Determinants of PTA Design

Insights from Machine Learning

Stepan Gordeev
UConn

Sandro Steinbach
NDSU

May 18, 2024

Midwest International Trade

Determinants of PTA Design

- Preferential Trade Agreements (PTAs) increasingly dominate international trade governance

Determinants of PTA Design

- Preferential Trade Agreements (PTAs) increasingly dominate international trade governance
  - **multi**lateral negotiations through WTO $\rightarrow$ **bi**lateral PTA negotiations

Determinants of PTA Design

- Preferential Trade Agreements (PTAs) increasingly dominate international trade governance
  - **multi**lateral negotiations through WTO $\rightarrow$ **bi**lateral PTA negotiations
- PTAs are becoming increasingly deep and varied in design

Determinants of PTA Design

- Preferential Trade Agreements (PTAs) increasingly dominate international trade governance
  - **multi**lateral negotiations through WTO $\rightarrow$ **bi**lateral PTA negotiations
- PTAs are becoming increasingly deep and varied in design
  - PTAs seek to improve market access

Determinants of PTA Design

- Preferential Trade Agreements (PTAs) increasingly dominate international trade governance
  - **multi**lateral negotiations through WTO $\rightarrow$ **bi**lateral PTA negotiations
- PTAs are becoming increasingly deep and varied in design
  - PTAs seek to improve market access
  - tariffs already low $\rightarrow$ # provisions, policy areas covered $\uparrow$

Determinants of PTA Design

- Preferential Trade Agreements (PTAs) increasingly dominate international trade governance
  - **multi**lateral negotiations through WTO $\rightarrow$ **bi**lateral PTA negotiations
- PTAs are becoming increasingly deep and varied in design
  - PTAs seek to improve market access
  - tariffs already low $\rightarrow$ # provisions, policy areas covered $\uparrow$
  - **non-tariff barriers**: antidumping, rules of origin, ...

Determinants of PTA Design

- Preferential Trade Agreements (PTAs) increasingly dominate international trade governance
  - **multi**lateral negotiations through WTO $\rightarrow$ **bi**lateral PTA negotiations
- PTAs are becoming increasingly deep and varied in design
  - PTAs seek to improve market access
  - tariffs already low $\rightarrow$ # provisions, policy areas covered $\uparrow$
  - **non-tariff barriers**: antidumping, rules of origin, ...
  - **behind-the-border policies**: IP protection, gov't procurement, ...

Determinants of PTA Design

- Preferential Trade Agreements (PTAs) increasingly dominate international trade governance
  - **multi**lateral negotiations through WTO → **bi**lateral PTA negotiations
- PTAs are becoming increasingly deep and varied in design
  - PTAs seek to improve market access
  - tariffs already low → # provisions, policy areas covered ↑
  - **non-tariff barriers**: antidumping, rules of origin, ...
  - **behind-the-border policies**: IP protection, gov't procurement, ...
- This Paper: what factors explain PTA formation, differences in PTA design?

Determinants of PTA Design

- Preferential Trade Agreements (PTAs) increasingly dominate international trade governance
  - **multi**lateral negotiations through WTO → **bi**lateral PTA negotiations
- PTAs are becoming increasingly deep and varied in design
  - PTAs seek to improve market access
  - tariffs already low → # provisions, policy areas covered ↑
  - **non-tariff barriers**: antidumping, rules of origin, ...
  - **behind-the-border policies**: IP protection, gov't procurement, ...
- This Paper: what factors explain PTA formation, differences in PTA design?
  - use an existing database of **classified PTA provisions**

Determinants of PTA Design

- Preferential Trade Agreements (PTAs) increasingly dominate international trade governance
  - **multi**lateral negotiations through WTO $\rightarrow$ **bi**lateral PTA negotiations
- PTAs are becoming increasingly deep and varied in design
  - PTAs seek to improve market access
  - tariffs already low $\rightarrow$ # provisions, policy areas covered $\uparrow$
  - **non-tariff barriers**: antidumping, rules of origin, ...
  - **behind-the-border policies**: IP protection, gov't procurement, ...
- This Paper: what factors explain PTA formation, differences in PTA design?
  - use an existing database of **classified PTA provisions**
  - collect a large # of country-pair-level observables as **potential determinants**

Determinants of PTA Design

- Preferential Trade Agreements (PTAs) increasingly dominate international trade governance
  - **multi**lateral negotiations through WTO $\rightarrow$ **bi**lateral PTA negotiations
- PTAs are becoming increasingly deep and varied in design
  - PTAs seek to improve market access
  - tariffs already low $\rightarrow$ # provisions, policy areas covered $\uparrow$
  - **non-tariff barriers**: antidumping, rules of origin, ...
  - **behind-the-border policies**: IP protection, gov't procurement, ...
- This Paper: what factors explain PTA formation, differences in PTA design?
  - use an existing database of **classified PTA provisions**
  - collect a large # of country-pair-level observables as **potential determinants**
  - problem is **high-dimensional** $\rightarrow$ exploit **machine learning**

Determinants of PTA Design

- Preferential Trade Agreements (PTAs) increasingly dominate international trade governance
  - **multi**lateral negotiations through WTO $\rightarrow$ **bi**lateral PTA negotiations
- PTAs are becoming increasingly deep and varied in design
  - PTAs seek to improve market access
  - tariffs already low $\rightarrow$ # provisions, policy areas covered $\uparrow$
  - **non-tariff barriers**: antidumping, rules of origin, ...
  - **behind-the-border policies**: IP protection, gov't procurement, ...
- **This Paper**: what factors explain PTA formation, differences in PTA design?
  - use an existing database of **classified PTA provisions**
  - collect a large # of country-pair-level observables as **potential determinants**
  - problem is **high-dimensional** $\rightarrow$ exploit **machine learning**
  - use random forests to identify important determinants of PTA formation, design
    - country-pair characteristics that are **most predictive** of PTA existence, inclusion of various provisions

Determinants of PTA Design

- Preferential Trade Agreements (PTAs) increasingly dominate international trade governance
  - **multi**lateral negotiations through WTO $\rightarrow$ **bi**lateral PTA negotiations
- PTAs are becoming increasingly deep and varied in design
  - PTAs seek to improve market access
  - tariffs already low $\rightarrow$ # provisions, policy areas covered $\uparrow$
  - **non-tariff barriers**: antidumping, rules of origin, ...
  - **behind-the-border policies**: IP protection, gov't procurement, ...
- **This Paper**: what factors explain PTA formation, differences in PTA design?
  - use an existing database of **classified PTA provisions**
  - collect a large # of country-pair-level observables as **potential determinants**
  - problem is **high-dimensional** $\rightarrow$ exploit **machine learning**
  - use random forests to identify important determinants of PTA formation, design
    - country-pair characteristics that are **most predictive** of PTA existence, inclusion of various provisions
  - consistently **important determinants**: interdependence, geography, governance

Literature

determinants of PTA formation

Literature

determinants of PTA formation

- explore a specific mechanism: *Baldwin and Jaimovich (2012), Mansfield and Milner (2012), Baccini and Urpelainen (2014)*

Literature

DETERMINANTS OF PTA FORMATION

- explore a specific mechanism: *Baldwin and Jaimovich (2012), Mansfield and Milner (2012), Baccini and Urpelainen (2014)*
- compare several determinants: *Baier and Bergstrand (2004), Bergstrand, Egger, and Larch (2016)*

Literature

determinants of PTA formation

- explore a specific mechanism: *Baldwin and Jaimovich (2012), Mansfield and Milner (2012), Baccini and Urpelainen (2014)*
- compare several determinants: *Baier and Bergstrand (2004), Bergstrand, Egger, and Larch (2016)*
- **This Paper**: consider hundreds of determinants, extend approach to PTA design

Literature

DETERMINANTS OF PTA FORMATION

- explore a specific mechanism: *Baldwin and Jaimovich (2012), Mansfield and Milner (2012), Baccini and Urpelainen (2014)*
- compare several determinants: *Baier and Bergstrand (2004), Bergstrand, Egger, and Larch (2016)*
- THIS PAPER: consider hundreds of determinants, extend approach to PTA design

DETERMINANTS OF PTA DESIGN

Literature

DETERMINANTS OF PTA FORMATION

- explore a specific mechanism: *Baldwin and Jaimovich (2012), Mansfield and Milner (2012), Baccini and Urpelainen (2014)*
- compare several determinants: *Baier and Bergstrand (2004), Bergstrand, Egger, and Larch (2016)*
- **This Paper**: consider hundreds of determinants, extend approach to PTA design

DETERMINANTS OF PTA DESIGN

- PTA depth: *Gamso and Grosse (2021), Mattoo et al. (2020)*

LITERATURE

DETERMINANTS OF PTA FORMATION

- explore a specific mechanism: *Baldwin and Jaimovich (2012), Mansfield and Milner (2012), Baccini and Urpelainen (2014)*
- compare several determinants: *Baier and Bergstrand (2004), Bergstrand, Egger, and Larch (2016)*
- **THIS PAPER**: consider hundreds of determinants, extend approach to PTA design

DETERMINANTS OF PTA DESIGN

- PTA depth: *Gamso and Grosse (2021), Mattoo et al. (2020)*
- explore a specific provision area: *Raess et al. (2018), Kucik (2012), Lechner (2016)*

Literature

determinants of PTA formation

- explore a specific mechanism: *Baldwin and Jaimovich (2012), Mansfield and Milner (2012), Baccini and Urpelainen (2014)*
- compare several determinants: *Baier and Bergstrand (2004), Bergstrand, Egger, and Larch (2016)*
- **This Paper**: consider hundreds of determinants, extend approach to PTA design

determinants of PTA design

- PTA depth: *Gamso and Grosse (2021), Mattoo et al. (2020)*
- explore a specific provision area: *Raess et al. (2018), Kucik (2012), Lechner (2016)*
- **This Paper**: explore determinants of hundreds of provisions simultaneously, identify most important mechanisms

Literature

determinants of PTA formation

- explore a specific mechanism: *Baldwin and Jaimovich (2012), Mansfield and Milner (2012), Baccini and Urpelainen (2014)*
- compare several determinants: *Baier and Bergstrand (2004), Bergstrand, Egger, and Larch (2016)*
- **This Paper**: consider hundreds of determinants, extend approach to PTA design

determinants of PTA design

- PTA depth: *Gamso and Grosse (2021), Mattoo et al. (2020)*
- explore a specific provision area: *Raess et al. (2018), Kucik (2012), Lechner (2016)*
- **This Paper**: explore determinants of hundreds of provisions simultaneously, identify most important mechanisms

applying machine learning (ML) to PTA context

Literature

DETERMINANTS OF PTA FORMATION

- explore a specific mechanism: *Baldwin and Jaimovich (2012), Mansfield and Milner (2012), Baccini and Urpelainen (2014)*
- compare several determinants: *Baier and Bergstrand (2004), Bergstrand, Egger, and Larch (2016)*
- **This Paper**: consider hundreds of determinants, extend approach to PTA design

DETERMINANTS OF PTA DESIGN

- PTA depth: *Gamso and Grosse (2021), Mattoo et al. (2020)*
- explore a specific provision area: *Raess et al. (2018), Kucik (2012), Lechner (2016)*
- **This Paper**: explore determinants of hundreds of provisions simultaneously, identify most important mechanisms

APPLYING MACHINE LEARNING (ML) TO PTA CONTEXT

- lasso to identify PTA provisions important for trade: *Breinlich et al. (2022), Kim (2023)*

Literature

Determinants of PTA formation

- explore a specific mechanism: *Baldwin and Jaimovich (2012), Mansfield and Milner (2012), Baccini and Urpelainen (2014)*
- compare several determinants: *Baier and Bergstrand (2004), Bergstrand, Egger, and Larch (2016)*
- **This Paper**: consider hundreds of determinants, extend approach to PTA design

Determinants of PTA design

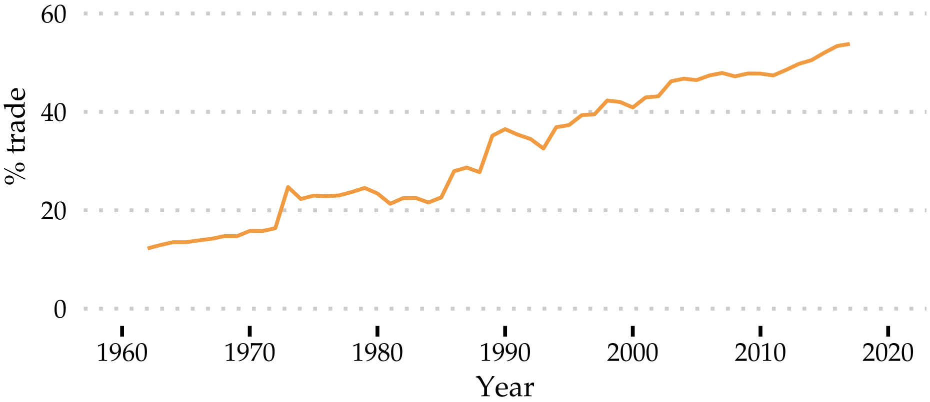
- PTA depth: *Gamso and Grosse (2021), Mattoo et al. (2020)*
- explore a specific provision area: *Raess et al. (2018), Kucik (2012), Lechner (2016)*
- **This Paper**: explore determinants of hundreds of provisions simultaneously, identify most important mechanisms

Applying machine learning (ML) to PTA context

- lasso to identify PTA provisions important for trade: *Breinlich et al. (2022), Kim (2023)*
- **This Paper:** random forest to identify factors important for provisions included in PTA

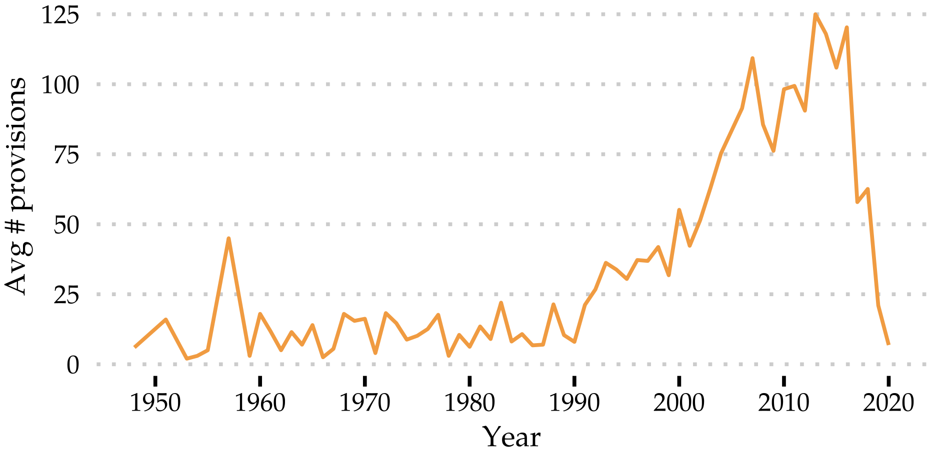
Background

of PTAs Has Skyrocketed



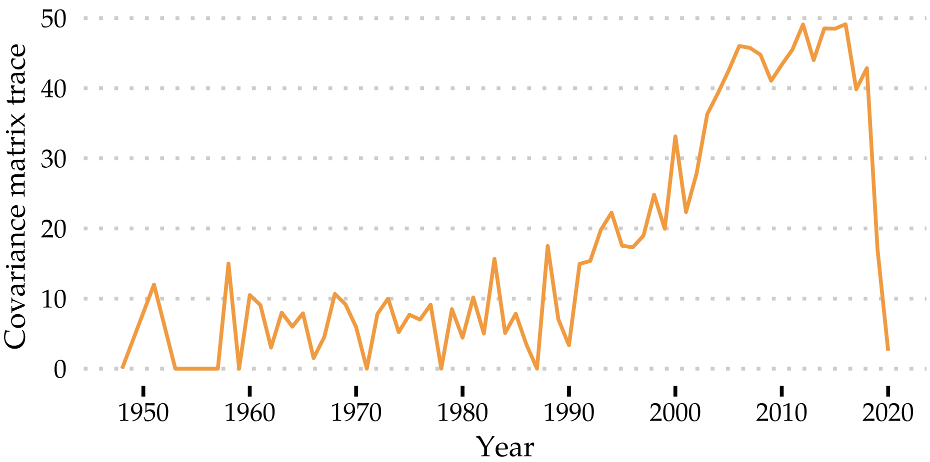
share of global trade happening between PTA members (data: EIA)

PTA Design Has Become More Complex



average number of provisions in a PTA over time (data: DESTA)

PTA Design Has Become More Diverse



variance of provisions included in PTAs over time (data: DESTA)

Understanding PTA Design Complexity

- estimated effects of PTAs on trade $\gg$ effects explainable by trade costs reductions
  - $\rightarrow$ inclusion of non-tariff provisions more important for boosting trade
  - $\rightarrow$ understand why they are included in some PTAs and not others

Understanding PTA Design Complexity

- estimated effects of PTAs on trade $\gg$ effects explainable by trade costs reductions
  - $\rightarrow$ inclusion of non-tariff provisions more important for boosting trade
  - $\rightarrow$ understand why they are included in some PTAs and not others
- different provisions improve different non-trade metrics
  - – even democracy & human rights
  - $\rightarrow$ understand determinants of design to better understand the breadth of outcomes signatories are seeking

Understanding PTA Design Complexity

- estimated effects of PTAs on trade $\gg$ effects explainable by trade costs reductions
  - $\rightarrow$ inclusion of non-tariff provisions more important for boosting trade
  - $\rightarrow$ understand why they are included in some PTAs and not others
- different provisions improve different non-trade metrics
  - – even democracy & human rights
  - $\rightarrow$ understand determinants of design to better understand the breadth of outcomes signatories are seeking
- understanding determinants of provisions matters for understanding their effects
  - – estimation of PTA effects suffers from their endogenous formation
  - – literature has instrumented for PTA formation with determinants exogeneous to studied outcome
  - $\rightarrow$ identify a wide range of design determinants to serve as IVs later

Empirical Approach

- which country-pair characteristics are highly predictive of...

Empirical Exercises

- which country-pair characteristics are highly predictive of...
  1. **PTA formation** between country pair?

Empirical Exercises

- which country-pair characteristics are highly predictive of...
  1. **PTA formation** between country pair?
  2. country pair's **PTA design** (inclusion of a particular provision in a PTA)?

Random Forest

- a supervised learning technique for classification and regression
  - *Breiman (2001)*

Random Forest

- a supervised learning technique for classification and regression
  - *Breiman (2001)*
- use RF to identify important determinants of PTA formation and design
  - **formation**: one RF model
  - **design**: one RF model per classified provision

Random Forest

- a supervised learning technique for classification and regression
  - *Breiman (2001)*
- use RF to identify important determinants of PTA formation and design
  - **formation**: one RF model
  - **design**: one RF model per classified provision

Procedure

Random Forest

- a supervised learning technique for classification and regression
  - *Breiman (2001)*
- use RF to identify important determinants of PTA formation and design
  - **formation**: one RF model
  - **design**: one RF model per classified provision

Procedure

- grow large # of decision trees, each using a bootstrap sample

- a supervised learning technique for classification and regression
  - *Breiman (2001)*
- use RF to identify important determinants of PTA formation and design
  - **formation**: one RF model
  - **design**: one RF model per classified provision

Procedure

- grow large # of decision trees, each using a bootstrap sample
- within each decision tree:

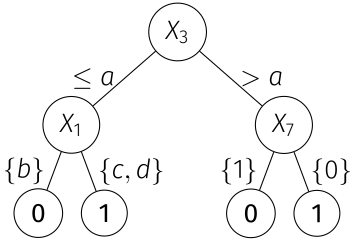


Figure 1: decision tree example

Random Forest

- a supervised learning technique for classification and regression
  - *Breiman (2001)*
- use RF to identify important determinants of PTA formation and design
  - **formation**: one RF model
  - **design**: one RF model per classified provision

Procedure

- grow large # of decision trees, each using a bootstrap sample
- within each decision tree:
  - each node splits the data on one of the x variables, optimizing some measure of fit

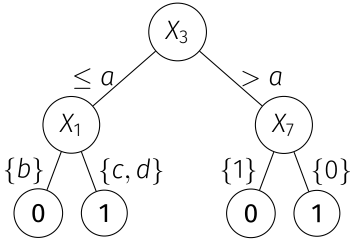


Figure 1: decision tree example

Random Forest

- a supervised learning technique for classification and regression
  - *Breiman (2001)*
- use RF to identify important determinants of PTA formation and design
  - **formation**: one RF model
  - **design**: one RF model per classified provision

Procedure

- grow large # of decision trees, each using a bootstrap sample
- within each decision tree:
  - each node splits the data on one of the x variables, optimizing some measure of fit
  - bottom "leaves" classify the observation into groups (0/1)

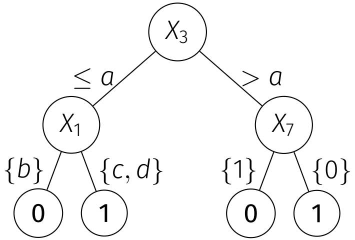


Figure 1: decision tree example

Random Forest

- a supervised learning technique for classification and regression
  - *Breiman (2001)*
- use RF to identify important determinants of PTA formation and design
  - **formation**: one RF model
  - **design**: one RF model per classified provision

Procedure

- grow large # of decision trees, each using a bootstrap sample
- within each decision tree:
  - each node splits the data on one of the *x* variables, optimizing some measure of fit
  - bottom "leaves" classify the observation into groups (0/1)
- to classify an observation: all trees classify it, majority wins

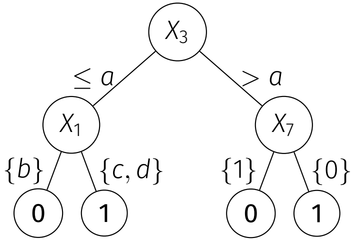


Figure 1: decision tree example

Why Random Forest?

- naturally adapts to non-linearities and interactions in the data

Why Random Forest?

- naturally adapts to non-linearities and interactions in the data
  - infeasible to pre-specify flexible interactions, non-linearities with OLS

Why Random Forest?

- naturally adapts to non-linearities and interactions in the data
  - infeasible to pre-specify flexible interactions, non-linearities with OLS
- methods for dealing with missings

Why Random Forest?

- naturally adapts to non-linearities and interactions in the data
  - infeasible to pre-specify flexible interactions, non-linearities with OLS
- methods for dealing with missings
  - requiring no missings would make considering 100s of determinants infeasible

Why Random Forest?

- naturally adapts to non-linearities and interactions in the data
  - infeasible to pre-specify flexible interactions, non-linearities with OLS
- methods for dealing with missings
  - requiring no missings would make considering 100s of determinants infeasible
  - imputing data would create spurious variable importance results

Why Random Forest?

- naturally adapts to non-linearities and interactions in the data
  - infeasible to pre-specify flexible interactions, non-linearities with OLS
- methods for dealing with missings
  - requiring no missings would make considering 100s of determinants infeasible
  - imputing data would create spurious variable importance results
  - $\rightarrow$ use *Tang, Ishwaran (2017)* method: only uses non-missing data for splits

Why Random Forest?

- naturally adapts to non-linearities and interactions in the data
  - infeasible to pre-specify flexible interactions, non-linearities with OLS
- methods for dealing with missings
  - requiring no missings would make considering 100s of determinants infeasible
  - imputing data would create spurious variable importance results
  - $\rightarrow$ use *Tang, Ishwaran (2017)* method: only uses non-missing data for splits
  - ▶ *details*

Why Random Forest?

- naturally adapts to non-linearities and interactions in the data
  - infeasible to pre-specify flexible interactions, non-linearities with OLS
- methods for dealing with missings
  - requiring no missings would make considering 100s of determinants infeasible
  - imputing data would create spurious variable importance results
  - $\rightarrow$ use *Tang, Ishwaran (2017)* method: only uses non-missing data for splits
  - ▶ *details*
- well-developed variable importance measures

Why Random Forest?

- naturally adapts to non-linearities and interactions in the data
  - infeasible to pre-specify flexible interactions, non-linearities with OLS
- methods for dealing with missings
  - requiring no missings would make considering 100s of determinants infeasible
  - imputing data would create spurious variable importance results
  - → use *Tang, Ishwaran (2017)* method: only uses non-missing data for splits
  - ▶ *details*
- well-developed variable importance measures
  - identifying important PTA provision determinants is the goal

Why Random Forest?

- naturally adapts to non-linearities and interactions in the data
  - infeasible to pre-specify flexible interactions, non-linearities with OLS
- methods for dealing with missings
  - requiring no missings would make considering 100s of determinants infeasible
  - imputing data would create spurious variable importance results
  - $\rightarrow$ use *Tang, Ishwaran (2017)* method: only uses non-missing data for splits
  - ▶ *details*
- well-developed variable importance measures
  - identifying important PTA provision determinants is the goal

▶ *tuning*

RF Implementation: Variable Importance

- use **Mean Decrease Accuracy** + *Altmann et al. (2010)* **Permutation Importance**

- use **Mean Decrease Accuracy** + *Altmann et al. (2010)* **Permutation Importance**
- Mean Decrease Accuracy:
  - conventional measure from *Breiman (2001)*
  - permutes variable's values, compares prediction error on this fake data
  - **problems**: mechanical biases, no absolute significance cutoff

RF Implementation: Variable Importance

- use **Mean Decrease Accuracy** + *Altmann et al. (2010)* **Permutation Importance**
- Mean Decrease Accuracy:
  - conventional measure from *Breiman (2001)*
  - permutes variable's values, compares prediction error on this fake data
  - **problems**: mechanical biases, no absolute significance cutoff
- *Altmann et al. (2010)*:
  - permute the outcome variable many times → re-construct the RF and VIMs for each permutation → obtain null distributions
  - use p-value as corrected variable importance measure
  - **benefits**: corrects biases, provides statistical significance

RF Implementation: Variable Importance

- use **Mean Decrease Accuracy** + *Altmann et al. (2010)* **Permutation Importance**
- Mean Decrease Accuracy:
  - conventional measure from *Breiman (2001)*
  - permutes variable's values, compares prediction error on this fake data
  - **problems**: mechanical biases, no absolute significance cutoff
- *Altmann et al. (2010)*:
  - permute the outcome variable many times $\rightarrow$ re-construct the RF and VIMs for each permutation $\rightarrow$ obtain null distributions
  - use p-value as corrected variable importance measure
  - **benefits**: corrects biases, provides statistical significance
- variable is an "important" predictor if Altmann permutation importance p-value $< 1\%$

DATA

Data: PTAs and Provisions

- **formation exercise**: Economic Integration Agreements (EIA), *NSF-Kellogg Inst., 2021*

Data: PTAs and Provisions

- **formation exercise**: Economic Integration Agreements (EIA), *NSF-Kellogg Inst., 2021*
  - most complete database of PTAs between country pairs

Data: PTAs and Provisions

- **formation exercise**: Economic Integration Agreements (EIA), *NSF-Kellogg Inst., 2021*
  - most complete database of PTAs between country pairs
- **main design exercise**: Design of Trade Agreements (DESTA), *Dür et al., 2014*

Data: PTAs and Provisions

- **formation exercise**: Economic Integration Agreements (EIA), *NSF-Kellogg Inst., 2021*
  - most complete database of PTAs between country pairs
- **main design exercise**: Design of Trade Agreements (DESTA), *Dür et al., 2014*
  - manually classify provisions in existing PTAs

Data: PTAs and Provisions

- **formation exercise**: Economic Integration Agreements (EIA), *NSF-Kellogg Inst., 2021*
  - most complete database of PTAs between country pairs
- **main design exercise**: Design of Trade Agreements (DESTA), *Dür et al., 2014*
  - manually classify provisions in existing PTAs
  - 710 PTAs

Data: PTAs and Provisions

- **formation exercise**: Economic Integration Agreements (EIA), *NSF-Kellogg Inst., 2021*
  - most complete database of PTAs between country pairs
- **main design exercise**: Design of Trade Agreements (DESTA), *Dür et al., 2014*
  - manually classify provisions in existing PTAs
  - 710 PTAs
  - ∼300 provisions, we use 119 with enough variation

Data: PTAs and Provisions

- **formation exercise**: Economic Integration Agreements (EIA), *NSF-Kellogg Inst., 2021*
  - most complete database of PTAs between country pairs
- **main design exercise**: Design of Trade Agreements (DESTA), *Dür et al., 2014*
  - manually classify provisions in existing PTAs
  - 710 PTAs
  - ~300 provisions, we use 119 with enough variation
- **robustness**: Deep Trade Agreements (DTA), *Mattoo et al., 2020*

Data: PTAs and Provisions

- **formation exercise**: Economic Integration Agreements (EIA), *NSF-Kellogg Inst., 2021*
  - most complete database of PTAs between country pairs
- **main design exercise**: Design of Trade Agreements (DESTA), *Dür et al., 2014*
  - manually classify provisions in existing PTAs
  - 710 PTAs
  - ∼300 provisions, we use 119 with enough variation
- **robustness**: Deep Trade Agreements (DTA), *Mattoo et al., 2020*
  - 274 PTAs

Data: PTAs and Provisions

- **formation exercise**: Economic Integration Agreements (EIA), *NSF-Kellogg Inst., 2021*
  - most complete database of PTAs between country pairs
- **main design exercise**: Design of Trade Agreements (DESTA), *Dür et al., 2014*
  - manually classify provisions in existing PTAs
  - 710 PTAs
  - ∼300 provisions, we use 119 with enough variation
- **robustness**: Deep Trade Agreements (DTA), *Mattoo et al., 2020*
  - 274 PTAs
  - 937 provisions

Data: Determinants

1. identify 287 economic, geographic, and political factors that potentially influence PTA design

Data: Determinants

1. identify 287 economic, geographic, and political factors that potentially influence PTA design
   - factors found to be important determinants of overall PTA formation
     - e.g. macroeconomic variables, PTA contagion, domestic politics

Data: Determinants

1. identify 287 economic, geographic, and political factors that potentially influence PTA design
   - factors found to be important determinants of overall PTA formation
     - e.g. macroeconomic variables, PTA contagion, domestic politics
   - factors related to particular PTA provision policy areas
     - e.g. innovation, energy use, labor markets, sectoral variables

Data: Determinants

1. identify 287 economic, geographic, and political factors that potentially influence PTA design
   - factors found to be important determinants of overall PTA formation
     - e.g. macroeconomic variables, PTA contagion, domestic politics
   - factors related to particular PTA provision policy areas
     - e.g. innovation, energy use, labor markets, sectoral variables
2. collect and merge data on potential determinants
   - **economy, society**: Penn World Tables, World Development Indicators
   - **proximity, culture**: CEPII Gravity, CEPII Language, GeoDist, UNCTADstat
   - **trade, FDI**: UN Comtrade, WITS, IMF CDIS, BACI
   - **politics**: Database of Political Institutions, Worldwide Governance Indicators

Data: Determinants

1. identify 287 economic, geographic, and political factors that potentially influence PTA design
   - factors found to be important determinants of overall PTA formation
     - e.g. macroeconomic variables, PTA contagion, domestic politics
   - factors related to particular PTA provision policy areas
     - e.g. innovation, energy use, labor markets, sectoral variables
2. collect and merge data on potential determinants
   - **economy, society**: Penn World Tables, World Development Indicators
   - **proximity, culture**: CEPII Gravity, CEPII Language, GeoDist, UNCTADstat
   - **trade, FDI**: UN Comtrade, WITS, IMF CDIS, BACI
   - **politics**: Database of Political Institutions, Worldwide Governance Indicators
3. construct all potential determinants at **country-pair** level

Data: Determinants

1. identify 287 economic, geographic, and political factors that potentially influence PTA design
   - factors found to be important determinants of overall PTA formation
     - e.g. macroeconomic variables, PTA contagion, domestic politics
   - factors related to particular PTA provision policy areas
     - e.g. innovation, energy use, labor markets, sectoral variables
2. collect and merge data on potential determinants
   - **economy, society**: Penn World Tables, World Development Indicators
   - **proximity, culture**: CEPII Gravity, CEPII Language, GeoDist, UNCTADstat
   - **trade, FDI**: UN Comtrade, WITS, IMF CDIS, BACI
   - **politics**: Database of Political Institutions, Worldwide Governance Indicators
3. construct all potential determinants at **country-pair** level
   - **country-pair**-level variables: as-is
     - e.g. country-pair's intra-industry trade index

Data: Determinants

1. identify 287 economic, geographic, and political factors that potentially influence PTA design
   - factors found to be important determinants of overall PTA formation
     - > e.g. macroeconomic variables, PTA contagion, domestic politics
   - factors related to particular PTA provision policy areas
     - > e.g. innovation, energy use, labor markets, sectoral variables
2. collect and merge data on potential determinants
   - **economy, society**: Penn World Tables, World Development Indicators
   - **proximity, culture**: CEPII Gravity, CEPII Language, GeoDist, UNCTADstat
   - **trade, FDI**: UN Comtrade, WITS, IMF CDIS, BACI
   - **politics**: Database of Political Institutions, Worldwide Governance Indicators
3. construct all potential determinants at **country-pair** level
   - **country-pair**-level variables: as-is
     - > e.g. country-pair's intra-industry trade index
   - **country**-level **continuous** variables: mean, difference within pair
     - > e.g. absolute log GDP difference within pair

Data: Determinants

1. identify 287 economic, geographic, and political factors that potentially influence PTA design
   - factors found to be important determinants of overall PTA formation
     - e.g. macroeconomic variables, PTA contagion, domestic politics
   - factors related to particular PTA provision policy areas
     - e.g. innovation, energy use, labor markets, sectoral variables
2. collect and merge data on potential determinants
   - **economy, society**: Penn World Tables, World Development Indicators
   - **proximity, culture**: CEPII Gravity, CEPII Language, GeoDist, UNCTADstat
   - **trade, FDI**: UN Comtrade, WITS, IMF CDIS, BACI
   - **politics**: Database of Political Institutions, Worldwide Governance Indicators
3. construct all potential determinants at **country-pair** level
   - **country-pair**-level variables: as-is
     - e.g. country-pair's intra-industry trade index
   - **country**-level **continuous** variables: mean, difference within pair
     - e.g. absolute log GDP difference within pair
   - **country**-level **factor** variables: same or not, combination
     - e.g. combination of political regimes within pair

Results: PTA Formation

Determinants of PTA Formation

- country-pair characteristics most predictive of PTAs:
  - **geographic proximity**: distance, continents
  - **contagion**: competition from third countries for export markets
  - **domestic politics**: executive and legislative composition, features, tenure
  - **regulatory quality**: accountability, ease of doing business
  - **trade**: bilateral trade volume, intra-industry trade

Determinants of PTA Formation

- country-pair characteristics most predictive of PTAs:
  - **geographic proximity**: distance, continents
  - **contagion**: competition from third countries for export markets
  - **domestic politics**: executive and legislative composition, features, tenure
  - **regulatory quality**: accountability, ease of doing business
  - **trade**: bilateral trade volume, intra-industry trade
- **overall**: support most findings of studies focusing on a particular mechanism
  - but difference in regulatory quality in development *not* important

Determinants of PTA Formation

- country-pair characteristics most predictive of PTAs:
  - **geographic proximity**: distance, continents
  - **contagion**: competition from third countries for export markets
  - **domestic politics**: executive and legislative composition, features, tenure
  - **regulatory quality**: accountability, ease of doing business
  - **trade**: bilateral trade volume, intra-industry trade
- **overall**: support most findings of studies focusing on a particular mechanism
  - but difference in regulatory quality in development *not* important

▶ *modifications & performance*

Most Likely New PTAs

- country pairs that are most likely to have a PTA according to RF, but don't

	Country Pair	PTA Probability
1	Dominican Republic, Panama	0.47
2	Colombia, Costa Rica	0.46
3	Bosnia & Herzegovina, Slovenia	0.45
4	Colombia, Dominican Republic	0.45
5	Norway, Russia	0.45
6	Albania, Greece	0.45
7	Ecuador, Panama	0.44
8	Australia, Germany	0.43
9	Albania, Spain	0.42
10	Austria, Bosnia & Herzegovina	0.42

Most Likely New PTAs

- country pairs that are most likely to have a PTA according to RF, but don't

	Country Pair	PTA Probability
1	Dominican Republic, Panama	0.47
2	Colombia, Costa Rica	0.46
3	Bosnia & Herzegovina, Slovenia	0.45
4	Colombia, Dominican Republic	0.45
5	Norway, Russia	0.45
6	Albania, Greece	0.45
7	Ecuador, Panama	0.44
8	Australia, Germany	0.43
9	Albania, Spain	0.42
10	Austria, Bosnia & Herzegovina	0.42

- 8/10 are making progress toward a PTA

Most Likely New PTAs

- country pairs that are most likely to have a PTA according to RF, but don't

	Country Pair	PTA Probability
1	Dominican Republic, Panama	0.47
2	Colombia, Costa Rica	0.46
3	Bosnia & Herzegovina, Slovenia	0.45
4	Colombia, Dominican Republic	0.45
5	Norway, Russia	0.45
6	Albania, Greece	0.45
7	Ecuador, Panama	0.44
8	Australia, Germany	0.43
9	Albania, Spain	0.42
10	Austria, Bosnia & Herzegovina	0.42

- 8/10 are making progress toward a PTA
  - 2: new PTA not recorded in EIA

Most Likely New PTAs

- country pairs that are most likely to have a PTA according to RF, but don't

	Country Pair	PTA Probability
1	Dominican Republic, Panama	0.47
2	Colombia, Costa Rica	0.46
3	Bosnia & Herzegovina, Slovenia	0.45
4	Colombia, Dominican Republic	0.45
5	Norway, Russia	0.45
6	Albania, Greece	0.45
7	Ecuador, Panama	0.44
8	Australia, Germany	0.43
9	Albania, Spain	0.42
10	Austria, Bosnia & Herzegovina	0.42

- 8/10 are making progress toward a PTA
  - 2: new PTA not recorded in EIA
  - 3, 6, 9, 10: candidates for EU accession

Most Likely New PTAs

- country pairs that are most likely to have a PTA according to RF, but don't

	Country Pair	PTA Probability
1	Dominican Republic, Panama	0.47
2	Colombia, Costa Rica	0.46
3	Bosnia & Herzegovina, Slovenia	0.45
4	Colombia, Dominican Republic	0.45
5	Norway, Russia	0.45
6	Albania, Greece	0.45
7	Ecuador, Panama	0.44
8	Australia, Germany	0.43
9	Albania, Spain	0.42
10	Austria, Bosnia & Herzegovina	0.42

- 8/10 are making progress toward a PTA
  - 2: new PTA not recorded in EIA
  - 3, 6, 9, 10: candidates for EU accession
  - 4: both in various stages of integration with CAPRICOM

Most Likely New PTAs

- country pairs that are most likely to have a PTA according to RF, but don't

	Country Pair	PTA Probability
1	Dominican Republic, Panama	0.47
2	Colombia, Costa Rica	0.46
3	Bosnia & Herzegovina, Slovenia	0.45
4	Colombia, Dominican Republic	0.45
5	Norway, Russia	0.45
6	Albania, Greece	0.45
7	Ecuador, Panama	0.44
8	Australia, Germany	0.43
9	Albania, Spain	0.42
10	Austria, Bosnia & Herzegovina	0.42

- 8/10 are making progress toward a PTA
  - 2: new PTA not recorded in EIA
  - 3, 6, 9, 10: candidates for EU accession
  - 4: both in various stages of integration with CAPRICOM
  - 7: negotiating a partial scope agreement

Most Likely New PTAs

- country pairs that are most likely to have a PTA according to RF, but don't

	Country Pair	PTA Probability
1	Dominican Republic, Panama	0.47
2	Colombia, Costa Rica	0.46
3	Bosnia & Herzegovina, Slovenia	0.45
4	Colombia, Dominican Republic	0.45
5	Norway, Russia	0.45
6	Albania, Greece	0.45
7	Ecuador, Panama	0.44
8	Australia, Germany	0.43
9	Albania, Spain	0.42
10	Austria, Bosnia & Herzegovina	0.42

- 8/10 are making progress toward a PTA
  - 2: new PTA not recorded in EIA
  - 3, 6, 9, 10: candidates for EU accession
  - 4: both in various stages of integration with CAPRICOM
  - 7: negotiating a partial scope agreement
  - 8: Australia negotiating FTA with EU

Results: PTA Design

Predictive Performance

- compare performance of RF models for each provision with alternative models
  - 2/3 training sample, 1/3 test sample
  - RF outperforms a single tree and conventional logistic regression

	Full Sample			Non-NA Sub-Sample		
	Random Forest	Tree	Logit	Random Forest	Tree	Logit
25th %-ile	0.123	0.212	0.432	0.055	0.107	0.079
median	**0.164**	**0.277**	**0.456**	**0.079**	**0.145**	**0.114**
75th %-ile	0.227	0.376	0.477	0.115	0.211	0.157

Predictive Performance

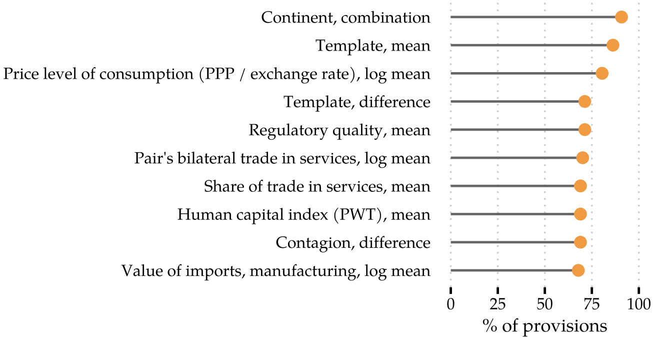
- compare performance of RF models for each provision with alternative models
  - 2/3 training sample, 1/3 test sample
  - RF outperforms a single tree and conventional logistic regression

	Full Sample			Non-NA Sub-Sample		
	Random Forest	Tree	Logit	Random Forest	Tree	Logit
25th %-ile	0.123	0.212	0.432	0.055	0.107	0.079
median	**0.164**	**0.277**	**0.456**	**0.079**	**0.145**	**0.114**
75th %-ile	0.227	0.376	0.477	0.115	0.211	0.157

▶ *OOB performance*

Results: Top 10 Determinants of Overall Design

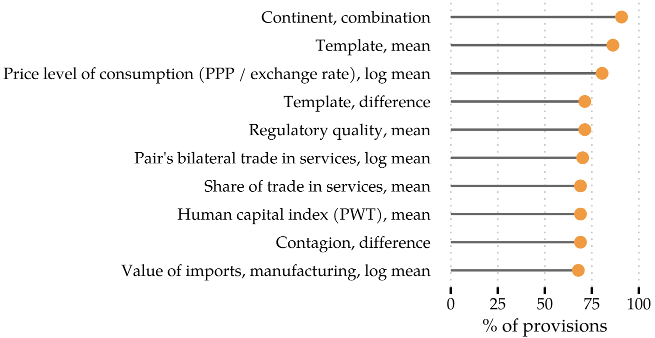
- rank country-pair characteristics by % of provisions each is a significant predictor of



- **template, contagion**: competitive pressure to copy provisions not to lose market share in an export market ▶ *details*
- **geography**: content of PTAs is different depending on proximity and location of partners

Results: Top 10 Determinants of Overall Design

- rank country-pair characteristics by % of provisions each is a significant predictor of



- **template, contagion**: competitive pressure to copy provisions not to lose market share in an export market ▶ *details*
- **geography**: content of PTAs is different depending on proximity and location of partners

▶ *alternative variable sets* ▶ *distribution of determinants* ▶ *DTA*

Conclusion

Conclusion

- use random forests to identify important determinants of PTA formation & design

Conclusion

- use random forests to identify important determinants of PTA formation & design
    - interdependence, geographic, regulatory measures emerge as leading

Conclusion

- use random forests to identify important determinants of PTA formation & design
  - interdependence, geographic, regulatory measures emerge as leading
- **limitations & future research**

Conclusion

- use random forests to identify important determinants of PTA formation & design
  - interdependence, geographic, regulatory measures emerge as leading
- **limitations & future research**
  - treat multilateral PTAs as sets of bilateral agreements between country pairs

Conclusion

- use random forests to identify important determinants of PTA formation & design
  - interdependence, geographic, regulatory measures emerge as leading
- limitations & future research
  - treat multilateral PTAs as sets of bilateral agreements between country pairs
    - for consistency with literature and to permit straightforward analysis on rectangular data

Conclusion

- use random forests to identify important determinants of PTA formation & design
  - interdependence, geographic, regulatory measures emerge as leading
- limitations & future research
  - treat multilateral PTAs as sets of bilateral agreements between country pairs
    - for consistency with literature and to permit straightforward analysis on rectangular data
    - but it ignores the *multi*lateral components: size and composition of a PTA being negotiated

Conclusion

- use random forests to identify important determinants of PTA formation & design
  - interdependence, geographic, regulatory measures emerge as leading
- **limitations & future research**
  - treat multilateral PTAs as sets of bilateral agreements between country pairs
    - for consistency with literature and to permit straightforward analysis on rectangular data
    - but it ignores the *multi*lateral components: size and composition of a PTA being negotiated

    → multilateral agreement level: OLS/RF using feature moments across members, or ML methods supporting variable feature sets

Conclusion

- use random forests to identify important determinants of PTA formation & design
  - interdependence, geographic, regulatory measures emerge as leading
- **limitations & future research**
  - treat multilateral PTAs as sets of bilateral agreements between country pairs
    - for consistency with literature and to permit straightforward analysis on rectangular data
    - but it ignores the *multi*lateral components: size and composition of a PTA being negotiated
    - → multilateral agreement level: OLS/RF using feature moments across members, or ML methods supporting variable feature sets
  - RFs highlight characteristics important for PTA design, but don't provide a coherent story of *why* they are important

Conclusion

- use random forests to identify important determinants of PTA formation & design
  - interdependence, geographic, regulatory measures emerge as leading
- **limitations & future research**
  - treat multilateral PTAs as sets of bilateral agreements between country pairs
    - for consistency with literature and to permit straightforward analysis on rectangular data
    - but it ignores the *multi*lateral components: size and composition of a PTA being negotiated
    - → multilateral agreement level: OLS/RF using feature moments across members, or ML methods supporting variable feature sets
  - RFs highlight characteristics important for PTA design, but don't provide a coherent story of *why* they are important
    - RFs don't impose linearity on the data → pick up non-linearities/interactions

Conclusion

- use random forests to identify important determinants of PTA formation & design
  - interdependence, geographic, regulatory measures emerge as leading
- **limitations & future research**
  - treat multilateral PTAs as sets of bilateral agreements between country pairs
    - for consistency with literature and to permit straightforward analysis on rectangular data
    - but it ignores the *multi*lateral components: size and composition of a PTA being negotiated
    - → multilateral agreement level: OLS/RF using feature moments across members, or ML methods supporting variable feature sets
  - RFs highlight characteristics important for PTA design, but don't provide a coherent story of *why* they are important
    - RFs don't impose linearity on the data → pick up non-linearities/interactions
    - but can't provide a single coefficient summarizing a variable's effect

Conclusion

- use random forests to identify important determinants of PTA formation & design
  - interdependence, geographic, regulatory measures emerge as leading
- **limitations & future research**
  - treat multilateral PTAs as sets of bilateral agreements between country pairs
    - for consistency with literature and to permit straightforward analysis on rectangular data
    - but it ignores the *multi*lateral components: size and composition of a PTA being negotiated
    - → multilateral agreement level: OLS/RF using feature moments across members, or ML methods supporting variable feature sets
  - RFs highlight characteristics important for PTA design, but don't provide a coherent story of *why* they are important
    - RFs don't impose linearity on the data → pick up non-linearities/interactions
    - but can't provide a single coefficient summarizing a variable's effect
    - → motivate future research to focus on individual mechanisms related to identified determinants

RF Tuning

- parameters
  - N (number of trees): 500
  - M_{try} (number of variables to consider at each split), *nodesize* (number of obs in terminal node): k-fold cross-validation

back

RF Tuning

- parameters
  - N (number of trees): 500
  - M_{try} (number of variables to consider at each split), *nodesize* (number of obs in terminal node): k-fold cross-validation
- splitting statistic: **AUC-ROC**
  - prob. that a random “true 0” and “true 1” are both classified correctly
  - due to imbalanced data, outperforms standard **misclassification rate**
    - even when overall performance measured with misclassification rate

back

RF Implementation: Missings

- **problem**: most observations have at least some missing predictors
  - collect hundreds of potential determinants from dozens of datasets
  - each dataset has its own data coverage

back

RF Implementation: Missings

- **problem**: most observations have at least some missing predictors
  - collect hundreds of potential determinants from dozens of datasets
  - each dataset has its own data coverage
- **standard solutions**:
  - throw away data (only use rows with no missing columns, or only use columns with no missing rows)
    - $\rightarrow$ throw away almost all data
  - or impute missings
    - $\rightarrow$ VIMs become unreliable: imputed variable contains info from other variables

back

RF Implementation: Missings

- **problem**: most observations have at least some missing predictors
  - collect hundreds of potential determinants from dozens of datasets
  - each dataset has its own data coverage
- **standard solutions**:
  - throw away data (only use rows with no missing columns, or only use columns with no missing rows)
    - $\rightarrow$ throw away almost all data
  - or impute missings
    - $\rightarrow$ VIMs become unreliable: imputed variable contains info from other variables
- *Tang, Ishwaran (2017)*: "on-the-fly imputation" for RFs
  - at each node, only non-missing data is used to come up with a split, then missings are split randomly
  - $\rightarrow$ uses all data without imputation
    - variables with many missings naturally get lower VIM values

back

Contagion

- large literature documents contagion/interdependence of PTAs

back

Contagion

- large literature documents contagion/interdependence of PTAs
- likely mechanism:
  - country j is an important export market for i, k
  - country k signs a PTA with j
  - $\rightarrow$ i has an incentive to also sign a PTA with j, or risk losing market share

back

Contagion

- large literature documents contagion/interdependence of PTAs
- likely mechanism:
  - country j is an important export market for i, k
  - country k signs a PTA with j
  - $\rightarrow$ i has an incentive to also sign a PTA with j, or risk losing market share
- we extend this to contagion of **provisions**, following the common specification of contagion:

back

Contagion

- large literature documents contagion / interdependence of PTAs
- likely mechanism:
  - country j is an important export market for i, k
  - country k signs a PTA with j
  - $\rightarrow$ i has an incentive to also sign a PTA with j, or risk losing market share
- we extend this to contagion of **provisions**, following the common specification of contagion:

$$\text{Contagion}_{p,ij,t} = \left(\frac{\text{bilateral exports}_{ij}}{\text{total exports}_i}\right) \sum_{k \neq i,j} \left(\frac{\text{bilateral exports}_{kj}}{\text{total imports}_j}\right) 1_{p,jkt}$$

back

Formation Details & Performance

- 480,738 country-pair-5yr observations $\rightarrow$ computational simplifications
  - 10 (rather than all) random splitting points considered at each node
  - use random 10% of sample for MDA values at each iteration

back

Formation Details & Performance

- 480,738 country-pair-5yr observations → computational simplifications
  - 10 (rather than all) random splitting points considered at each node
  - use random 10% of sample for MDA values at each iteration
- highly imbalanced data (3.5% minority class) → *O'Brien and Ishwaran (2019)* quantile classifier
  - effectively boosts the predictions of the minority class

back

Formation Details & Performance

- 480,738 country-pair-5yr observations → computational simplifications
  - 10 (rather than all) random splitting points considered at each node
  - use random 10% of sample for MDA values at each iteration
- highly imbalanced data (3.5% minority class) → *O'Brien and Ishwaran (2019)* quantile classifier
  - effectively boosts the predictions of the minority class
- evaluate performance with **out-of-bag** misclassification
  - for each observation, only the trees that did not have it in their bootstrap sample are used

OOB Misclassification			
Overall	0 (Absent)	1 (Present)	Share of 1s
0.255	0.264	0.021	0.035

back

OOB Performance

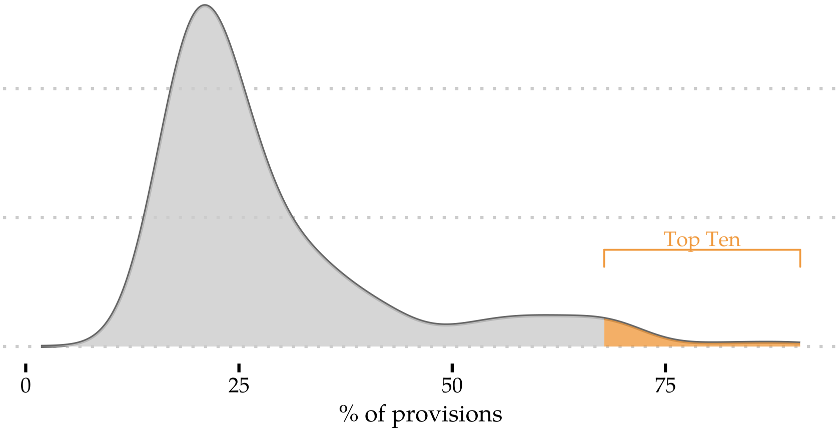
- evaluate performance with **out-of-bag** misclassification
  - for each observation, only the trees that did not have it in their bootstrap sample are used

	OOB Misclassification			
	Overall	0 (Absent)	1 (Present)	Share of 1s
25th %-ile	0.120	0.000	0.390	0.148
median	**0.160**	**0.006**	**0.702**	**0.212**
75th %-ile	0.226	0.034	0.948	0.359

back

Results: Distribution of Determinant Importances

- distribution of variables by the % of provisions each is a significant predictor of



back

Results: Alternative Variable Sets

- alternative way to isolate important variables: run RFs with subsets of country-pair characteristics

All	Excluding Top 10	Economy	Geography	History, Culture	Interdependence	Politics	Trade
0.160	0.186	0.191	0.148	0.210	0.137	0.179	0.187

back

Results: Alternative Variable Sets

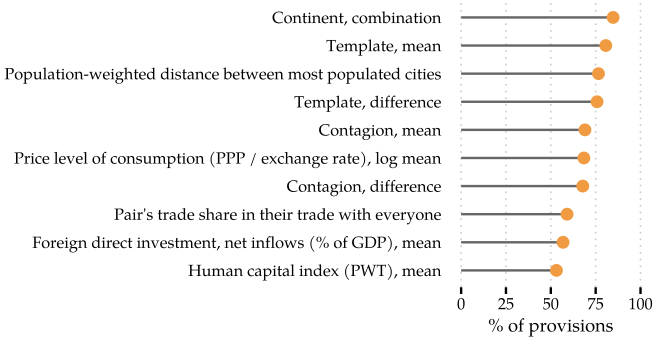
- alternative way to isolate important variables: run RFs with subsets of country-pair characteristics

All	Excluding Top 10	Economy	Geography	History, Culture	Interdependence	Politics	Trade
0.160	0.186	0.191	0.148	0.210	0.137	0.179	0.187

- **interdependence**, **geography** again have highest predictive power
  - even outperform the original RF with complete set of variables

back

Results: Top 10 Determinants of Overall Design, DTA



back