Gaussian Processes

David Carlson

May 26, 2021

Overview

- Gaussian process regression (GPR) for time-series cross sectional (TSCS) analyses

Overview

- Gaussian process regression (GPR) for time-series cross sectional (TSCS) analyses
- GPR for prediction / projection

Overview

- Gaussian process regression (GPR) for time-series cross sectional (TSCS) analyses
- GPR for prediction / projection
- Other uses of GPs

Motivation of TSCS

- Try to relate explanatory to outcome

Motivation of TSCS

- Try to relate explanatory to outcome
- Variables often violate modeling assumptions

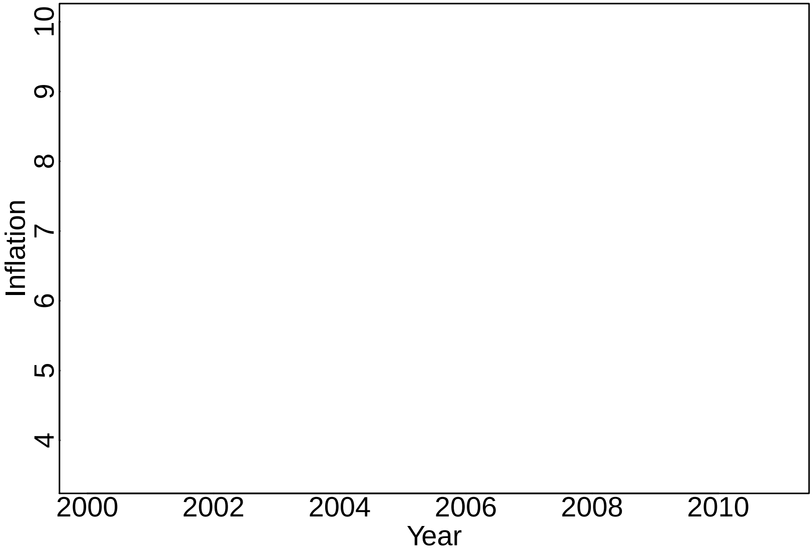
Motivation of TSCS

- Try to relate explanatory to outcome
- Variables often violate modeling assumptions
- Observations not conditionally independent

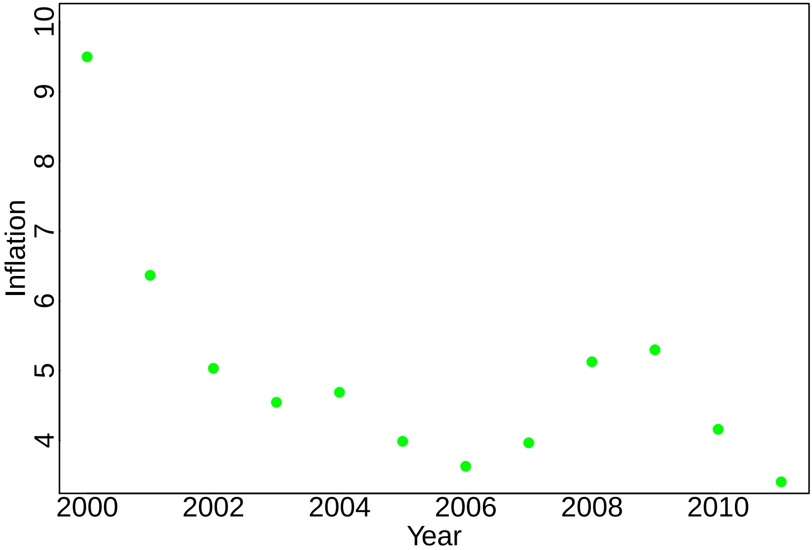
Motivation of TSCS

- Try to relate explanatory to outcome
- Variables often violate modeling assumptions
- Observations not conditionally independent
- Example: How does inflation explain anti-Americanism?

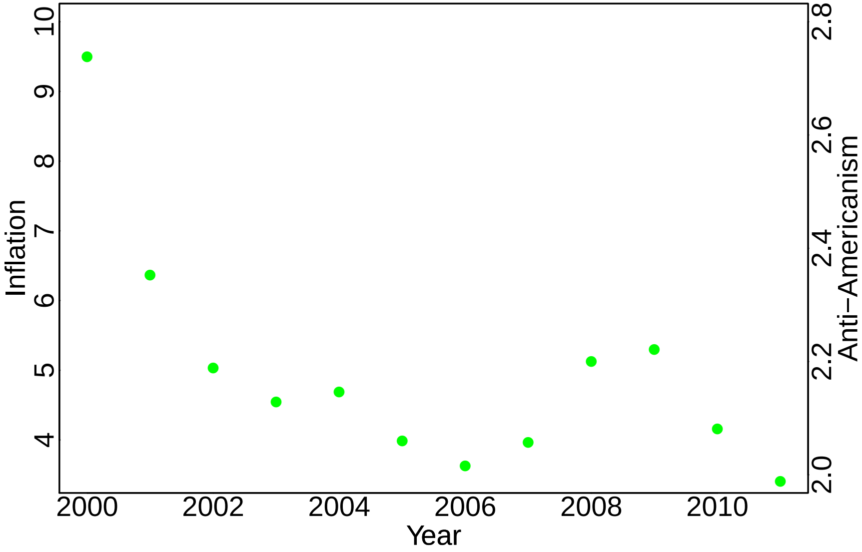
Inflation and Anti-Americanism in Mexico



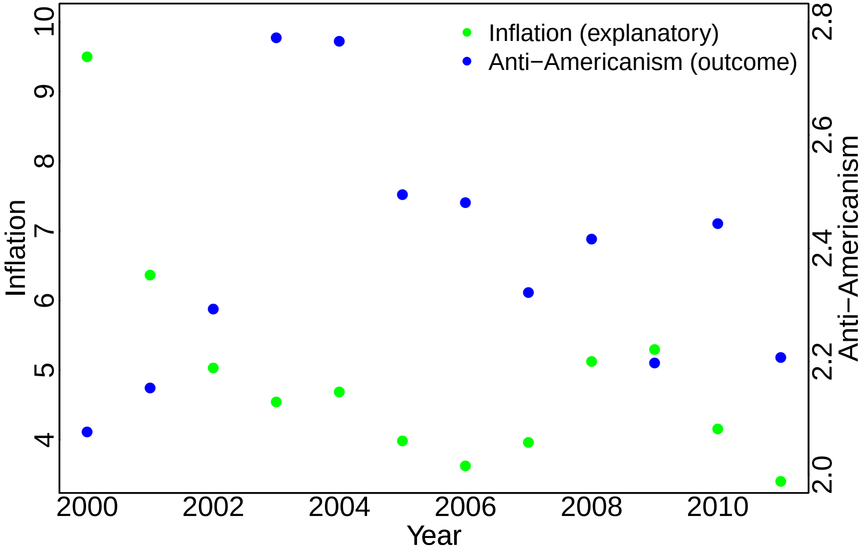
Inflation and Anti-Americanism in Mexico



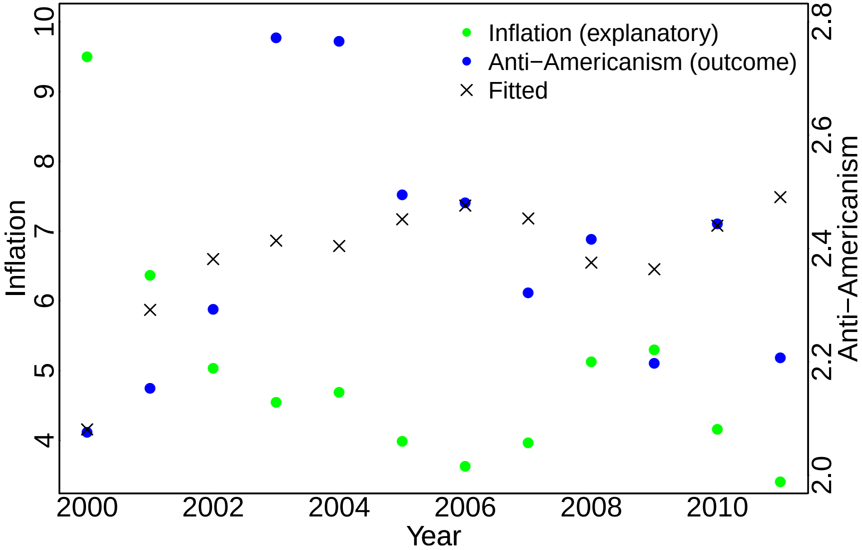
Inflation and Anti-Americanism in Mexico



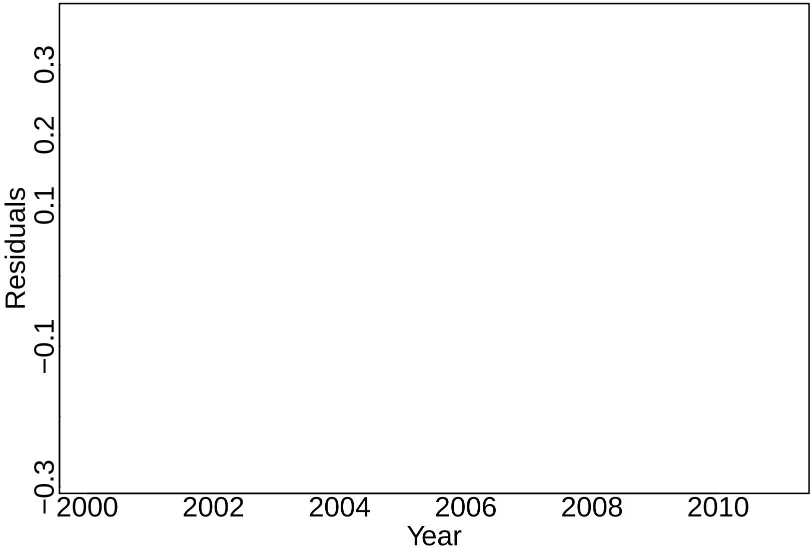
Inflation and Anti-Americanism in Mexico



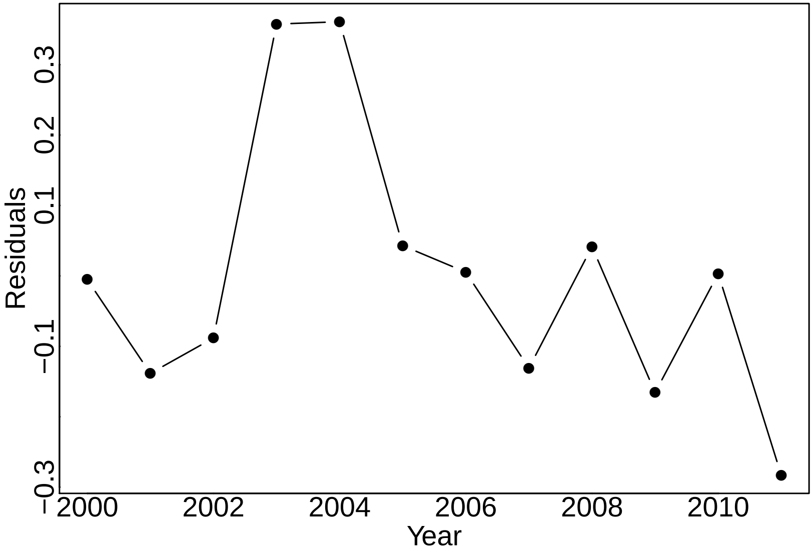
Inflation and Anti-Americanism in Mexico



Inflation and Anti-Americanism in Mexico



Inflation and Anti-Americanism in Mexico



TSCS Data

- Time-series cross-sectional (panel) analyses

TSCS Data

- Time-series cross-sectional (panel) analyses
- Issues:

TSCS Data

- Time-series cross-sectional (panel) analyses
- Issues:
  - Serial correlation

TSCS Data

- Time-series cross-sectional (panel) analyses
- Issues:
  - Serial correlation
  - Unit heterogeneity

TSCS Data

- Time-series cross-sectional (panel) analyses
- Issues:
  - Serial correlation
  - Unit heterogeneity
  - Correlation between and within units

TSCS Data

- Time-series cross-sectional (panel) analyses
- Issues:
  - Serial correlation
  - Unit heterogeneity
  - Correlation between and within units
  - Small number of units (countries)

TSCS Data

- Time-series cross-sectional (panel) analyses
- Issues:
  - Serial correlation
  - Unit heterogeneity
  - Correlation between and within units
  - Small number of units (countries)
  - Small number of observations (time)

TSCS Data

- Time-series cross-sectional (panel) analyses
- Issues:
  - Serial correlation
  - Unit heterogeneity
  - Correlation between and within units
  - Small number of units (countries)
  - Small number of observations (time)
  - Not conditionally independent observations

Modeling TSCS Data

- Model as if independent, then adjust

Modeling TSCS Data

- Model as if independent, then adjust
- No default model choice

Modeling TSCS Data

- Model as if independent, then adjust
- No default model choice
- Solution: GPR

Modeling TSCS Data

- Model as if independent, then adjust
- No default model choice
- Solution: GPR
- Machine learning algorithm models outcomes jointly as a process

Current Practices

TSCS 3-year review: *APSR*, *AJPS*, *JOP*, *CPS* (320 articles)

Current Practices

TSCS 3-year review: *APSR, AJPS, JOP, CPS* (320 articles)

- Group and/or time fixed effects (FE) $\approx 48\%$ ($n = 154$)

Current Practices

TSCS 3-year review: *APSR, AJPS, JOP, CPS* (320 articles)

- Group and/or time fixed effects (FE) $\approx 48\%$ ($n = 154$)
- Robust clustered standard errors (RCSE) $\approx 54\%$ ($n = 172$)

Current Practices

TSCS 3-year review: *APSR, AJPS, JOP, CPS* (320 articles)

- Group and/or time fixed effects (FE) $\approx$ 48% ($n = 154$)
- Robust clustered standard errors (RCSE) $\approx$ 54% ($n = 172$)
- Panel corrected standard errors (PCSE) $\approx$ 5% ($n = 17$)

Current Practices

TSCS 3-year review: *APSR, AJPS, JOP, CPS* (320 articles)

- Group and/or time fixed effects (FE) $\approx$ 48% ($n = 154$)
- Robust clustered standard errors (RCSE) $\approx$ 54% ($n = 172$)
- Panel corrected standard errors (PCSE) $\approx$ 5% ($n = 17$)
- Lagged dependent variable (LDV) $\approx$ 7% ($n = 23$)

Current Practices

TSCS 3-year review: *APSR, AJPS, JOP, CPS* (320 articles)

- Group and/or time fixed effects (FE) $\approx$ 48% ($n = 154$)
- Robust clustered standard errors (RCSE) $\approx$ 54% ($n = 172$)
- Panel corrected standard errors (PCSE) $\approx$ 5% ($n = 17$)
- Lagged dependent variable (LDV) $\approx$ 7% ($n = 23$)
- Random effects (RE) $\approx$ 15% ($n = 48$)

Issues with Current Practice

- Serial correlation:

Issues with Current Practice

- Serial correlation:
  - Error correction does not change point estimates

Issues with Current Practice

- Serial correlation:
  - Error correction does not change point estimates
  - Fixed-effects are inefficient

Issues with Current Practice

- Serial correlation:
  - Error correction does not change point estimates
  - Fixed-effects are inefficient
  - LDV performs poorly

Issues with Current Practice

- Serial correlation:
  - Error correction does not change point estimates
  - Fixed-effects are inefficient
  - LDV performs poorly
  - RE only accounts for clustering

Issues with Current Practice

- Serial correlation:
  - Error correction does not change point estimates
  - Fixed-effects are inefficient
  - LDV performs poorly
  - RE only accounts for clustering
- Correlation between explanatory variable and unit intercept

Issues with Current Practice

- Serial correlation:
  - Error correction does not change point estimates
  - Fixed-effects are inefficient
  - LDV performs poorly
  - RE only accounts for clustering
- Correlation between explanatory variable and unit intercept
  - RE $\rightarrow$ false positives

Issues with Current Practice

- Serial correlation:
  - Error correction does not change point estimates
  - Fixed-effects are inefficient
  - LDV performs poorly
  - RE only accounts for clustering
- Correlation between explanatory variable and unit intercept
  - RE $\rightarrow$ false positives
- Small number of units

Issues with Current Practice

- Serial correlation:
  - Error correction does not change point estimates
  - Fixed-effects are inefficient
  - LDV performs poorly
  - RE only accounts for clustering
- Correlation between explanatory variable and unit intercept
  - RE $\rightarrow$ false positives
- Small number of units
  - FE and LDV $\rightarrow$ false negatives

Issues with Current Practice

- Serial correlation:
  - Error correction does not change point estimates
  - Fixed-effects are inefficient
  - LDV performs poorly
  - RE only accounts for clustering
- Correlation between explanatory variable and unit intercept
  - RE $\rightarrow$ false positives
- Small number of units
  - FE and LDV $\rightarrow$ false negatives
- No default for common issues

GPR Discussion

- Model outcomes jointly as a process

GPR Discussion

- Model outcomes jointly as a process
- Every point is a realization from a normal distribution

GPR Discussion

- Model outcomes jointly as a process
- Every point is a realization from a normal distribution
- Collection of outcomes is a multivariate normal

GPR Discussion

- Model outcomes jointly as a process
- Every point is a realization from a normal distribution
- Collection of outcomes is a multivariate normal
- Points close in the covariate space (e.g., time) share correlation

GPR Discussion

- Model outcomes jointly as a process
- Every point is a realization from a normal distribution
- Collection of outcomes is a multivariate normal
- Points close in the covariate space (e.g., time) share correlation
- Further distance $\rightarrow$ less correlation

GPR Discussion

- Model outcomes jointly as a process
- Every point is a realization from a normal distribution
- Collection of outcomes is a multivariate normal
- Points close in the covariate space (e.g., time) share correlation
- Further distance $\rightarrow$ less correlation
- Smooth surface

GPR Discussion

- Model outcomes jointly as a process
- Every point is a realization from a normal distribution
- Collection of outcomes is a multivariate normal
- Points close in the covariate space (e.g., time) share correlation
- Further distance $\rightarrow$ less correlation
- Smooth surface
- Correlation is unknown

GPR Discussion

- Model outcomes jointly as a process
- Every point is a realization from a normal distribution
- Collection of outcomes is a multivariate normal
- Points close in the covariate space (e.g., time) share correlation
- Further distance $\rightarrow$ less correlation
- Smooth surface
- Correlation is unknown
- Estimate variance-covariance in a learning kernel

GPR Discussion

- Model outcomes jointly as a process
- Every point is a realization from a normal distribution
- Collection of outcomes is a multivariate normal
- Points close in the covariate space (e.g., time) share correlation
- Further distance $\rightarrow$ less correlation
- Smooth surface
- Correlation is unknown
- Estimate variance-covariance in a learning kernel
- Very flexible, still interpretable

GPR Specification

- Typical: $\underbrace{y_i}_{\text{scalar}} \sim \mathcal{N}(\underbrace{\mathbf{x}_i}_{\text{vector}} \boldsymbol{\beta}, \underbrace{\sigma^2}_{\text{scalar}})$

GPR Specification

- Typical: $\underbrace{y_i}_{\text{scalar}} \sim \mathcal{N}(\underbrace{\mathbf{x}_i}_{\text{vector}} \boldsymbol{\beta}, \underbrace{\sigma^2}_{\text{scalar}})$
- GPR: $\underbrace{\mathbf{y}}_{\text{vector}} \sim \mathcal{MVN}(\underbrace{\mathbf{X}}_{\text{matrix}} \boldsymbol{\beta}, \sigma^2 \underbrace{\boldsymbol{\Omega}}_{\text{matrix}})$

GPR Specification

- Typical: $\underbrace{y_i}_{\text{scalar}} \sim \mathcal{N}(\underbrace{\mathbf{x}_i}_{\text{vector}} \boldsymbol{\beta}, \underbrace{\sigma^2}_{\text{scalar}})$
- GPR: $\underbrace{\mathbf{y}}_{\text{vector}} \sim \mathcal{MVN}(\underbrace{\mathbf{X}}_{\text{matrix}} \boldsymbol{\beta}, \sigma^2 \underbrace{\boldsymbol{\Omega}}_{\text{matrix}})$
- Mean is linear ($\mathbf{X}\boldsymbol{\beta}$)

GPR Specification

- Typical: $\underbrace{y_i}_{\text{scalar}} \sim \mathcal{N}(\underbrace{\mathbf{x}_i}_{\text{vector}}\boldsymbol{\beta}, \underbrace{\sigma^2}_{\text{scalar}})$
- GPR: $\underbrace{\mathbf{y}}_{\text{vector}} \sim \mathcal{MVN}(\underbrace{\mathbf{X}}_{\text{matrix}}\boldsymbol{\beta}, \sigma^2\underbrace{\boldsymbol{\Omega}}_{\text{matrix}})$
- Mean is linear ($\mathbf{X}\boldsymbol{\beta}$)
- Include unit indicators in $\mathbf{X}$

GPR Specification

- Typical: $\underbrace{y_i}_{\text{scalar}} \sim \mathcal{N}(\underbrace{\mathbf{x}_i}_{\text{vector}} \boldsymbol{\beta}, \underbrace{\sigma^2}_{\text{scalar}})$
- GPR: $\underbrace{\mathbf{y}}_{\text{vector}} \sim \mathcal{MVN}(\underbrace{\mathbf{X}}_{\text{matrix}} \boldsymbol{\beta}, \sigma^2 \underbrace{\boldsymbol{\Omega}}_{\text{matrix}})$
- Mean is linear ($\mathbf{X}\boldsymbol{\beta}$)
- Include unit indicators in $\mathbf{X}$
- Posteriors on $\boldsymbol{\beta}$ standard inferences

GPR Specification

- Typical: $\underbrace{y_i}_{\text{scalar}} \sim \mathcal{N}(\underbrace{\mathbf{x}_i}_{\text{vector}} \boldsymbol{\beta}, \underbrace{\sigma^2}_{\text{scalar}})$
- GPR: $\underbrace{\mathbf{y}}_{\text{vector}} \sim \mathcal{MVN}(\underbrace{\mathbf{X}}_{\text{matrix}} \boldsymbol{\beta}, \sigma^2 \underbrace{\boldsymbol{\Omega}}_{\text{matrix}})$
- Mean is linear ($\mathbf{X}\boldsymbol{\beta}$)
- Include unit indicators in $\mathbf{X}$
- Posteriors on $\boldsymbol{\beta}$ standard inferences
- Need to estimate variance-covariance $\boldsymbol{\Omega}$ from data

GPR Specification

- Typical: $\underbrace{y_i}_{\text{scalar}} \sim \mathcal{N}(\underbrace{\mathbf{x}_i}_{\text{vector}} \boldsymbol{\beta}, \underbrace{\sigma^2}_{\text{scalar}})$
- GPR: $\underbrace{\mathbf{y}}_{\text{vector}} \sim \mathcal{MVN}(\underbrace{\mathbf{X}}_{\text{matrix}} \boldsymbol{\beta}, \sigma^2 \underbrace{\boldsymbol{\Omega}}_{\text{matrix}})$
- Mean is linear ($\mathbf{X}\boldsymbol{\beta}$)
- Include unit indicators in $\mathbf{X}$
- Posteriors on $\boldsymbol{\beta}$ standard inferences
- Need to estimate variance-covariance $\boldsymbol{\Omega}$ from data
- Data in mean and $\boldsymbol{\Omega}$ do not have to be the same

Gaussian Kernel

- Variance-covariance estimated from data:

$$\mathbf{y} \sim \mathcal{MVN}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{\Omega}),$$

Gaussian Kernel

- Variance-covariance estimated from data:

$$\mathbf{y} \sim \mathcal{MVN}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{\Omega}),$$

$$\boldsymbol{\Omega}(\mathbf{x}_j, \mathbf{x}_i) = \exp\left\{-\sum_{p=1}^{m} |x_{pj} - x_{pi}|^2\right\}.$$

Gaussian Kernel

- Variance-covariance estimated from data:

$$\mathbf{y} \sim \mathcal{MVN}(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{\Omega}),$$

$$\boldsymbol{\Omega}(\mathbf{x}_j, \mathbf{x}_i | \zeta) = \exp\left\{ -\sum_{p=1}^{m} \frac{|x_{pj} - x_{pi}|^2}{\zeta_p} \right\}.$$

Gaussian Kernel

- Variance-covariance estimated from data:

$$\mathbf{y} \sim \mathcal{MVN}(\tilde{\mathbf{X}}\beta, \sigma^2\boldsymbol{\Omega}),$$

$$\boldsymbol{\Omega}(\mathbf{x}_j^*, \mathbf{x}_i^* | \zeta) = \exp\left\{ -\sum_{p=1}^{m} \frac{|x_{pj}^* - x_{pi}^*|^2}{\zeta_p} \right\}.$$

Motivating Example

- Consider inflation in Latin America

Motivating Example

- Consider inflation in Latin America
- Fit a two-way fixed effects model, an LDV model, and a GPR

Motivating Example

- Consider inflation in Latin America
- Fit a two-way fixed effects model, an LDV model, and a GPR
- Only use as inputs country indicator and time

Motivating Example

- Consider inflation in Latin America
- Fit a two-way fixed effects model, an LDV model, and a GPR
- Only use as inputs country indicator and time
- Inflation

Motivating Example

- Consider inflation in Latin America
- Fit a two-way fixed effects model, an LDV model, and a GPR
- Only use as inputs country indicator and time
- Inflation
  - Different baselines

Motivating Example

- Consider inflation in Latin America
- Fit a two-way fixed effects model, an LDV model, and a GPR
- Only use as inputs country indicator and time
- Inflation
  - Different baselines
  - Time trends

Motivating Example

- Consider inflation in Latin America
- Fit a two-way fixed effects model, an LDV model, and a GPR
- Only use as inputs country indicator and time
- Inflation
  - Different baselines
  - Time trends

Motivating Example

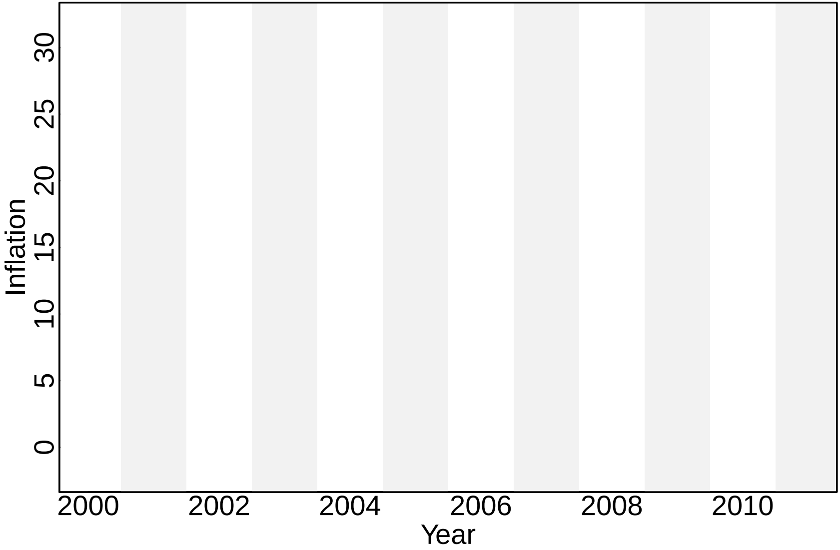
- Consider inflation in Latin America
- Fit a two-way fixed effects model, an LDV model, and a GPR
- Only use as inputs country indicator and time
- Inflation
  - Different baselines
  - Time trends . . . but

Motivating Example

- Consider inflation in Latin America
- Fit a two-way fixed effects model, an LDV model, and a GPR
- Only use as inputs country indicator and time
- Inflation
  - Different baselines
  - Time trends ... but different trends within countries

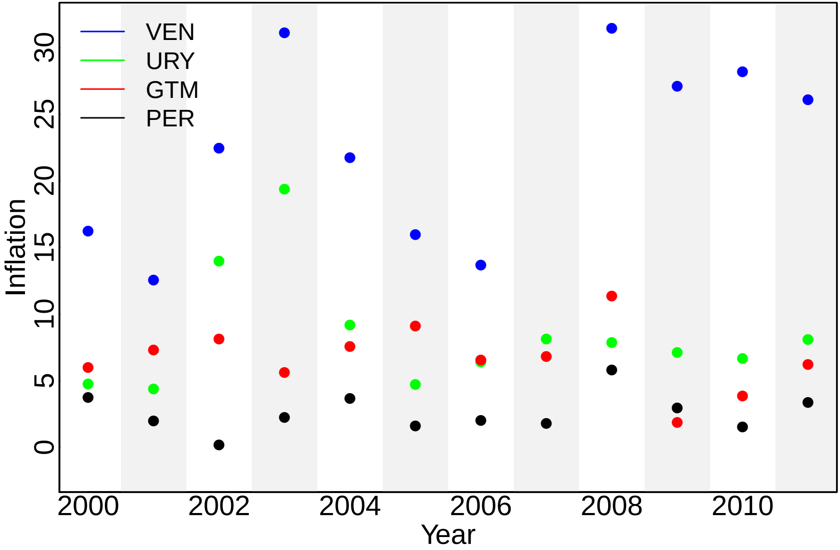
Inflation in Latin America

Two-way fixed effects models assume same time trends



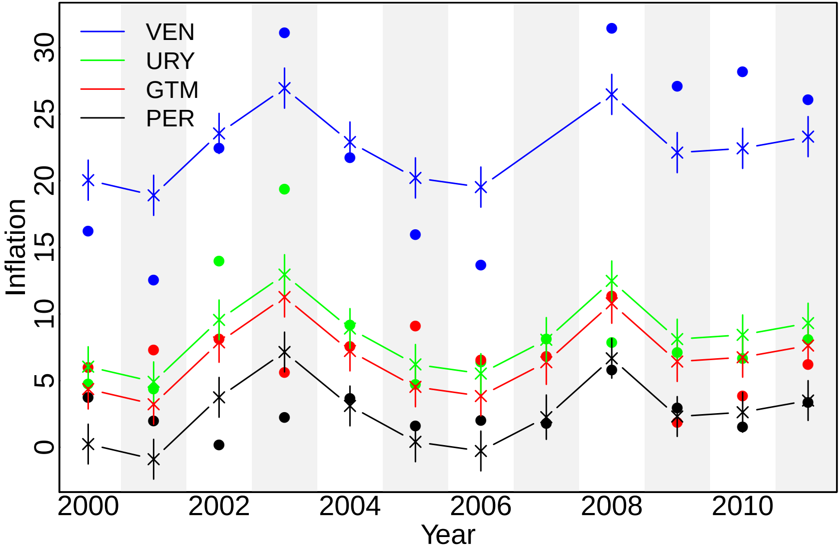
Inflation in Latin America

Two-way fixed effects models assume same time trends



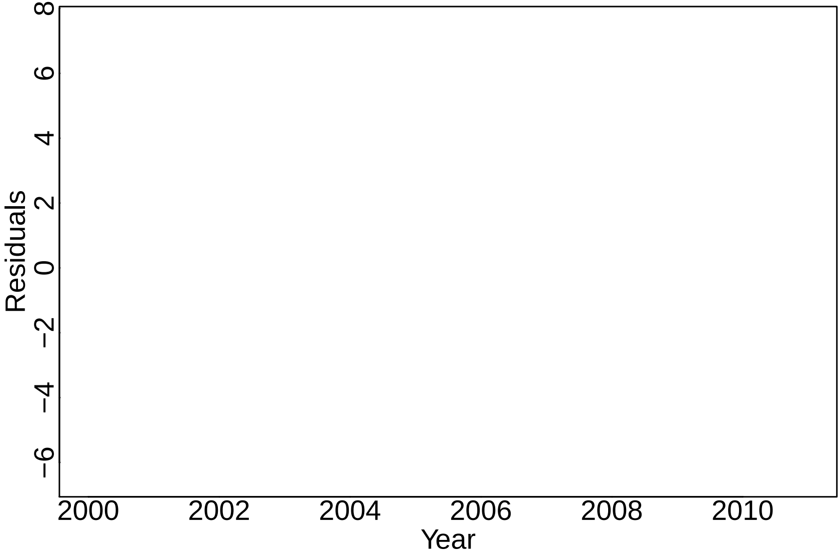
Inflation in Latin America

Two-way fixed effects models assume same time trends



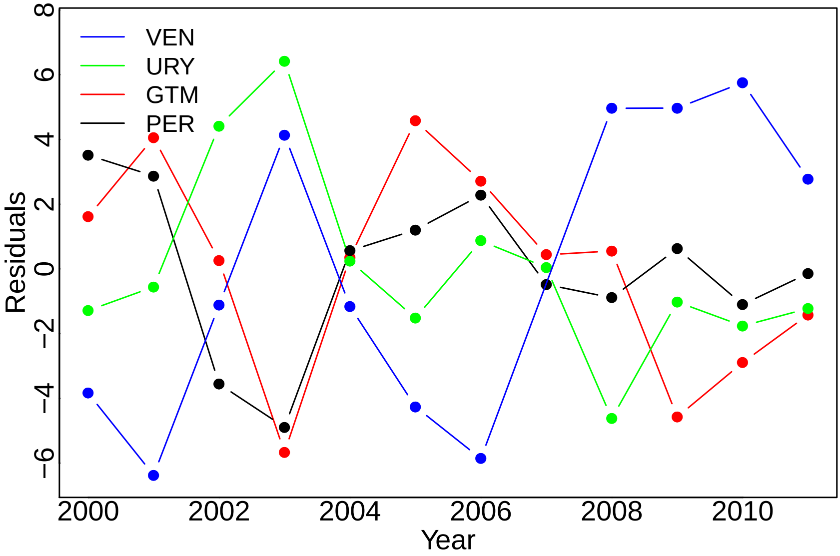
Inflation in Latin America

Residuals in fixed effects model show trends



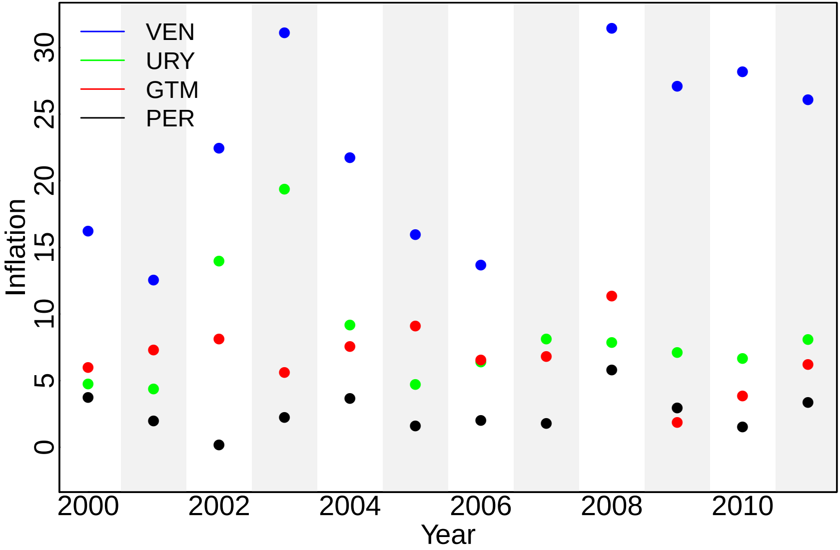
Inflation in Latin America

Residuals in fixed effects model show trends



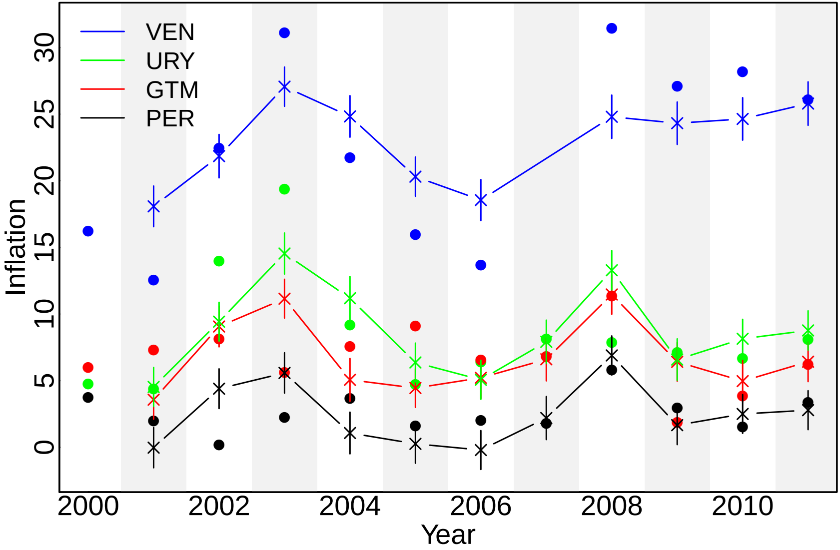
Inflation in Latin America

Two-way fixed effects models with LDV inaccurate



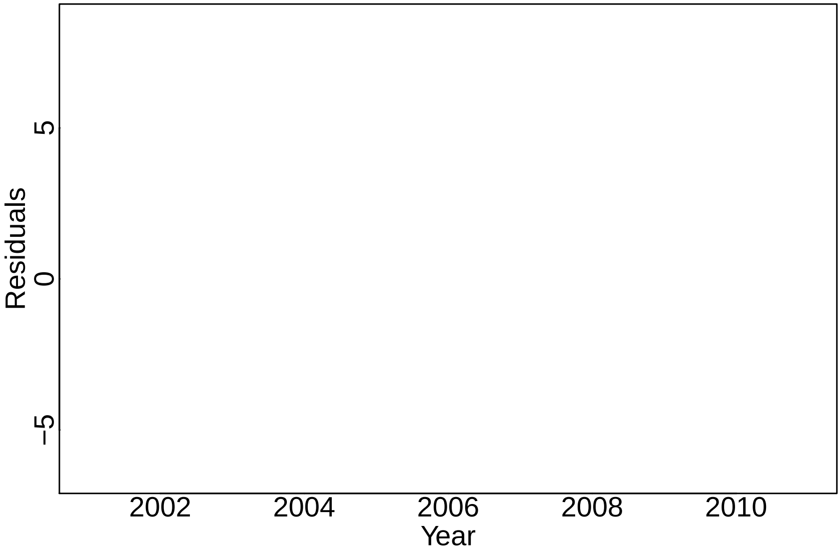
Inflation in Latin America

Two-way fixed effects models with LDV inaccurate



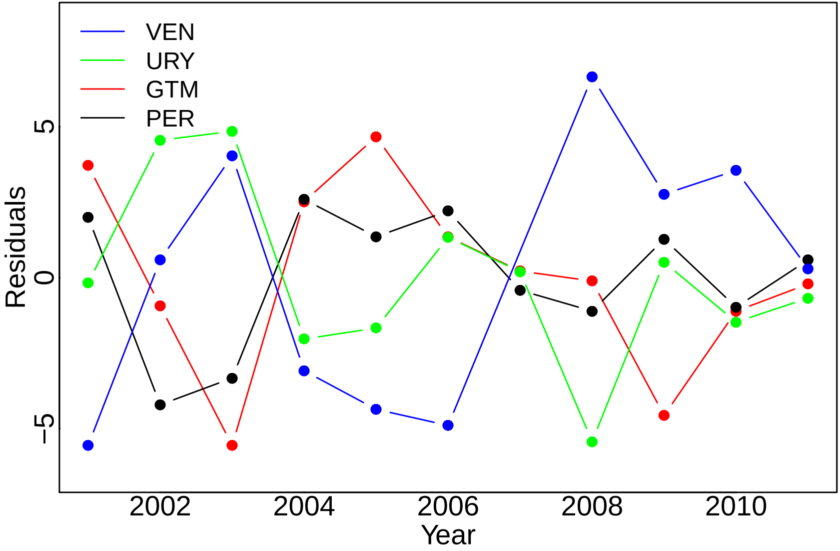
Inflation in Latin America

Residuals in LDV model show same trends



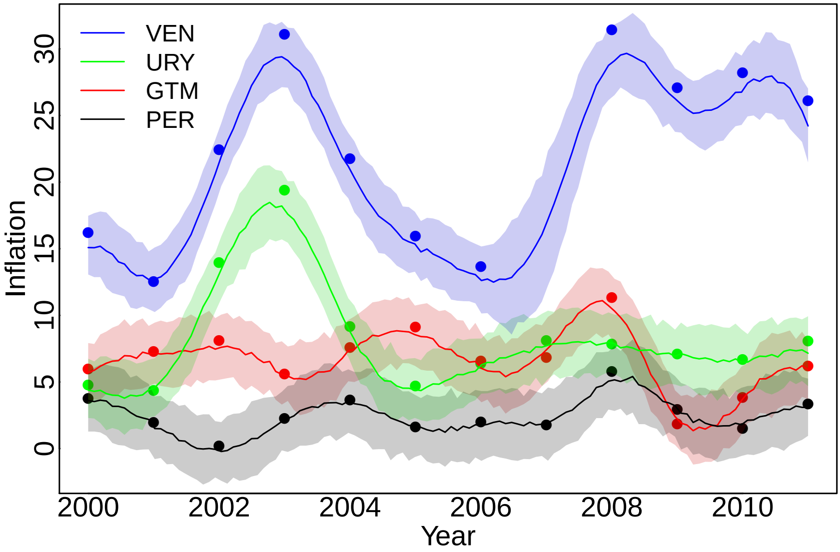
Inflation in Latin America

Residuals in LDV model show same trends



Inflation in Latin America

GPR allows correlated but different time trends



GPR for Projection

- Predict counter-factual covariates

GPR for Projection

- Predict counter-factual covariates
- Predict counter-factual outcome

GPR for Projection

- Predict counter-factual covariates
- Predict counter-factual outcome
- Compare counter-factual to actual outcome

GPR for Projection

- Predict counter-factual covariates
- Predict counter-factual outcome
- Compare counter-factual to actual outcome
- Model the explanatory variables and missing explanatory variables with random effects and a GP

GPR for Projection

- Predict counter-factual covariates
- Predict counter-factual outcome
- Compare counter-factual to actual outcome
- Model the explanatory variables and missing explanatory variables with random effects and a GP
- Model the counter-factual outcome similarly

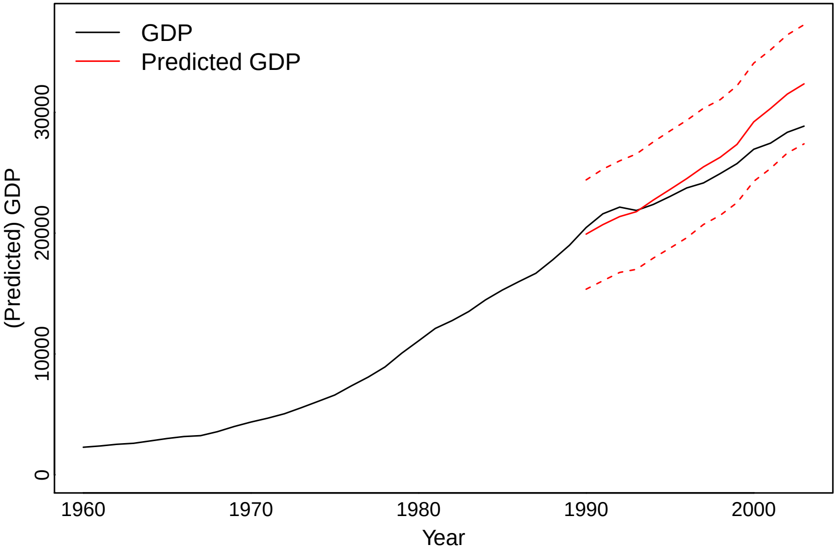
GPR for Projection

- Predict counter-factual covariates
- Predict counter-factual outcome
- Compare counter-factual to actual outcome
- Model the explanatory variables and missing explanatory variables with random effects and a GP
- Model the counter-factual outcome similarly
- Uncertainty carries over due to joint modeling

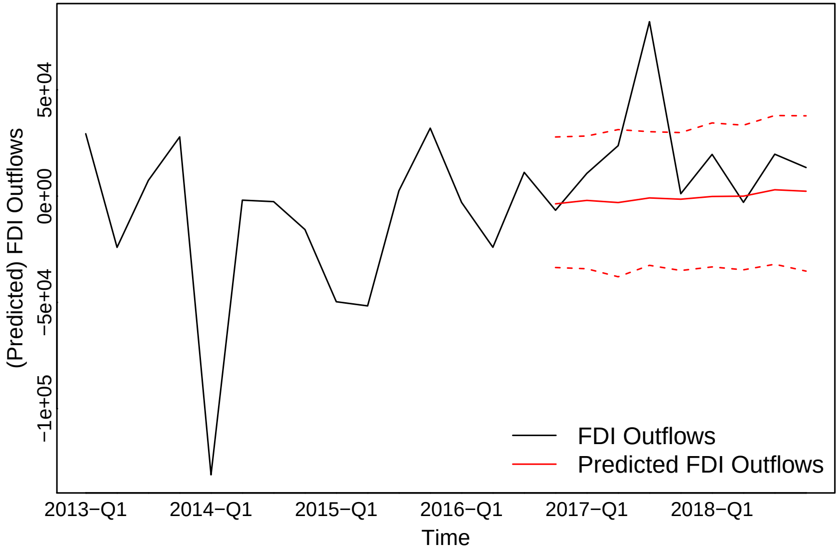
GPR for Projection

- Predict counter-factual covariates
- Predict counter-factual outcome
- Compare counter-factual to actual outcome
- Model the explanatory variables and missing explanatory variables with random effects and a GP
- Model the counter-factual outcome similarly
- Uncertainty carries over due to joint modeling
- For the applications, stationary kernel for explanatory, non-stationary kernel for outcome

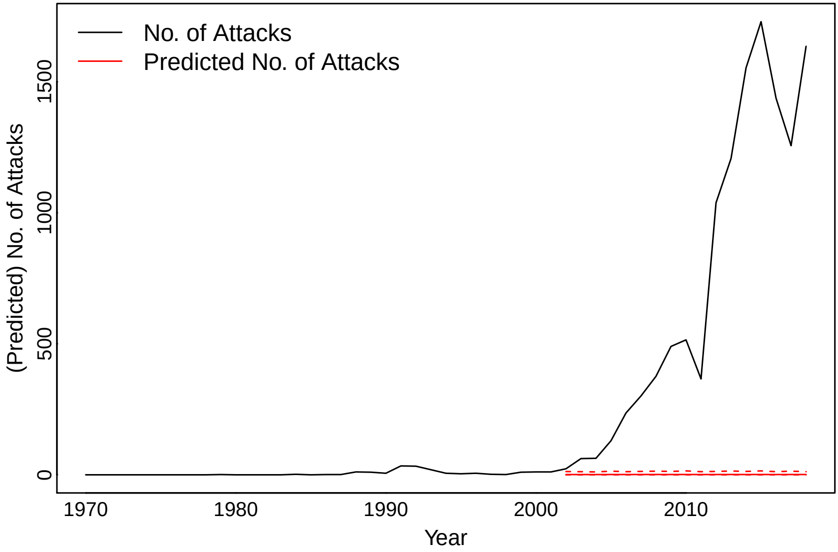
GPP Findings on Reunification



GPP Findings on UK Capital Flight



GPP Findings on Afghan Terror Attacks Following US War



Other GP Uses

- Standard ML regression or classification tasks

Other GP Uses

- Standard ML regression or classification tasks
- Geographic smoothing (Bayesian kriging)

Other GP Uses

- Standard ML regression or classification tasks
- Geographic smoothing (Bayesian kriging)
- Measurement models

Other GP Uses

- Standard ML regression or classification tasks
- Geographic smoothing (Bayesian kriging)
- Measurement models
- Entropy estimation in hypothetical information tasks

Other GP Uses

- Standard ML regression or classification tasks
- Geographic smoothing (Bayesian kriging)
- Measurement models
- Entropy estimation in hypothetical information tasks
- Ranking tasks

Other GP Uses

- Standard ML regression or classification tasks
- Geographic smoothing (Bayesian kriging)
- Measurement models
- Entropy estimation in hypothetical information tasks
- Ranking tasks
- Causal ordering tasks

Other GP Uses

- Standard ML regression or classification tasks
- Geographic smoothing (Bayesian kriging)
- Measurement models
- Entropy estimation in hypothetical information tasks
- Ranking tasks
- Causal ordering tasks
- Mapping