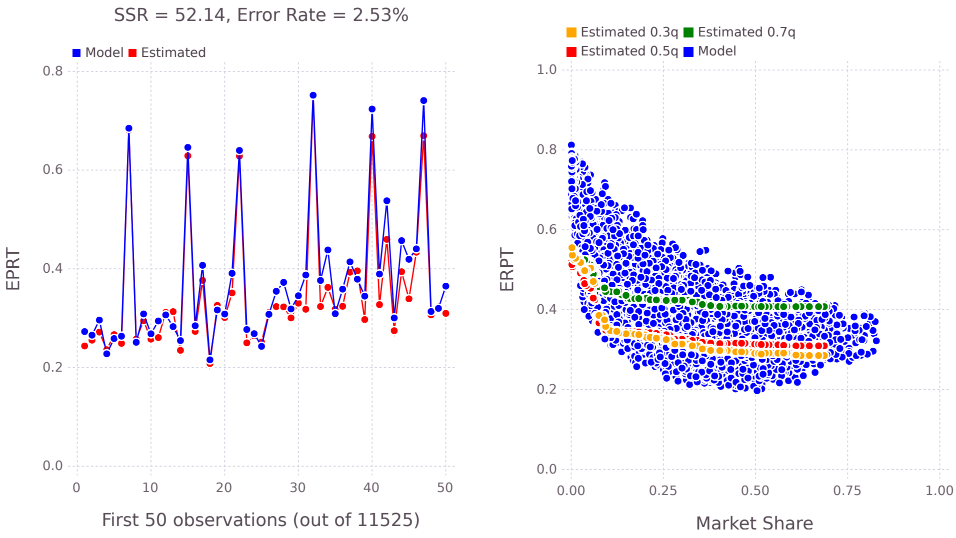
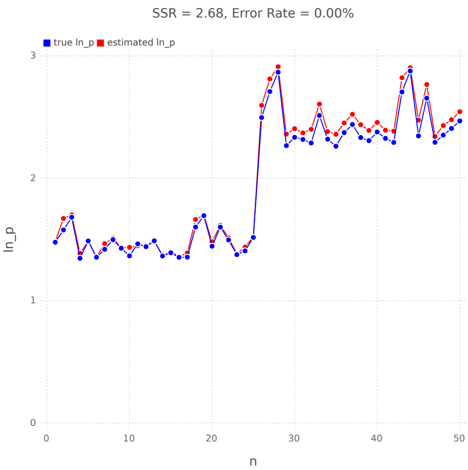


Fig. 3.D.5 Case 1: Point estimates of the proposed algorithm compared to true counterfactual environments

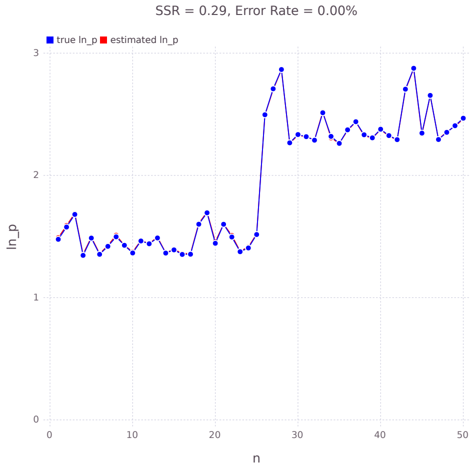


3.D.2 Case 2: adding productivity shocks

Fig. 3.D.6 Case 2: Precision on price predictions



(a) GBRT without adding regression coefficients



(b) the proposed algorithm

Fig. 3.D.7 Case 2: Point estimates of the proposed algorithm compared to true counterfactual environments

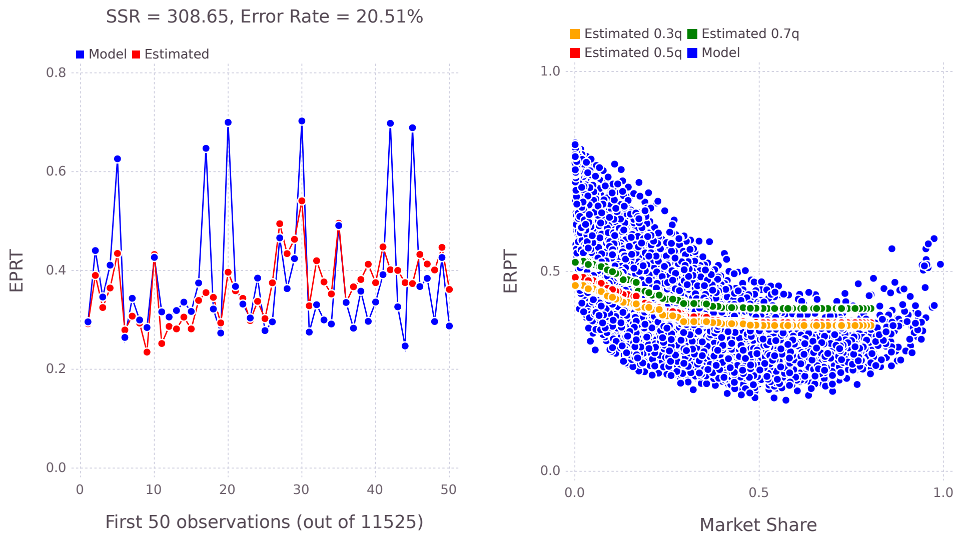
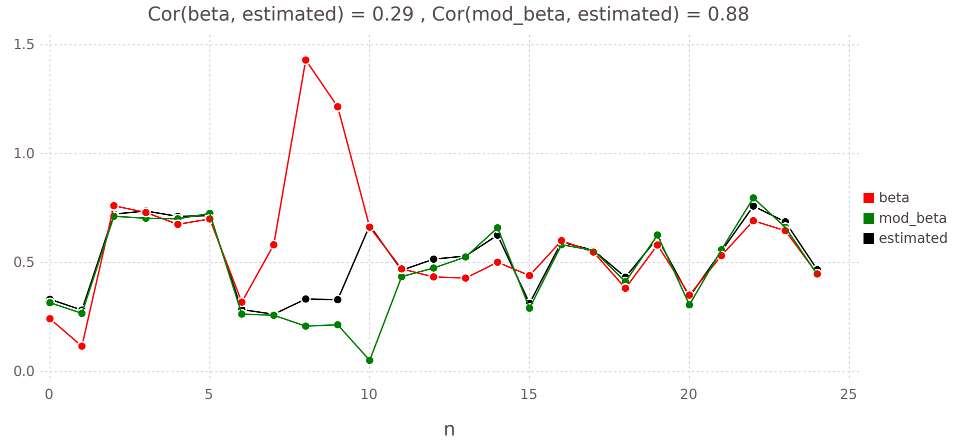


Fig. 3.D.8 Comparing naive, counterfactual and algorithm predicted ERPT estimates



Note: Firm's productivity is assumed to follow an AR(1) process with a persistence of 0.95. The red line presents the ERPT estimates calculated using actual price changes of the simulated model. The green line represents the model implied ERPT estimates in a counterfactual equilibrium where there is no productivity shock in the next period. The black line represents ERPT estimates predicted by the proposed algorithm.

Algorithm 1 The Proposed Algorithm

Input data $\mathbf{I}, \mathbf{y}, \mathbf{X}, \mathbf{e}$

Obtain variable names of the index matrix $\mathbf{I}$ and the feature variable matrix $\mathbf{X}$ and save them as i_{names} and x_{names} respectively.
Calculate all non-repetitive combinations of dimension indices in i_{names} and save as S_i.
for s in S_i **do**
  $\mathbf{I_s} \leftarrow \mathbf{I}[i_{names} \in s]$
  $\widetilde{\mathbf{I}}_s \leftarrow unique(\mathbf{I_s})$
  for x in x_{names} **do**
    $x_s \leftarrow \mathbf{0}$
    for i_s in 1 to nrow($\widetilde{\mathbf{I}}_s$) **do**
      $x_s[\mathbf{I_s} = \widetilde{\mathbf{I}}[i_s]] \leftarrow mean(x|\mathbf{I_s} = \widetilde{\mathbf{I}}[i_s])$
    end for
  end for
end for
Calculate all non-repetitive binary combinations of S_i and save as S_{share}.
for s in S_{share} **do**
  $(s_a, s_b) \leftarrow s[sort(length(s[1], s[2]))]$
  for x in x_{names} **do**
    $x_{s_a,s_b} \leftarrow \frac{x_{s_a}}{x_{s_b}}$
  end for
end for
Observe dimensions in which the supervisor $\mathbf{y}$ and the policy/treatment variable $\mathbf{e}$ vary. Identify a subset available for controlling unobserved variables and save as S_{id}.
for s in S_{id} **do**
  Assume a possible (linear) structural equation based on economic rationale.
  for j in 1:(number of parameters in the structural model) **do**
    $coef_s^j \leftarrow \mathbf{0}$
  end for
  for d_s in 1 to nrow($\widetilde{\mathbf{I}}_s$) **do**
    Estimate the structural regression for the subset of data where $\mathbf{I_s} = \widetilde{\mathbf{I}}[i_s]$
    for j in 1:(number of parameters in the structural model) **do**
      $coef_s^j[\mathbf{I_s} = \widetilde{\mathbf{I}}[i_s]] \leftarrow parameter^j$
    end for
  end for
end for
Run GBRT with supervisor $\mathbf{y}$ on $\mathbf{e}, \mathbf{X}, \mathbf{X_{s_a,s_b}}, \mathbf{coef_{id}^j}$ and obtain $model_1$.
$\mathbf{y^{Est1}} \leftarrow model_1(\mathbf{e} - 0.5std(\mathbf{e}, \mathbf{X}, \mathbf{X_s}, \mathbf{coef_s^j})$
$\mathbf{y^{Est2}} \leftarrow model_1(\mathbf{e} + 0.5std(\mathbf{e}), \mathbf{X}, \mathbf{X_s}, \mathbf{coef_s^j})$
$\mathbf{beta^{Est}} \leftarrow \frac{\mathbf{y^{Est2}} - \mathbf{y^{Est1}}}{std(\mathbf{e})}$
Run GBRT again with supervisor $\mathbf{beta^{Est}}$ on $\mathbf{e}, \mathbf{X}, \mathbf{X_{s_a,s_b}}, \mathbf{coef_{id}^j}$ and obtain $model_2$.

Output: $model_1, model_2, \mathbf{beta^{Est}}$

References

Ahn, J., A. K. Khandelwal, and S.-J. Wei (2011). The role of intermediaries in facilitating trade. *Journal of International Economics 84*(1), 73–85.

Alessandria, G. (2009). Consumer search, price dispersion, and international relative price fluctuations. *International Economic Review 50*(3), 803–829.

Alvarez, F. and R. E. Lucas (2007). General equilibrium analysis of the eaton–kortum model of international trade. *Journal of Monetary Economics 54*(6), 1726–1768.

Amiti, M., O. Itskhoki, and J. Konings (2014). Importers, exporters, and exchange rate disconnect. *The American Economic Review 104*(7), 1942–1978.

Armington, P. S. (1969). A theory of demand for products distinguished by place of production. *IMF Staff Papers 16*(1), 159–178.

Athey, S. and G. Imbens (2015). A measure of robustness to misspecification. *The American Economic Review 105*(5), 476–480.

Athey, S. and G. Imbens (2016). The state of applied econometrics-causality and policy evaluation. *arXiv preprint arXiv:1607.00699*.

Athey, S., G. Imbens, T. Pham, and S. Wager (2017). Estimating average treatment effects: Supplementary analyses and remaining challenges. *The American Economics Review (forthcoming in May)*.

Athey, S., J. Tibshirani, and S. Wager (2016). Solving heterogeneous estimating equations with gradient forests. *arXiv preprint arXiv:1610.01271*.

Atkeson, A. and A. Burstein (2007). Pricing-to-market in a ricardian model of international trade. *The American Economic Review 97*(2), 362–367.

Atkeson, A. and A. Burstein (2008). Pricing-to-market, trade costs, and international relative prices. *The American Economic Review 98*(5), 1998–2031.

Auer, R. A. and R. S. Schoenle (2016). Market structure and exchange rate pass-through. *Journal of International Economics 98*, 60–77.

Backus, D. K., P. J. Kehoe, and F. E. Kydland (1992). International real business cycles. *Journal of Political Economy*, 745–775.

Backus, D. K., P. J. Kehoe, and F. E. Kydland (1994, March). Dynamics of the trade balance and the terms of trade: The J-Curve? *The American Economic Review 84*(1), 84–103.

Bajari, P., D. Nekipelov, S. P. Ryan, and M. Yang (2015). Machine learning methods for demand estimation. *The American Economic Review 105*(5), 481–485.

Baker, W., J. W. Hutchinson, D. Moore, and P. Nedungadi (1986). Brand familiarity and advertising: Effects on the evoked set and brand preference. *Advances in Consumer Research 13*(1), 637–642.

Berman, N., P. Martin, and T. Mayer (2012). How do different exporters react to exchange rate changes? *The Quarterly Journal of Economics 127*(1), 437–492.

Betts, C. and M. B. Devereux (2000). Exchange rate dynamics in a model of pricing-to-market. *Journal of International Economics 50*(1), 215–244.

Bhatt, S., P. W. Gething, O. J. Brady, J. P. Messina, A. W. Farlow, C. L. Moyes, J. M. Drake, J. S. Brownstein, A. G. Hoen, O. Sankoh, et al. (2013). The global distribution and burden of dengue. *Nature 496*(7446), 504–507.

Borzekowski, D. L. and T. N. Robinson (2001). The 30-second effect: An experiment revealing the impact of television commercials on food preferences of preschoolers. *Journal of the American Dietetic Association 101*(1), 42–46.

Breiman, L. (1996). Bagging predictors. *Machine learning 24*(2), 123–140.

Breiman, L. (2001). Random forests. *Machine learning 45*(1), 5–32.

Breiman, L., J. Friedman, C. J. Stone, and R. A. Olshen (1984). *Classification and regression trees*. CRC press.

Burstein, A., J. Neves, and S. Rebelo (2003). Distribution costs and real exchange rate dynamics during exchange-rate-based stabilizations. *Journal of Monetary Economics 50*(6), 1189–1214.

Chatterjee, A., R. Dix-Carneiro, and J. Vichyanond (2013). Multi-product firms and exchange rate fluctuations. *American Economic Journal: Economic Policy 5*(2), 77–110.

Chen, N., J. Imbs, and A. Scott (2009). The dynamics of trade and competition. *Journal of International Economics 77*(1), 50–62.

Chen, N. and L. Juvenal (2016). Quality, trade, and exchange rate pass-through. *Journal of International Economics 100*, 61–80.

Cheng, L. L.-S. and R. Sybesma (1999). Bare and not-so-bare nouns and the structure of np. *Linguistic inquiry 30*(4), 509–542.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, et al. (2016). Double machine learning for treatment and causal parameters. *arXiv preprint arXiv:1608.00060*.

Chernozhukov, V., C. Hansen, and M. Spindler (2015, May). Post-selection and post-regularization inference in linear models with many controls and instruments. *American Economic Review 105*(5), 486–90.

Correia, S. (2016). A feasible estimator for linear models with multi-way fixed effects. *Duke University Mimeo*.

Corsetti, G. and L. Dedola (2005). A macroeconomic model of international price discrimination. *Journal of International Economics 67*(1), 129–155.

Corsetti, G., L. Dedola, and S. Leduc (2007). Optimal monetary policy and the sources of local-currency price stability. In *International Dimensions of Monetary Policy*, pp. 319–367. University of Chicago Press.

Corsetti, G., L. Dedola, and S. Leduc (2008a). High exchange-rate volatility and low pass-through. *Journal of Monetary Economics 55*(6), 1113–1128.

Corsetti, G., L. Dedola, and S. Leduc (2008b). International risk sharing and the transmission of productivity shocks. *Review of Economic Studies 75*(2), 443–473.

Corsetti, G. and P. Pesenti (2005). International dimensions of optimal monetary policy. *Journal of Monetary economics 52*(2), 281–305.

Cox, P. M., D. Pearson, B. B. Booth, P. Friedlingstein, C. Huntingford, C. D. Jones, and C. M. Luke (2013). Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. *Nature 494*(7437), 341–344.

Crowley, M., H. Song, and N. Meng (2016). Tariff scares: trade policy uncertainty and foreign market entry by chinese firms. Discussion Papers 111722, Centre of Economic Policy Research.

De Blas, B. and K. N. Russ (2015). Understanding markups in the open economy. *American Economic Journal: Macroeconomics 7*(2), 157–180.

Dornbusch, R. (1987). Exchange rates and prices. *The American Economic Review 77*(1), 93–106.

Drozd, L. A. and J. B. Nosal (2012). Understanding international prices: Customers as capital. *The American Economic Review 102*(1), 364–395.

Eaton, J. and S. Kortum (2002). Technology, geography, and trade. *Econometrica 70*(5), 1741–1779.

Elith, J., J. R. Leathwick, and T. Hastie (2008). A working guide to boosted regression trees. *Journal of Animal Ecology 77*(4), 802–813.

Fitzgerald, D. and S. Haller (2014). Pricing-to-market: evidence from plant-level prices. *The Review of Economic Studies 81*(2), 761–786.

Forbes, K., I. Hjortsoe, and T. Nenova (2015). The shocks matter: improving our estimates of exchange rate pass-through. Discussion Papers 43, Monetary Policy Committee Unit, Bank of England.

Freund, Y. and R. E. Schapire (1996). Experiments with a new boosting algorithm. In *Proceedings of the Thirteenth International Conference on Machine Learning (ICML'96)*, pp. 148–156. Morgan Kaufmann.

Friedman, J., T. Hastie, and R. Tibshirani (2001). *The elements of statistical learning*, Volume 1. Springer series in statistics Springer, Berlin.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis 38*(4), 367–378.

Gagnon, J. E. and M. M. Knetter (1995). Markup adjustment and exchange rate fluctuations: evidence from panel data on automobile exports. *Journal of International Money and Finance 14*(2), 289–310.

Ghironi, F. and M. Melitz (2005). International trade and macroeconomic dynamics with heterogeneous firms. *The Quarterly Journal of Economics 120*(3), 865–915.

Goldberg, L. S. and J. M. Campa (2005). Exchange rate pass-through into import prices. *Review of Economics and Statistics 87*(4), 679–690.

Goldberg, L. S. and J. M. Campa (2006). Differences in exchange rate pass-through in the euro area. *European Economic Review 50*(1), 121–145.

Goldberg, L. S. and J. M. Campa (2010). The sensitivity of the cpi to exchange rates: Distribution margins, imported inputs, and trade exposure. *The Review of Economics and Statistics 92*(2), 392–407.

Goldberg, P. and F. Verboven (2001). The evolution of price dispersion in the european car market. *The Review of Economic Studies 68*(4), 811–848.

Goldberg, P. K. and M. M. Knetter (1997). Goods prices and exchange rates: what have we learned? *Journal of Economic Literature 35*(3), 1243–1272.

Goldberg, P. K. and F. Verboven (2005). Market integration and convergence to the law of one price: evidence from the european car market. *Journal of international Economics 65*(1), 49–73.

Gopinath, G. (2015). The international price system. *National Bureau of Economic Research*.

Gopinath, G., O. Itskhoki, and R. Rigobon (2010). Currency choice and exchange rate pass-through. *The American Economic Review 100*(1), 304–336.

Gopinath, G. and R. Rigobon (2008). Sticky borders. *The Quarterly Journal of Economics*, 531–575.

Halpern, L., M. Koren, A. Szeidl, et al. (2011). Imported inputs and productivity. *The American Economic Review, R&R 2*(3), 9.

Hancock, T., T. Jiang, M. Li, and J. Tromp (1996). Lower bounds on learning decision lists and trees. *Information and Computation 126*(2), 114–122.

Heathcote, J. and F. Perri (2002). Financial autarky and international business cycles. *Journal of Monetary Economics 49*(3), 601–627.

Hellerstein, R. (2008). Who bears the cost of a change in the exchange rate? pass-through accounting for the case of beer. *Journal of international economics 76*(1), 14–32.

Hyafil, L. and R. L. Rivest (1976). Constructing optimal binary decision trees is np-complete. *Information processing letters 5*(1), 15–17.

Kleinberg, J., J. Ludwig, S. Mullainathan, and Z. Obermeyer (2015). Prediction policy problems. *The American economic review 105*(5), 491–495.

Knetter, M. M. (1989). Price discrimination by us and german exporters. *The American Economic Review 79*(1), 198–210.

Knetter, M. M. (1993). International comparisons of pricing-to-market behavior. *The American Economic Review 83*(3), 473–486.

Krugman, P. (1986, May). Pricing to market when the exchange rate changes. Working Paper 1926, National Bureau of Economic Research.

Lu, Y., Z. Tao, and Y. Zhang (2013). How do exporters respond to antidumping investigations? *Journal of International Economics 91*(2), 290–300.

Manova, K. and Z. Zhang (2012). Export prices across firms and destinations*. *The Quarterly Journal of Economics 127*(1), 379.

Marazzi, M., N. Sheets, R. Vigfusson, J. Faust, J. Gagnon, J. Marquez, R. Martin, T. Reeve, and J. Rogers (2005). Exchange rate pass-through to us import prices: some new evidence. *International Finance Discussion Papers*.

Marquez, J. (1990). Bilateral trade elasticities. *The Review of Economics and Statistics*, 70–77.

Marston, R. (1990). Pricing to market in japanese manufacturing. *Journal of International Economics 29*(3), 217–236.

Mayer, T., M. J. Melitz, and G. I. Ottaviano (2014). Market size, competition, and the product mix of exporters. *The American Economic Review 104*(2), 495–536.

Melitz, M. J. and G. I. Ottaviano (2008). Market size, trade, and productivity. *The review of economic studies 75*(1), 295–316.

Mendoza, E. G. (1991). Real business cycles in a small open economy. *The American Economic Review*, 797–818.

Mumtaz, H., Ö. Oomen, and J. Wang (2008). Exchange rate pass-through into uk import prices. *Bank of Englind Working Paper No.312*.

Obstfeld, M. and K. Rogoff (1995). Exchange rate dynamics redux. *The Journal of Political Economy 103*(3), 624–660.

Obstfeld, M. and K. Rogoff (2000). New directions for stochastic open economy models. *Journal of international economics 50*(1), 117–153.

Olley, S. and A. Pakes (1996). The dynamics of productivity in the telecomunications equipment industry. *Econometrica 64*, 1263–97.

Randall, C. and R. Van Woesik (2015). Contemporary white-band disease in caribbean corals driven by climate change. *Nature Climate Change 5*(4), 375–379.

Ravn, M., S. Schmitt-Grohé, and M. Uribe (2006). Deep habits. *The Review of Economic Studies 73*(1), 195–218.

Ridgeway, G. (2007). Generalized boosted models: A guide to the gbm package. *R "gbm" Package User Manual*.

Rogoff, K. (1996). The purchasing power parity puzzle. *Journal of Economic literature 34*(2), 647–668.

Rokach, L. and O. Maimon (2005). Top-down induction of decision trees classifiers-a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 35*(4), 476–487.

Ruhl, K. J. (2008). The international elasticity puzzle. *Unpublished scripts*.

Schmitt-Grohé, S. and M. Uribe (2003). Closing small open economy models. *Journal of international Economics 61*(1), 163–185.

Simonovska, I. and M. E. Waugh (2014). The elasticity of trade: Estimates and evidence. *Journal of international Economics 92*(1), 34–50.

Stockman, A. C. and L. L. Tesar (1995). Tastes and technology in a two-country model of the business cycle: Explaining international comovements. *The American Economic Review*, 168–185.

Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE transactions on neural networks 10*(5), 988–999.

Varian, H. R. (2014). Big data: New tricks for econometrics. *The Journal of Economic Perspectives 28*(2), 3–27.

Wager, S. and S. Athey (2015). Estimation and inference of heterogeneous treatment effects using random forests. *arXiv preprint arXiv:1510.04342*.