Firm Level Pass Through: A Machine Learning Approach

Lu Han
University of Cambridge

29 Nov 2017

Introduction

- A consistent finding of the exchange rate pass through (ERPT) literature is that export prices denominated in destination currency do not react one-to-one to bilateral exchange rate movements.
- Among many theoretical and empirical studies, one approach is to use micro-level firm data to understand ERPT.
- This paper proposes an innovative method to systematically detect features affecting ERPT and predict ERPT at the firm-level.
- The proposed method works with micro-founded macro models and is directly applicable to highly dis-aggregated large scale custom datasets.

Exchange Rate Pass Through

The price of an exporter denominated in the destination currency at the border can be expressed as

$$p_{dt}^{destination} = MKP_{dt} \cdot e_{dt} \cdot MC_t$$

- MKP_{dt} is the exporter's markup at destination d.
- e_{dt} is the bilateral exchange rate between the origin country and destination d ($\uparrow$ means origin appreciation).
- MC_t is the marginal cost of the firm.

Alternatively, write into the price denominated in the origin currency.

$$p_{dt}^{origin} = MKP_{dt} \cdot MC_t$$

Exchange rate pass through to the import price at the destination d is defined as

$$\beta_{dt} := \frac{\partial log(p_{dt}^{destination})}{\partial log(e_{dt})}$$

Empirical Findings

Macro level (the puzzle):

1. ERPT is far from complete;
   - Campa and Goldberg (2005) estimate $\beta = 0.46$ [one month] and $\beta = 0.64$ [4 months] for OECD countries.
2. different across industries and countries;
3. changes over different time horizons;
   - Marazzi et al. (2005); Gust, Leduc and Vigfusson (2010)

Micro firm-level studies addressing the puzzle find evidences cannot be easily reconciled with macro findings:

1. more complete ERPT using unit values and firm level dataset.
   - $\beta = 1 - 0.08 = 0.92$, Berman, Martin, and Mayer (2012), France.
   - $\beta = 1 - 0.20 = 0.80$, Amiti, Itskhoki, and Konings (2014), Belgium.
   - $\beta = 1 - 0.05 = 0.95$, Corsetti, Crowley, Han, and Song (WP), China.
2. Lewis (2016) discusses the reconcilability of ERPT estimates for UK import prices.

Missing Components

1. The source of the exchange rate shock matters.
   - Theory: Corsetti, Dedola, and Leduc (2008)
   - Empirical: Forbes, Hjortsoe, Nenova (2015)
2. Heterogeneity in ERPT
   - Competition and market structure matters. ERPT is U-shaped in market share: Dornbusch (1987), Atkeson and Burstein (2008), Auer and Schoenle (2016)
   - Other stuff matters. Lower ERPT if more imported inputs; higher productivity; higher distribution margin; low quality; invoiced in international currencies; differentiated products; general trade v.s. processing materials; trading v.s. non-trading exporters;
3. Not only bilateral but also multi-lateral exchange rates.
   - Trade flows: Bown and Crowley (2007)
4. Nominal rigidities.
   - Conditional on price changes: Gopinath and Itskhoki (2010), Fitzgerald and Haller (2013)

A method customized for big data

- Unlike micro studies in other fields, international trade firm-level datasets recently made available contain a significant portion of firms in an economy and almost all custom transactions at firm product (8-digits) level in a given period.
- Although significant progress has been made in understanding the heterogeneity of ERPT, the literature fails to make scenario dependent predictions based on the macro environment and observed firm, product and market structure features.

This paper tries to construct a method to systematically detect features affecting ERPT and predict ERPT at the firm-level.

Challenges

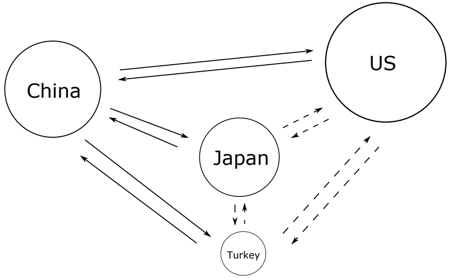


Figure 1 A multi-country trade model with multilateral exchange rate shocks

1. Unobserved variables
    - marginal cost of firms
    - trade flows between trade partners of China
2. Nonlinearity
    - ERPT may be a nonlinear function of firm characteristics and market structures, e.g. U-shaped in market share.

This paper

- Proposes an innovative method that estimates determinants of ERPT at the firm-level accounting for
    - nonlinearity
    - unobserved variables not varying at all dimensions.
- Extends the two-country Atkeson and Burstein (2008) model into a multi-country framework and studies how markets reach equilibrium under multilateral exchange rate shocks.
- Using the simulated data from my model, I test the accuracy of the method in identifying ERPT at the firm-level.
- Applying my method with China's custom data, I document evidences on the nonlinear relationship between market shares and ERPT.

Outline

1. Question
2. Algorithm
3. Tested on a Multi-country Atkeson and Burstein (2008) Model
4. Conclusions

A general pricing equation for exports

$$
\begin{aligned}
p_{i,f,d,t} \;=\; g(&\mathcal{F}_i,\ \mathcal{F}_f,\ \mathcal{F}_d,\ \mathcal{F}_t,\\
&\mathcal{F}_{i,f},\ \mathcal{F}_{i,d},\ \mathcal{F}_{i,t},\ \mathcal{F}_{f,d},\ \mathcal{F}_{f,t},\ \mathcal{F}_{d,t},\\
&\mathcal{F}_{f,d,t},\ \mathcal{F}_{i,d,t},\ \mathcal{F}_{i,f,t},\ \mathcal{F}_{i,f,d},\\
&\mathcal{F}_{i,f,d,t},\ \varepsilon_{i,f,d,t})
\end{aligned}
$$

Where i for product, f for firm, d for destination.

- $\mathcal{F}_i$: product specific factors/properties, etc.
- $\mathcal{F}_{d,t}$: exchange rate, CPI, and other macro variables.
- $\mathcal{F}_{i,f,t}$: marginal cost, etc.
- $\mathcal{F}_{f,d,t}$: firm destination market share, etc.
- $\mathcal{F}_{i,f,d,t}$: firm-product destination market share, etc.

Question

$$p_{\mathcal{I}} = g\left(e_{d,t}, X_{\mathcal{I}}, M_{\mathcal{Q}}, \epsilon_{\mathcal{I}}\right)$$

where

- $\mathcal{I} = \{i, f, d, t\}$; $\mathcal{Q} \subset \mathcal{I}$
- g: an unknown function
- p: the exporter's price
- e: the bilateral exchange rate
- X: a vector of observed feature variables
- M: a vector of unobserved variables that correlate with e
- ϵ: the error term that does not correlate with e

Objective:

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

Outline

1. Question
2. Algorithm
3. Tested on a Multi-country Atkeson and Burstein (2008) Model
4. Conclusions

An algorithm to learn the behavior of exporters

- Machine Learning in Economics:
  - Hal R. Varian introduced various machine learning methods in 2014.
  - Athey and Imbens (2015), Bajari, et al. (2015), Chernozhukov, Hansen, and Spindler (2015), Kleinberg, et al. (2015), Wager and Athey (2015), Athey and Imbens (2016), Chernozhukov, et al. (2016)
- Gradient Boosting Regression Trees (GBRT):
  - Recent applications of GBRT:
    - Bhatt, et al. (2013, Nature); Pearson, et al. (2014, Nature Climate Change); Doench, et al (2016, Nature Biotechnology) , etc.
  - High precision in solving practical estimation problems:
    - the main estimation method of winners of many international data science competitions, e.g. Kaggle competitions.
- Critiques of machine learning methods:
  - Completely data driven
  - Good at making predictions but does not identify causal relationships and enhance our economic understanding.

My contribution to the ML literature

- The proposed algorithm represents an attempt to control unobserved variables by feeding additional structural information implied by economic models into the machine learning algorithm.
  - no longer completely data driven but relies on the supplied external structural information.
- The core:
  1. I show a unique property of tree based algorithms that can be exploited to control unobserved components.
  2. In a multi-dimensional panel, parameter estimates from a structural model in a range of limited-dimensional spaces can help to restrain the variation of unobserved components.

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

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

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

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

3. Approximating p. Run gradient boosting regression tree algorithm entering coefficients $\{b_t^0, b_t^1, b_d^0, b_d^1\}$ as additional feature variables. Obtain

$$g_1 : (e_{d,t}, ms_{d,t}, b_t^0, b_t^1, b_d^0, b_d^1) \rightarrow p_{d,t}$$

4. Numerical differentiation. Use g_1 to construct counter-factual predictions **conditional on the values of** $ms_{d,t}, b_t^0, b_t^1, b_d^0, b_d^1$ and calculate:
   - $p_{d,t}^{Est1} = g_1(e_{d,t}, ms_{d,t}, b_t^0, b_t^1, b_d^0, b_d^1)$
   - $p_{d,t}^{Est2} = g_1(e_{d,t} + std(e_{d,t}), ms_{d,t}, b_t^0, b_t^1, b_d^0, b_d^1)$
   - $\beta_{d,t}^{Est} = \frac{p_{d,t}^{Est2} - p_{d,t}^{Est1}}{std(e_{d,t})}$

5. 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

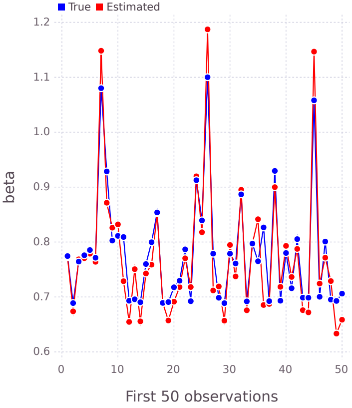
$$g_2 : (e_{d,t}, ms_{d,t}, b_t^0, b_t^1, b_d^0, b_d^1) \rightarrow \beta_{d,t}^{Est}$$

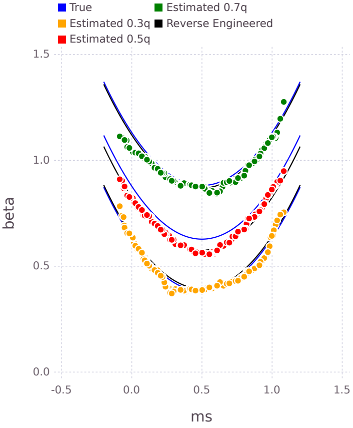
▸ full algorithm ▸ Explanation on the mechanisms

The Proposed Algorithm

$$p_{d,t} = 10 + \underbrace{\left[(ms_{d,t} - 0.5)^2 + mc_t\right]}_{beta} e_{d,t} + ms_{d,t} + mc_t + \epsilon_{d,t}$$

SSR = 83.71, Error Rate = 0.54%





spec2 | spec3 | spec1 Dummies | spec2 Dummies | spec3 Dummies | spec1 NoAddiInfo | spec2 NoAddiInfo | spec3 NoAddiInfo

Outline

A Multi-country Atkeson and Burstein (2008) Model

- Three countries
- Market structure assumptions as in Atkeson and Burstein (2008)
- Large number of sectors and N firms competing in each sector
- Only the best firm in each sector exports.
- N-2 number of destination competitors
- Productivity distributions are different across sectors, countries.
- Simplification:
  - No imported inputs
  - No nominal rigidities
  - No distribution cost

▸ Details of Model Specifications

Recover ERPT for Simulated Exporters

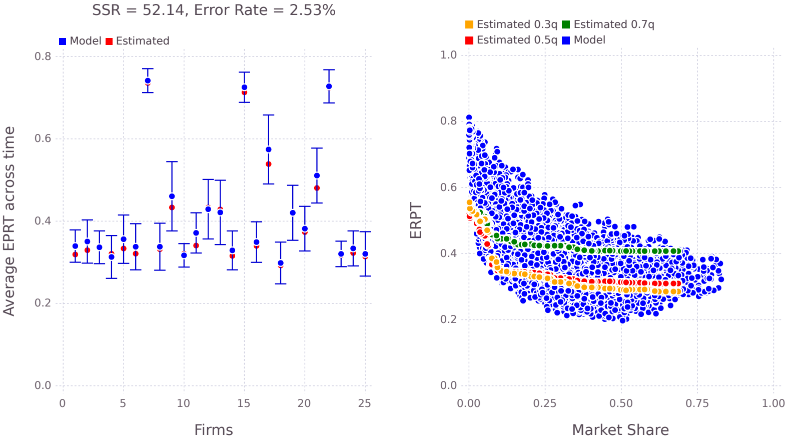
- Simulate the model for 240 periods (20 years).
- Construct an environment similar to my empirical dataset:
  1. Select an origin country
  2. Record trade flows of exports
     $f, s, d, t, p_{f,s,d,t}, q_{f,s,d,t}, P_{s,d,t}, D_{s,d,t}, e_{d,t}, P_{d,t}, L_{d,t}, C_{d,t}$

Objective: Use only information from recorded variables to

1. Learn trade and pricing patterns of exporters.
2. Given market conditions at $t-1$, estimate price changes under a bilateral exchange rate shock at period t.
3. Recover ERPT for simulated firms.

Case 1: Only Exchange Rate Shocks

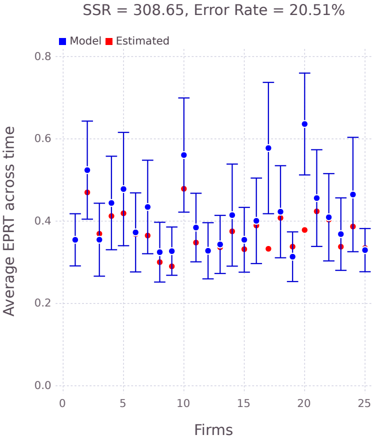
My algorithm compared to true counter-factual environments

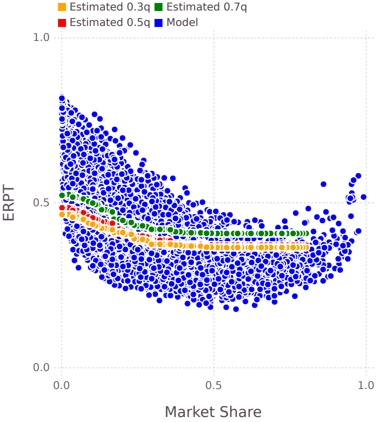


▸ case 1 point estimates

Case 2: Add productivity shocks

My algorithm compared to true counter-factual environments





▸ case 2 point estimates

Outline

Conclusions

- Proposes an innovative method that works with micro-founded macro models and is directly applicable to highly dis-aggregated large scale custom datasets. It estimates determinants of ERPT at the firm-level accounting for
    - nonlinearity
    - unobserved variables not varying at all dimensions.
- Extends the two-country Atkeson and Burstein (2008) model into a multi-country framework and studies how markets reach equilibrium under multilateral exchange rate shocks.
- Using the simulated data from my model, I test the accuracy of the method in identifying ERPT at the firm-level.
- Applying my method with China's custom data, I document evidences on the nonlinear relationship between market shares and exchange rate pass through.

Appendix

Outline

5. Appendix
   - Model
   - Classification and Regression Trees
   - One Dimensional Examples
   - Drawbacks of Conventional Regression Approaches
   - Alternative Specifications
   - Full Algorithm

Firm's Problem

Cournot competition in quantities within sector s:

$$\max_{q_{f,s,o,d,t}} q_{f,s,o,d,t}(p_{f,s,o,d,t} - mc_{f,s,o,t}) \tag{1}$$

subject to

$$q_{f,s,o,d,t} = \left(\frac{p_{f,s,o,d,t}}{P_{s,d,t}}\right)^{-\rho_s} \left(\frac{P_{s,d,t}}{P_{d,t}}\right)^{-\eta} D_{d,t} \tag{2}$$

$$D_{s,d,t} \equiv \left[\sum_o \sum_{f\in s, f\in I_o} (q_{f,s,o,d,t})^{\frac{\rho_s - 1}{\rho_s}}\right]^{\frac{\rho_s}{\rho_s - 1}} \quad , \quad D_{d,t} \equiv \left[\sum_s (D_{s,d,t})^{\frac{\eta-1}{\eta}}\right]^{\frac{\eta}{\eta-1}}$$

$$P_{s,d,t} \equiv \left[\sum_o \sum_{f\in s, f\in I_o} (p_{f,s,o,d,t})^{1-\rho_s}\right]^{\frac{1}{1-\rho_s}} \quad , \quad P_{d,t} \equiv \left[\sum_s (p_{s,d,t})^{1-\eta}\right]^{\frac{1}{1-\eta}}$$

where $f, s, o, d, t = firm, sector, origin, destination, time$

Price, Market Share and Demand Elasticity

The usual pricing equation

$$p_{f,s,o,d,t} = \frac{\varepsilon_{f,s,o,d,t}(ms_{f,s,o,d,t})}{\varepsilon_{f,s,o,d,t}(ms_{f,s,o,d,t}) - 1} \frac{mc_{f,s,o,t}}{e_{d,t}}$$

Demand elasticity ε is a nonlinear function of market share.

$$\varepsilon = \frac{1}{\frac{1}{\rho}(1 - ms) + \frac{1}{\eta} ms}$$

where

$$ms = \frac{pq}{\sum_f pq} = \frac{p^{1-\rho}}{\sum_f (p)^{1-\rho}}$$

Subscripts f, s, o, d, t is omitted for simplicity.

Equilibrium Effect of Market Share Matters

$$\widehat{p}_{k,s,o,d,t} = \kappa_{k,s,o,d,t}\widehat{ms}_{k,s,o,d,t} + \widehat{mc}_{k,s,o,t} + \widehat{e}_{o,d,t}$$

where $\kappa_{f,s,o,d,t}$ is the price elasticity with respect to a firm's own market share:

$$\kappa_{f,s,o,d,t} \equiv \left(\frac{\varepsilon_{f,s,o,d,t}}{\varepsilon_{f,s,o,d,t}-1}\right)\left(-\frac{1}{\rho_s}+\frac{1}{\eta}\right)$$

Equilibrium effects of market share under a multi-country context:

$$\frac{\widehat{ms}_{k,s,o,d,t}}{\lambda_{f,s,o,d,t}} = (1-ms_{k,s,o,d,t})\left\{(1-\rho_s)\left[\widehat{mc}_{k,s,o,t}+\widehat{e}_{o,d,t}\right]\right\} - \sum_{o'}\sum_{f\neq k} ms_{f,s,o',d,t}\left\{(1-\rho_s)\left[\widehat{mc}_{f,s,o',t}+\widehat{e}_{o',d,t}-\kappa_{f,s,o',d,t}\widehat{ms}_{f,s,o',d,t}\right]\right\}$$

where $\lambda_{k,s,o,d,t}$ is the theoretical exchange rate pass through and it is U-shaped in market share:

$$\lambda_{f,s,o,d,t} = \frac{1}{1-(1-ms_{f,s,o,d,t})(1-\rho_s)\kappa_{f,s,o,d,t}}$$

The Theoretical U-shape and More

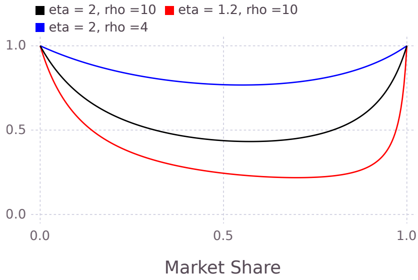
$$\widehat{p}_{k,s,o,d,t} = \lambda_{k,s,o,d,t} \left[\widehat{mc}_{k,s,o,t} + \widehat{e}_{o,d,t} - \kappa_{k,s,o,d,t} \widehat{CE}_{k,s,o,d,t} \right]$$

where $\widehat{CE}_{k,s,o,d,t}$ is the total effect of competitors' reactions.

$$\widehat{CE}_{k,s,o,d,t} = \sum_{o'} \sum_{f \neq k} s_{f,s,o',d,t}(1-\rho_s) \left[\widehat{mc}_{f,s,o',t} + \widehat{e}_{o',d,t} - \kappa_{f,s,o',d,t} \widehat{ms}_{f,s,o',d,t} \right]$$

The theoretical U-shape of $\lambda \rightarrow$

$$\lambda = \frac{1}{1 - \frac{(1-ms)(1-\rho)ms}{1 - \frac{1-ms}{\rho} - \frac{ms}{\eta}} (1/\eta - 1/\rho)}$$



Household's Problem

$$\max_{C_{d,t},L_{d,t}} E_0 \sum_{t=0}^{\infty} \beta^t U(C_{d,t}, L_{d,t})$$

subject to

$$U_{d,t} = log[C_{d,t}^{\mu}(1 - L_{d,t})^{1-\mu}]$$

As in Atkeson and Burstein (2008), household can trade a complete set of international assets.

$$P_{d,t}C_{d,t} + \sum_{o}\left[\sum_{\nu} p_{o,t}^{B}(\nu)B_{o,t}(\nu) - (1 + i_{o,t-1})B_{o,t-1}\right] * e_{o,d,t} = W_{d,t}L_{d,t} + \Pi_{d,t}$$

The optimal solution of household's problem are given by

$$\frac{1-\mu}{\mu}\frac{C_{d,t}}{1-L_{d,t}} = \frac{W_{d,t}}{P_{d,t}} \tag{3}$$

$$\frac{C_{o,t}P_{o,t}}{e_{o,d,t}C_{d,t}P_{d,t}} = \frac{C_{o,t+1}(\nu)P_{o,t+1}(\nu)}{e_{o,d,t+1}(\nu)C_{d,t+1}(\nu)P_{d,t+1}(\nu)} \tag{4}$$

Other Equilibrium Conditions

$$mc_{f,s,o,t} = \frac{W_{o,t}}{\Omega_{f,s,o,t}}$$

$$\Omega_{f,s,o,t} l_{f,s,o,t} = \sum_d q_{f,s,o,d,t}$$

$$\sum_{f,s} l_{f,s,o,t} = L_{o,t}$$

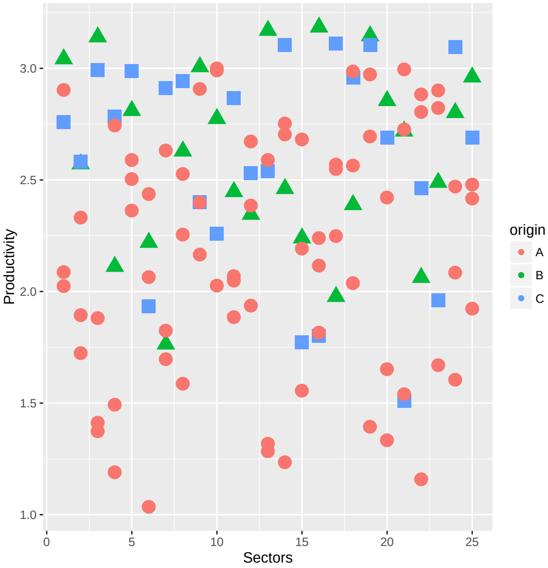
Steady state condition: balance of trade

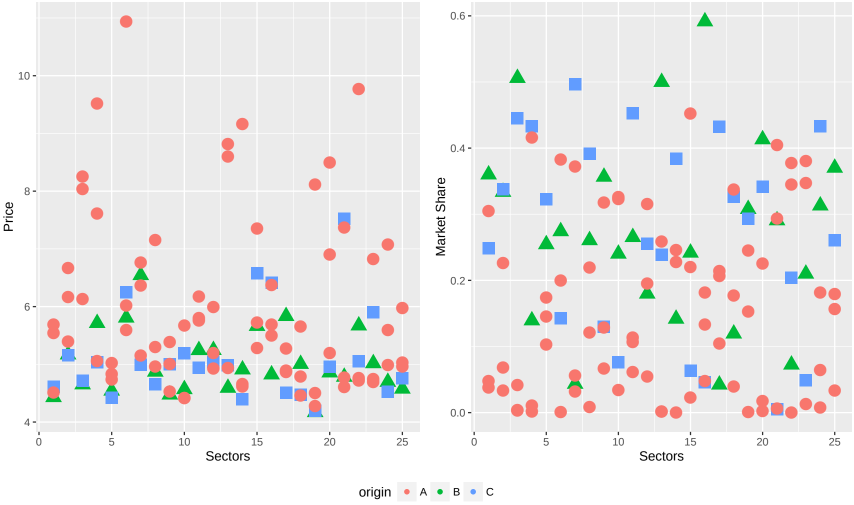
$$\sum_{f,s} p_{f,s,d,o} q_{f,s,d,o} = \sum_{f,s} p_{f,s,o,d} q_{f,s,o,d} * e_{o,d} \quad \textit{for } o \neq d$$

where $e_{o,d}$ is defined as units of currency o per units of currency d.

	Countries	S	N	ρ	η	$\Phi(\Omega)$
Benchmark	3	25	3+2	10	2	Uniform

Distribution of Simulated Productivities in Country A

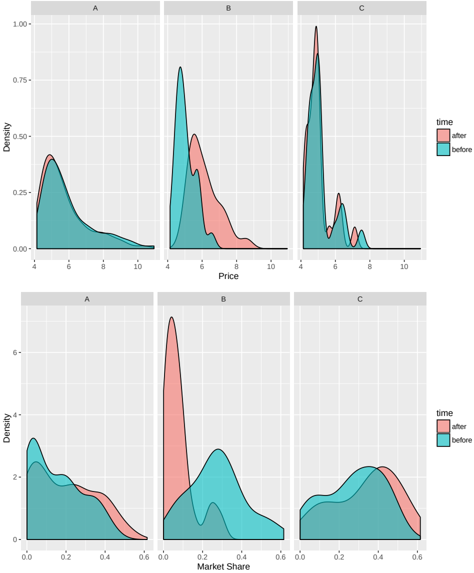



Price
10
8
6
4
0
5
10
15
20
25
Sectors
Market Share
0.6
0.4
0.2
0.0
0
5
10
15
20
25
Sectors
origin
A
B
C

Back

Appreciation of Country B

Change in Price and Market Share Dist in A



Estimation Procedure

1. Make assumptions based on the structural model. In general, the structural estimation can take any form. Here, I will use the linear approximation of the pricing equation:

$$log(p_{f,s,d,t}) = a + b * log(e_{d,t}) + c * log(p_{f,s,d,t-1})$$

2. Identify dimensions to be fixed as $\{s, t, d, sd\}$. Run regressions and collect coefficients $a_s, b_s, c_s, a_t, b_t, c_t, \ldots$
3. Create market share measure $ms_{fsdt,sdt} = \frac{p_{f,s,d,t} q_{f,s,d,t}}{P_{s,d,t} D_{s,d,t}}$
4. Run GBRT with the dependent variable $log(p_{f,s,d,t})$ on $log(e_{d,t})$ and feature variables $\mathcal{F}$

$$\begin{aligned}\mathcal{F} := \; & log(e_{d,t-1}), log(e_{d',t}), log(e_{d',t-1}), log(ms_{fsdt-1,sdt-1}), \\ & log(D_{s,d,t-1}), log(P_{s,d,t-1}), log(P_{d,t-1}), log(L_{d,t-1}), log(C_{d,t-1}), \\ & a_s, b_s, c_s, a_t, b_t, c_t, \ldots\end{aligned}$$

and obtain g_1

Estimation Procedure (cont.)

5. Numerical differentiation on predicted counter-factual bilateral exchange rates

$$p^{Est1}_{f,s,d,t} = g_1[log(e_{d,t}), \mathcal{F}]$$
$$p^{Est2}_{f,s,d,t} = g_1[log(e_{d,t-1}), \mathcal{F}]$$
$$ERPT^{Est}_{f,s,d,t} = \frac{log(p^{Est1}_{f,s,d,t}) - log(p^{Est2}_{f,s,d,t})}{log(e_{d,t}) - log(e_{d,t-1})}$$

6. Run GBRT again with supervisor $ERPT^{Est}_{f,s,d,t}$ on $log(e_{d,t})$, $\mathcal{F}$ and obtain g_2.

Case 1: Only Exchange Rate Shocks

- No financial market exchange rate arbitrage condition holds

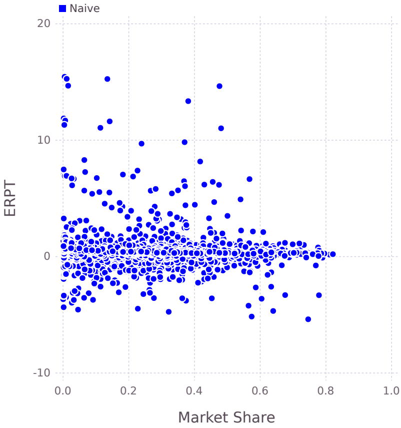
$$e_{1,2,t} = \frac{e_{1,3,t}}{e_{2,3,t}}$$

- Max 2 shocks (the third bilateral exchange rate is determined by the no arbitrary condition)

$$\begin{aligned} e_{1,2,t} &= \xi_{1,t} e_{1,2,ss} \\ e_{3,2,t} &= \xi_{3,t} e_{3,2,ss} \\ \xi_{i,t} &\sim U(0.8, 1.2) \end{aligned}$$

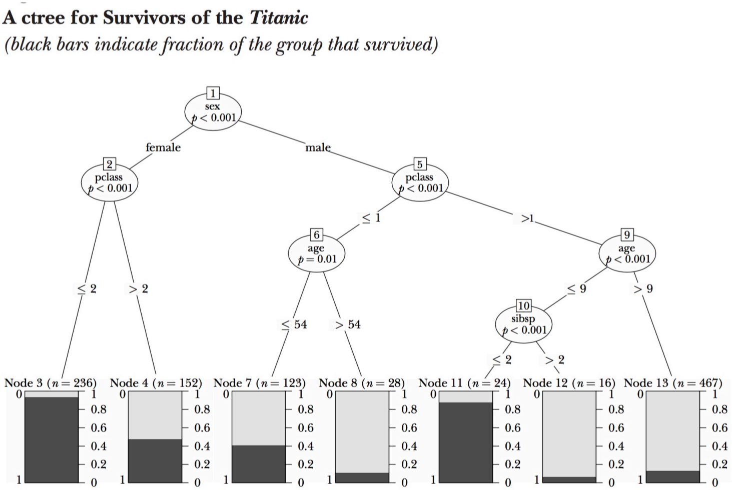
- Counter-factual state if there was no exchange rate shocks between 1 and 2

$$\begin{aligned} e^c_{1,2,t} &= e_{1,2,t-1} \\ e^c_{3,2,t} &= \xi_{3,t} e_{3,2,ss} \\ e^c_{1,3,t} &= \frac{e^c_{1,2,t}}{e^c_{3,2,t}} \end{aligned}$$



$$ERPT^{Naive}_{f,s,o,d,t} = \frac{log(p_{f,s,o,d,t}) - log(p_{f,s,o,d,t-1})}{log(e_{o,d,t}) - log(e_{o,d,t-1})}$$

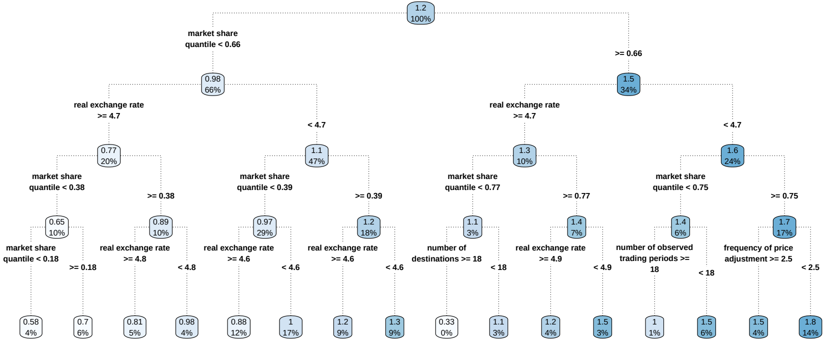
Illustration: Classification and Regression Trees



Source: Varian (2014)

Classification and Regression Trees

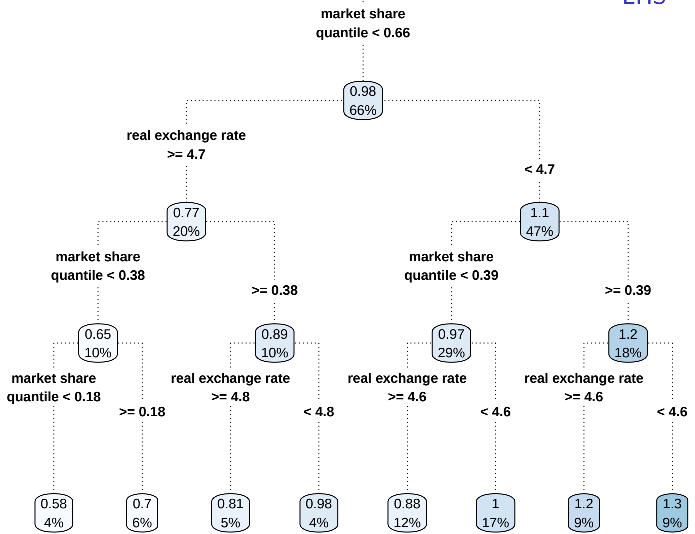
Predicting Export Price



Source: my own calculation from China's import and export database.

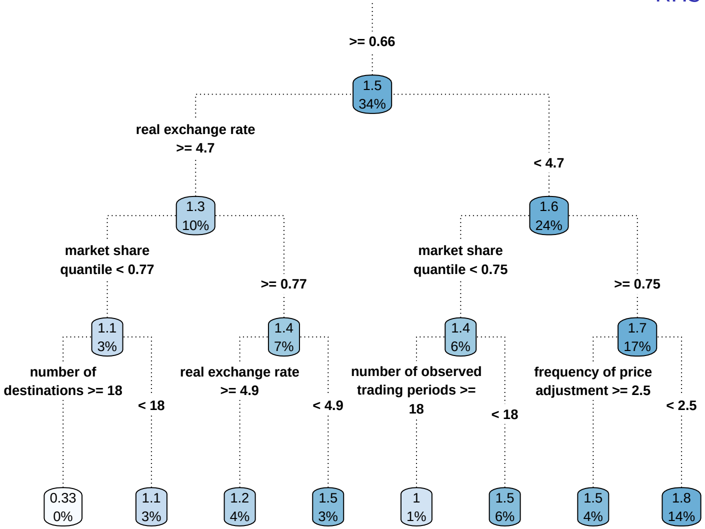
Classification and Regression Trees

LHS



Classification and Regression Trees

RHS



Gradient Boosting Models

Based on Friedman (2001), to minimize the objective function

$$\widehat{f}(\mathbf{x}) = arg \min_{f(\mathbf{x})} E_{y,\mathbf{x}} \Psi(y, f(\mathbf{x})) \tag{5}$$

1. a loss function (distribution Φ)
2. the number of iterations, `iter`
3. the depth of each tree, `inter.depth`
4. the shrinkage or the learning rate, `lr`
5. sampling rate (bagging fraction), `bf`

Gradient Boosting Models

Based on Friedman (2001), to minimize the objective function

$$\widehat{f}(\mathbf{x}) = arg \min_{f(\mathbf{x})} E_{y,\mathbf{x}} \Psi(y, f(\mathbf{x})) \tag{5}$$

1. a loss function (distribution Φ)
2. the number of iterations, `iter`
3. the depth of each tree, `inter.depth`
4. the shrinkage or the learning rate, `lr`
5. sampling rate (bagging fraction), `bf`

	Φ	`iter`	`inter.depth`	`lr`	`bf`
Benchmark	Normal	Cross Validation	8	0.005	0.5
Robustness	Laplace, Quantile	5000	1-10	0.01, 0.001	0.3, 1

Gradient Boosting Models (cont.)

Initialize $\widehat{f}(\mathbf{x})$ to be a constant.
For i in 1,..., $Iter_{max}$

1. Compute the negative gradient as the working response

$$h_i = -\frac{\partial}{\partial f(\mathbf{x_i})} \Psi(y_i, f(\mathbf{x_i})) \Big|_{f(\mathbf{x_i})=\widehat{f}(\mathbf{x_i})} \quad (6)$$

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

$$\widehat{f}(\mathbf{x}) \rightarrow \widehat{f}(\mathbf{x}) + \texttt{lr} * g(\mathbf{x}) \quad (7)$$

5. Repeat step 1-4 until $Iter_{max}$
6. Cross validation method to determine the optimal `iter`

Back

One dimensional example

Consider the case of identifying the individual treatment effect.

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

- where D_i is a treatment indicator and β_i is the treatment effect for individual i.
- The objective is to find β_i given data of individual outcomes y_i and its treatment indicator D_i.
- The data generating process (the functional form of each variable) is unknown. M_i is unobserved.

Suppose M_i is observed, we can estimate the individual treatment effect β_{it} using the following two-step procedure:

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

$$g_1 : (D_i, M_i) \rightarrow y_i$$

2. Use g_1 to construct counter-factual predictions **conditional on the value of** M_i and calculate individual treatment effect.

$$\beta_i^{Est} = g_1(1, M_i) - g_1(0, M_i)$$

- In most cases, we do not observe M_i. But it may be possible to have/create a variable $\mathfrak{M}_i$ that preserves some structural information of M_i.
- If we could construct counter-factuals conditional 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 by the alternative variable $\mathfrak{M}_i$ could be highly nonlinear.
- I find decision tree based algorithms have a unique advantage in addressing this type of problems.

One dimensional example (cont.)

Simulate 200 individuals:

$$y_i = \beta_i(M_i)D_i + M_i$$
$$\beta_i(M_i) = M_i$$

Table 1: Assignment of M_i

M_i	i
0	1-100
1	101-200

Table 2: Values of y_i

y_i	β_i	M_i	D_i
0	0	0	0
0	0	0	1
1	1	1	0
2	1	1	1

- the assignment of M_i is constructed to be ordered
- I want to utilize the information provided by the index i to estimate the individual treatment effect β_i.

A Basic Classification and Regression Tree (CART) Algorithm

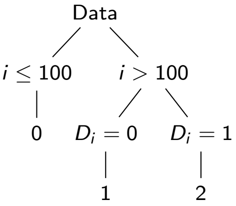
A CART algorithm recursively binary splits/partitions data at the point which minimizes the mean prediction error (MPE) measured by the loss function $h(.)$

$$MPE = \sum_{\tau \in leaves(T)} \sum_{i \in \tau} h(y - m_c)$$

$$m_c = \frac{1}{n_c} \sum_{i \in \tau} y_i$$

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 will reduce MPE as much as possible. If the largest decrease in MPE would be less than some threshold, or one of the resulting nodes would contain less than q points, stop. Otherwise, take that split, creating two new nodes.
3. In each new node, back to step 1.

Results

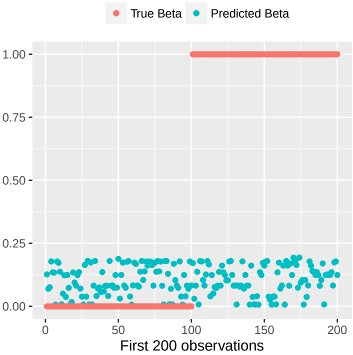


Stage 1

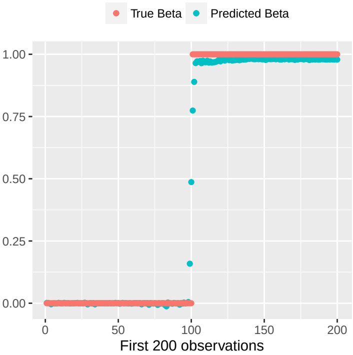
i	$\widehat{y}_i$ Evaluated at		Estimated β_i	True β_i
	$D_i = 1$	$D_i = 0$		
1-50	0	0	0	0
51-100	0	0	0	0
101-150	2	1	1	1
151-200	2	1	1	1

Stage 2

Example 1: the ordered case

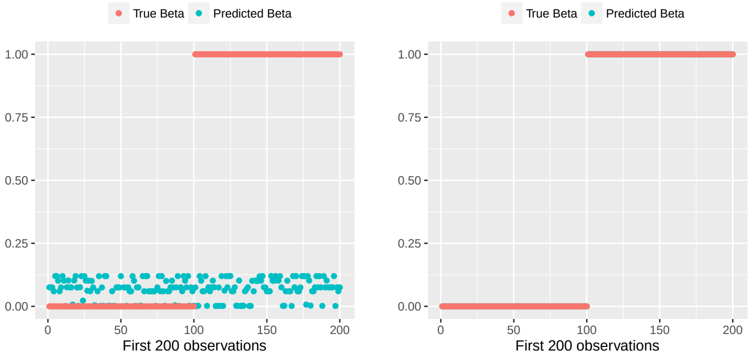


(a) No index variable



(b) Add index *i*

Example 1: the ordered case (cont.)



(c) Add dummy variables of i

(d) Add true β_i

The weak monotonic transformation property

Definition 1

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

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

Proposition 1

For a large number of observations n, entering $\{a_n\}$ as a feature variable is equivalent to entering any $\daleth(\{a_n\})$ in decision tree based algorithms.

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 gradient boosting regression algorithm.

Example 2: Weak and conditional weak monotonic transformation

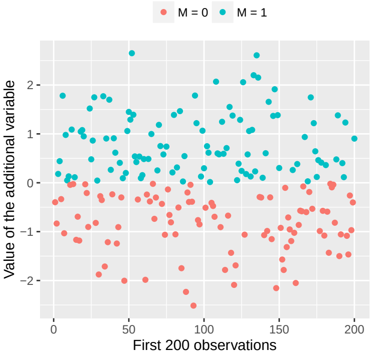
$$y_i = \beta_i(X_i, M_i)D_i + M_i$$
$$\beta_i(X_i, M_i) = X_i + M_i$$
$$D_i \in \{0,1\},\ X_i \in \{0,1\}$$

where M_i is randomly drawn from $\{0,1\}$ with equal probability for each individual i. y_i, X_i, D_i are observed variables. I experiment on the following two transformations of M_i:

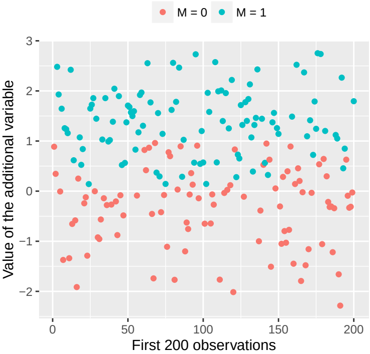
$$\mathfrak{M}_i^1 = \begin{cases} -|\epsilon_i| & if\ \ M_i = 0 \\ |\epsilon_i| & if\ \ M_i = 1 \end{cases};\qquad \mathfrak{M}_i^2 = \begin{cases} -|\epsilon_i| + X_i & if\ \ M_i = 0 \\ |\epsilon_i| + X_i & if\ \ M_i = 1 \end{cases};$$

Where $\epsilon_i \sim N(0,1)$

Weak (a) and conditional weak (b) monotonic transformations of M_i



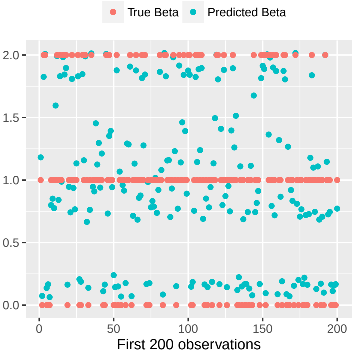
(a) Value of $\mathfrak{M}_i^1$



(b) Value of $\mathfrak{M}_i^2$

Estimates, n = 200

(a) Add $\mathfrak{M}_i^1$ as control



(b) Add $\mathfrak{M}_i^2$ as control

Estimates, n = 2000

(a) Add $\mathfrak{M}_i^1$ as control

(b) Add $\mathfrak{M}_i^2$ as control

[redacted]

Two dimensional examples

$$
\begin{aligned}
p_{d,t} &= 10 + \beta_{d,t} e_{d,t} + ms_{d,t} + mc_t + \epsilon_{d,t} \\
ms_{d,t} &= u_{d,t} + 0.1 e_{d,t} \\
mc_t &= u_t - 0.1 \overline{e_t}; \quad \overline{e_t} := \frac{\sum_d e_{d,t}}{n_d} \\
e_{d,t} \sim & N(0,1),\ u_{d,t} \sim Uniform(0,1),\ \epsilon_{d,t} \sim N(0,0.01)
\end{aligned}
$$

Simulate 3 different cases of the underlying ERPT function

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

Objective:

- $d, t, p_{d,t}, e_{d,t}, ms_{d,t}$ observed; mc_t unobserved.
- Estimate $\beta_{d,t}$ and understand what explains the heterogeneity of $\beta_{d,t}$.

It may be difficult to reach the correct regression specification, column (5)

Simulation Spec1: $p_{d,t} = \mu + \left[(ms_{d,t} - 0.5)^2 + mc_t\right] e_{d,t} + ms_{d,t} + mc_t + \epsilon_{d,t}$

	(1)	(2)	(3)	(4)	(5)
$e_{d,t}$	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)
Time FE	yes	yes	yes	yes	yes
Individual FE	yes	yes	yes	yes	yes
Adjusted R^2	0.663	0.865	0.722	0.865	0.879
Observations	80,000	80,000	80,000	80,000	80,000

Even at the correct regression specification column (5), results are not very informative about the underlying structure driving the heterogeneity of ERPT, $\beta_{d,t}$.

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

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

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

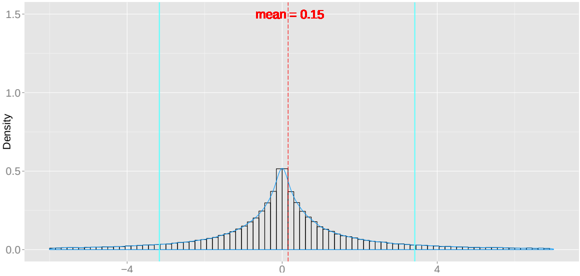
	Spec1	Spec2	Spec3
$e_{d,t}$	0.796***	0.242***	-0.088***
	(0.003)	(0.001)	(0.019)
$ms_{d,t}$	0.995***	1.002***	1.038***
	(0.005)	(0.002)	(0.016)
$e_{d,t} * ms_{d,t}$	-1.019***	-0.962***	0.301***
	(0.011)	(0.004)	(0.016)
$e_{d,t} * ms^2_{d,t}$	1.016***	0.957***	-0.190***
	(0.011)	(0.004)	(0.011)
Time FE	yes	yes	yes
Individual FE	yes	yes	yes
Adjusted R^2	0.879	0.885	0.056
Observations	80,000	80,000	80,000

Conventional estimation methods of ERPT trade-off between controlling unobserved variables and flexibility of functional forms

1. Interaction terms with $e_{d,t}$ and multiple fixed effects

$$p_{i,f,d,t} = \beta e_{d,t} + InteractionTerms + Controls + FEs + \epsilon_{i,f,d,t}$$

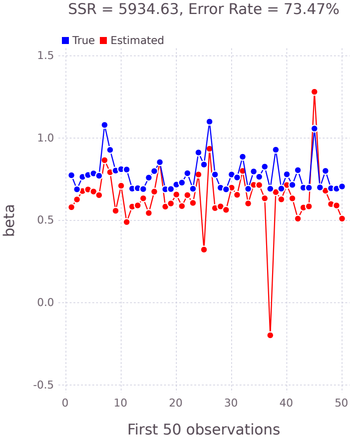
2. Or more flexible settings by categorizing data into several bins, e.g. by destinations, quantiles of market share, etc.

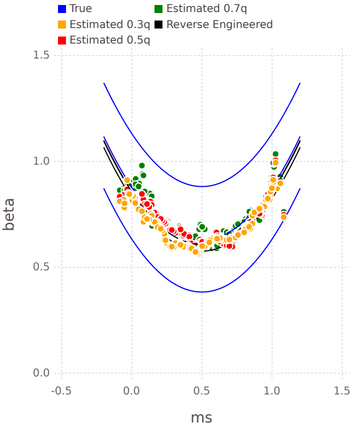


Distribution of Quarterly ERPT Estimates for China's Exporters by Firm-product Bins

Misalignment of beta due to unobserved mc_t

$$p_{d,t} = 10 + \underbrace{\left[(ms_{d,t} - 0.5)^2 + mc_t\right]}_{beta} e_{d,t} + ms_{d,t} + mc_t + \epsilon_{d,t}$$





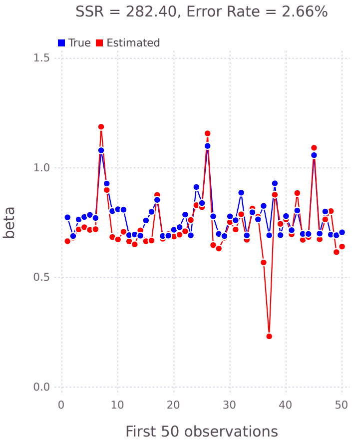
▸ Back

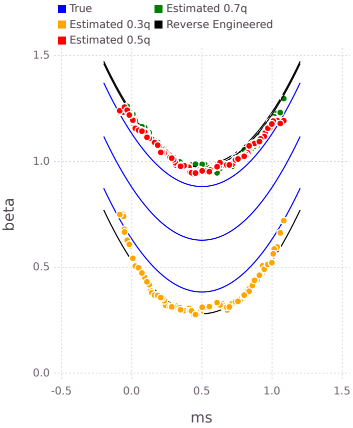
RHS Graph

1. True
   - Plot $f(ms) = (ms - 0.5)^2 + mc$
   - where $mc^q := quantile(mc, q)$ from data; $q \in [0.3, 0.5, 0.7]$
2. Estimated
   - Calculate q quantile of $e_{d,t}, b_t^0, b_t^1, b_d^0, b_d^1$
   - Plot $f(ms) = g_2((e_{d,t})^q, ms, (b_t^0)^q, (b_t^1)^q, (b_d^0)^q, (b_d^1)^q)$
3. Reverse Engineered
   1. Use $mc^q := quantile(mc, q)$ from reverse engineering; $q \in [0.3, 0.5, 0.7]$
      - Run GBRT with mc_t as the dependent variables on feature variables $e_{d,t}, b_t^0, b_t^1, b_d^0, b_d^1$ and get $M_3(.)$
      - Estimate $mc^q = M_3((e_{d,t})^q, (b_t^0)^q, (b_t^1)^q, (b_d^0)^q, (b_d^1)^q)$
   2. Plot $f(ms) = (ms - 0.5)^2 + mc$

Adding Control Dummies

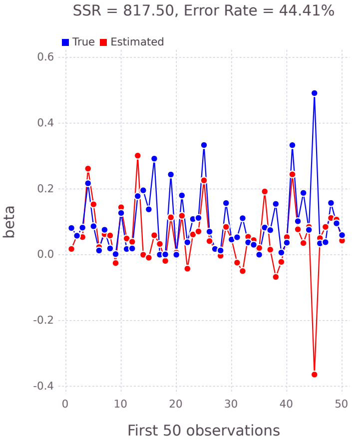
$$p_{d,t} = 10 + \underbrace{\left[(ms_{d,t} - 0.5)^2 + mc_t\right]}_{beta} e_{d,t} + ms_{d,t} + mc_t + \epsilon_{d,t}$$

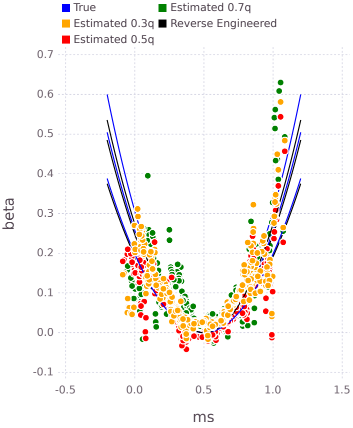




▸ Back

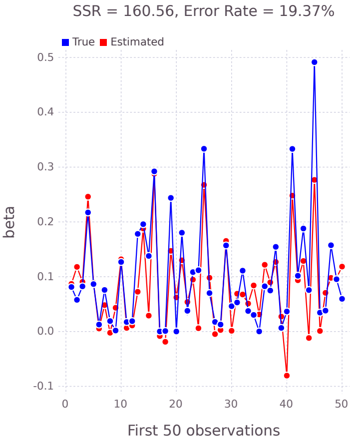
Spec2: Misalignment of beta due to unobserved mc_t

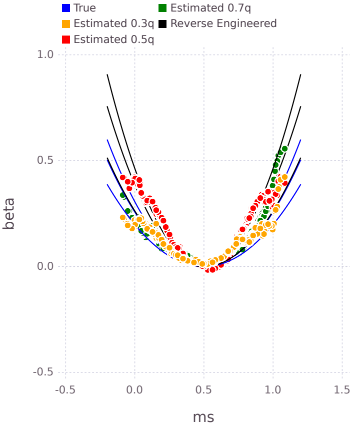




▸ Back

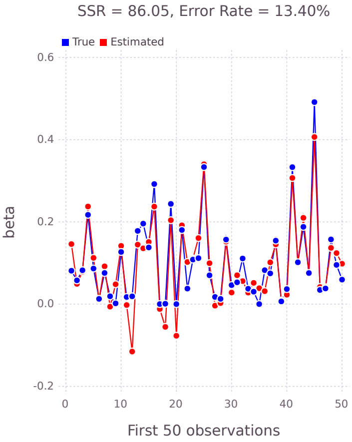
Spec2: Adding Control Dummies

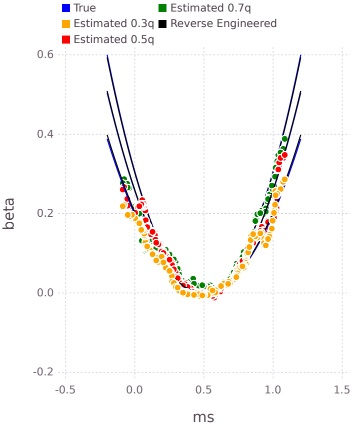




▸ Back

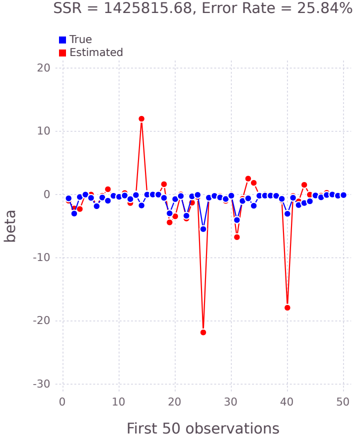
Spec2: The Proposed Algorithm

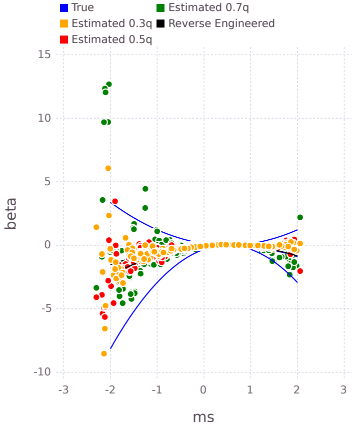




▸ Back

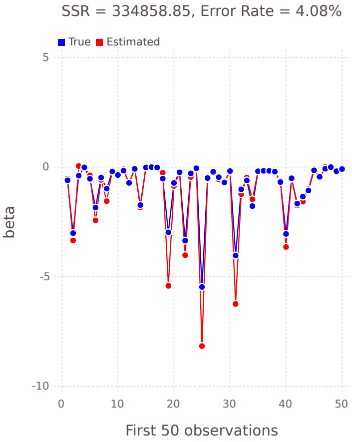
Spec3: Misalignment of beta due to unobserved mc_t

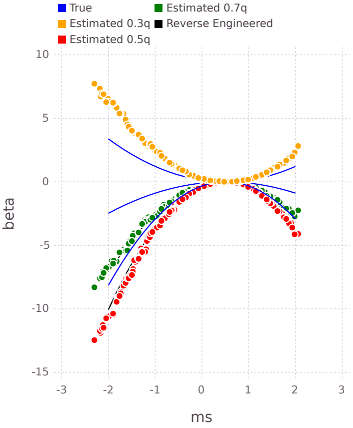




Back

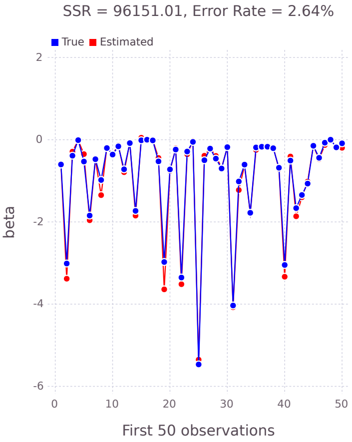
Spec3: Adding Control Dummies

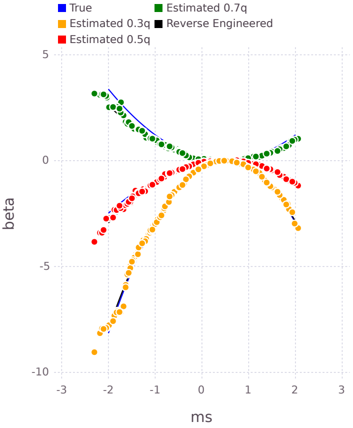




Back

Spec3: The Proposed Algorithm





▸ Back

High Nonlinearity

Setting:

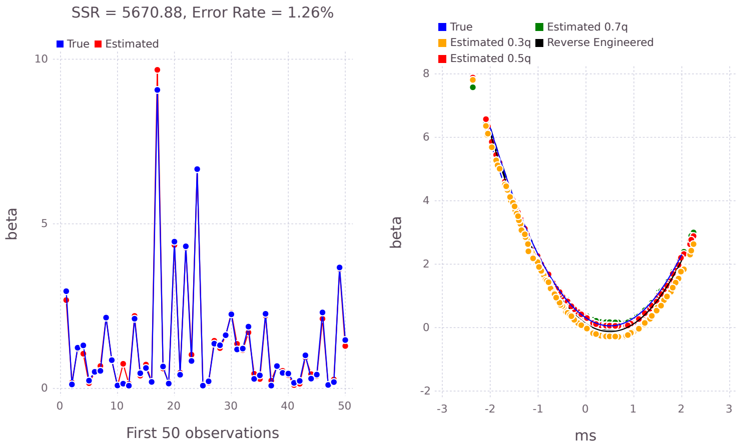
$$
\begin{aligned}
p_{d,t} &= 10 + \beta_{d,t} e_{d,t} + ms_{d,t} - mc_t + \epsilon_{d,t} \\
\beta_{d,t} &= (ms_{d,t} - 0.5)^2 + sin(1000 mc_t) mc_t \\
ms_{d,t} &= u_{d,t} + 0.1 e_{d,t} \\
mc_t &= u_t - 0.1 \overline{e_t} \\
\overline{e_t} &= \frac{\sum_d e_{d,t}}{n_d} \\
n_d &= 2000;\ n_t = 40 \\
u_{d,t} &\sim N(0,1),\ u_t \sim N(0,1),\ \epsilon_{d,t} \sim N(0,1)
\end{aligned}
$$

Objective:

- Read records $d, t, p_{d,t}, e_{d,t}, ms_{d,t}$
- Estimate $\beta_{d,t}$

High Nonlinearity

My Algorithm



▶ Back

Not Identifiable

Setting:

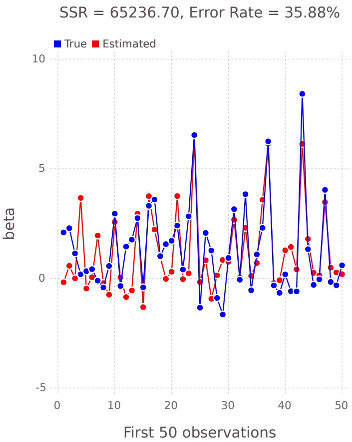
$$
\begin{aligned}
p_{d,t} &= 10 + \beta_{d,t} e_{d,t} + ms_{d,t} - mc_{d,t} + \epsilon_{d,t} \\
\beta_{d,t} &= (ms_{d,t} - 0.5)^2 + mc_{d,t} \\
ms_{d,t} &= u_{d,t} + 0.1 e_{d,t} \\
mc_{d,t} &= u_{d,t} - 0.1 e_{d,t} \\
n_d &= 2000;\ n_t = 40 \\
u_{d,t} &\sim N(0,1),\ u_t \sim N(0,1),\ \epsilon_{d,t} \sim N(0,1)
\end{aligned}
$$

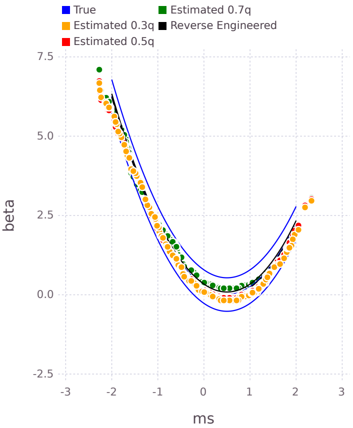
Objective:

- Read records $d, t, p_{d,t}, e_{d,t}, ms_{d,t}$
- Estimate $\beta_{d,t}$

Not Identifiable

My algorithm





▸ Back

Larger Correlation

Setting:

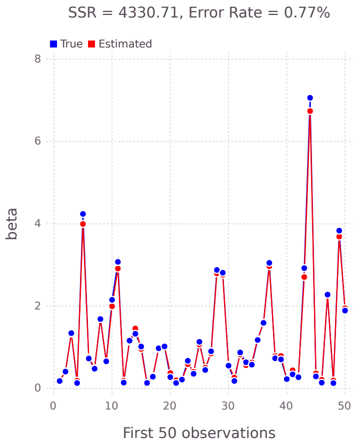
$$
\begin{aligned}
p_{d,t} &= 10 + \beta_{d,t} e_{d,t} + ms_{d,t} - mc_t + \epsilon_{d,t} \\
\beta_{d,t} &= (ms_{d,t} - 0.5)^2 + mc_t \\
ms_{d,t} &= u_{d,t} + 0.1 e_{d,t} \\
mc_t &= u_t - 1\overline{e_t} \\
\overline{e_t} &= \frac{\sum_d e_{d,t}}{n_d} \\
n_d &= 2000;\ n_t = 40 \\
u_{d,t} &\sim N(0,1),\ u_t \sim N(0,1),\ \epsilon_{d,t} \sim N(0,1)
\end{aligned}
$$

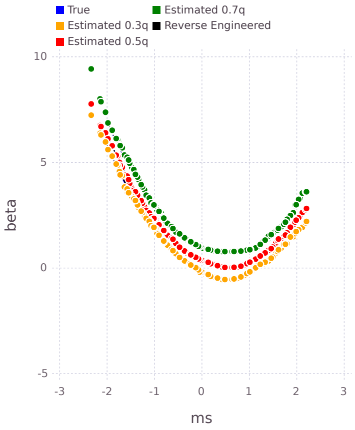
Objective:

- Read records $d, t, p_{d,t}, e_{d,t}, ms_{d,t}$
- Estimate $\beta_{d,t}$

Larger Correlation

My algorithm





▸ Back

Different Function of the Outer Part

Setting:

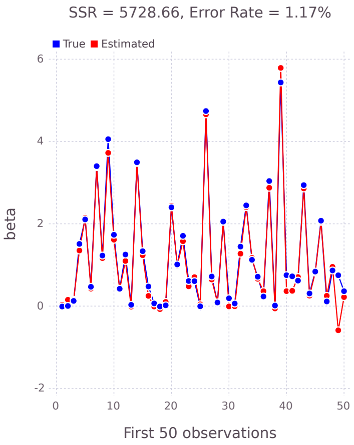
$$
\begin{aligned}
p_{d,t} &= 10 + \beta_{d,t} e_{d,t} + ms_{d,t} - mc_t^2 + \epsilon_{d,t} \\
\beta_{d,t} &= (ms_{d,t} - 0.5)^2 + mc_t \\
ms_{d,t} &= u_{d,t} + 0.1 e_{d,t} \\
mc_t &= u_t - 0.1 \overline{e_t} \\
\overline{e_t} &= \frac{\sum_d e_{d,t}}{n_d} \\
n_d &= 2000;\ n_t = 40 \\
u_{d,t} &\sim N(0,1),\ u_t \sim N(0,1),\ \epsilon_{d,t} \sim N(0,1)
\end{aligned}
$$

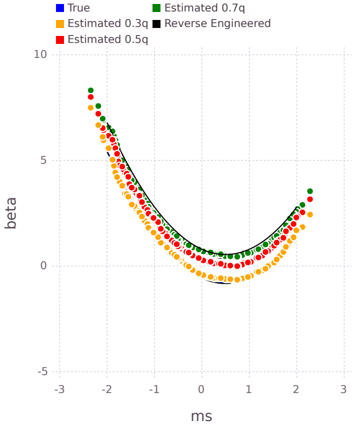
Objective:

- Read records $d, t, p_{d,t}, e_{d,t}, ms_{d,t}$
- Estimate $\beta_{d,t}$

Different Function of the Outer Part

My algorithm





▸ Back

Arellano and Bond

Setting:

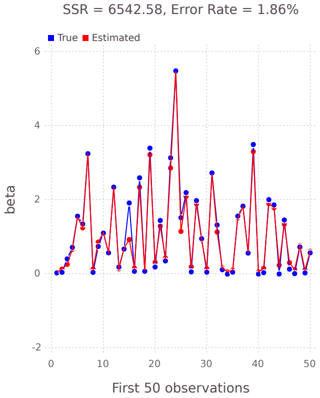
$$
\begin{aligned}
p_{d,t} &= 0.95 p_{d,t-1} + \beta_{d,t} e_{d,t} + ms_{d,t} - mc_t \\
\beta_{d,t} &= (ms_{d,t} - 0.5)^2 + mc_t \\
ms_{d,t} &= u_{d,t} + 0.1 e_{d,t} \\
mc_t &= u_t - 0.1 \overline{e_t} \\
\overline{e_t} &= \frac{\sum_d e_{d,t}}{n_d} \\
n_d &= 2000;\ n_t = 40 \\
u_{d,t} &\sim N(0,1),\ u_t \sim N(0,1),\ \epsilon_{d,t} \sim N(0,1)
\end{aligned}
$$

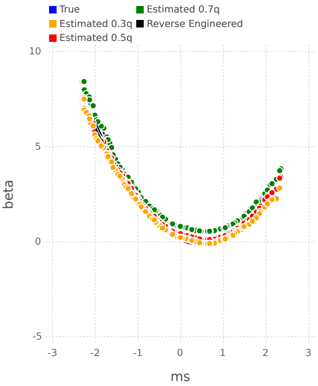
Objective:

- Read records $d, t, p_{d,t}, e_{d,t}, ms_{d,t}$
- Estimate $\beta_{d,t}$

Arellano and Bond

My Algorithm





▸ Back

Reduce Sample Size

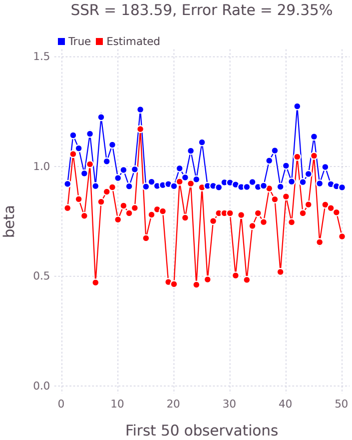
Setting:

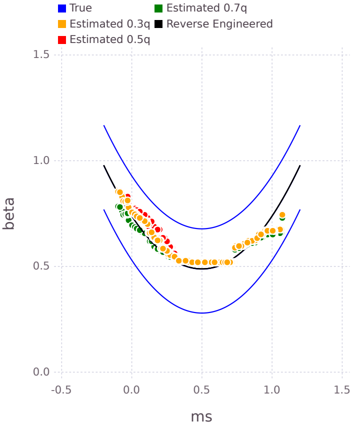
$$
\begin{aligned}
p_{d,t} &= 10 + \beta_{d,t} e_{d,t} + ms_{d,t} - mc_t + \epsilon_{d,t} \\
\beta_{d,t} &= (ms_{d,t} - 0.5)^2 + mc_t \\
ms_{d,t} &= u_{d,t} + 0.1 e_{d,t} \\
mc_t &= u_t - 0.1 \overline{e_t} \\
\overline{e_t} &= \frac{\sum_d e_{d,t}}{n_d} \\
n_d &= 200;\ n_t = 40 \\
u_{d,t} &\sim uniform,\ u_t \sim uniform,\ \epsilon_{d,t} \sim N(0,1)
\end{aligned}
$$

Objective:

- Read records $d, t, p_{d,t}, e_{d,t}, ms_{d,t}$
- Estimate $\beta_{d,t}$

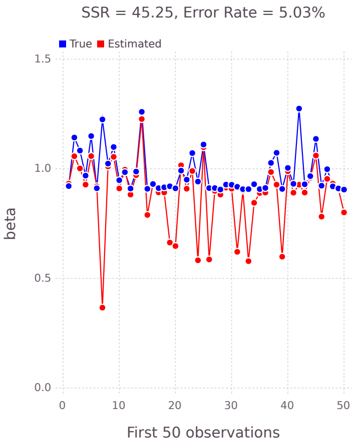
Reduce Sample Size
Dummies

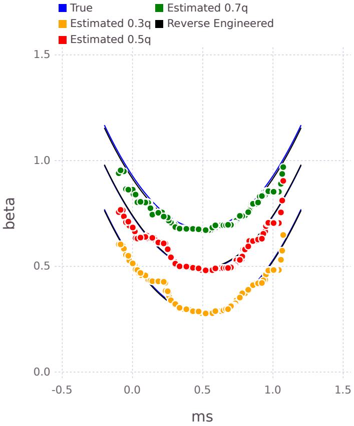




Reduce Sample Size

My Algorithm





▸ Back

Precision on Estimating Price

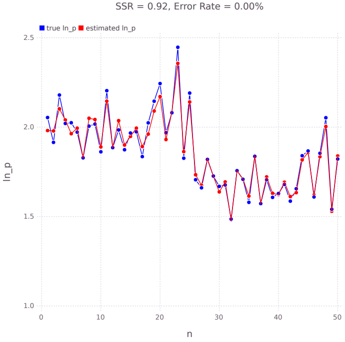


Figure 6: Decision Tree GBM

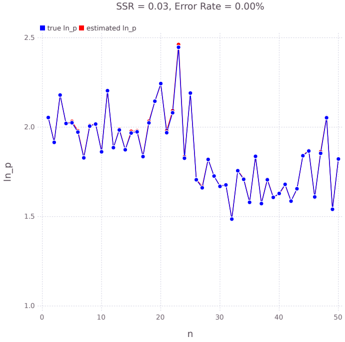
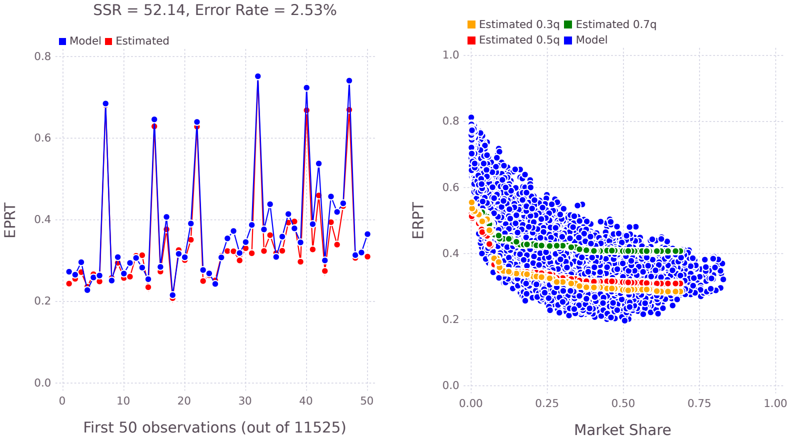


Figure 7: My Algorithm

Case 1: Only Exchange Rate Shocks

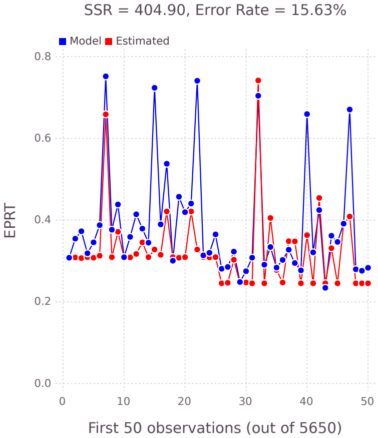
My algorithm compared to true counter-factual environments

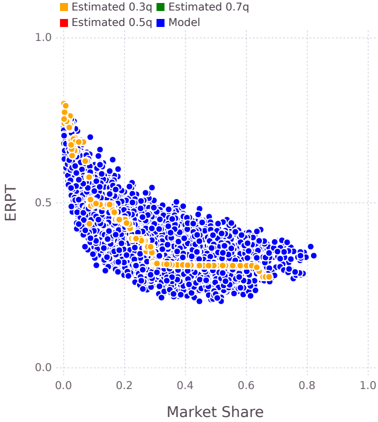


▸ case 1

Case 1: Only Exchange Rate Shocks

Without adding regression coefficients

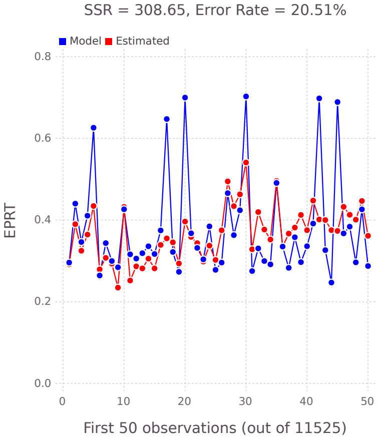


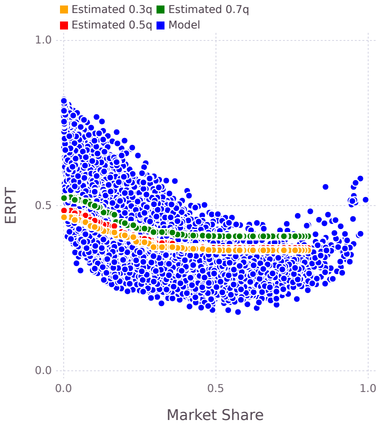


▸ case 1

Case 2: Add Productivity Shocks

My algorithm compared to true counter-factual environments

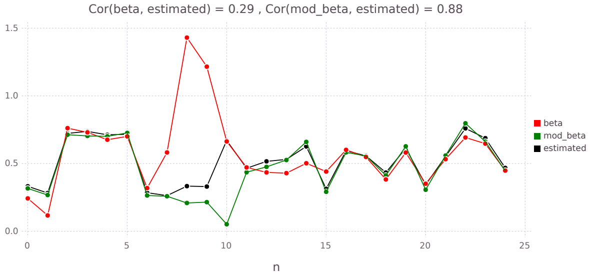




▸ case 2

Recover ERPT for Simulated Exporters

Case 3: add productivity shocks to the exporting country



Firm's productivity is assumed to follow a AR(1) process with a persistence of $\rho = 0.95$. 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 counter-factual equilibrium where there is no productivity shock in the next period. The black line represents ERPT estimates predicted by the algorithm.

More Experiments

1. High nonlinear function of the unobserved variable
2. Large correlation between the unobserved variable and exchange rates
3. Not identifiable case , unobserved variables vary over all dimensions
4. Arellano and Bond dynamic panel
5. Reduce sample size

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

Algorithm (cont.)

Observe dimensions 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 g_1.
$\mathbf{y^{Est1}} \leftarrow g_1(\mathbf{e}, \mathbf{X}, \mathbf{X_s}, \mathbf{coef_s^j})$
$\mathbf{y^{Est2}} \leftarrow g_1(\mathbf{e} + std(\mathbf{e}), \mathbf{X}, \mathbf{X_s}, \mathbf{coef_s^j})$
$\mathbf{beta^{Est}} \leftarrow \frac{\mathbf{y^{Est2}} - \mathbf{y^{Est1}}}{std(\mathbf{e})}$
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 g_2.

Output: $g_1, g_2, \mathbf{beta^{Est}}$ ▸ Back