Estimating a Counter-Factual with Uncertainty Through Gaussian Process Projection

David Carlson

Koç University

dcarlson@ku.edu.tr

ABSTRACT

Estimating a counter-factual in which a treatment did not occur allows political science researchers to better understand the effect of an intervention. To date, the most prominent attempt in the literature has been the introduction of the synthetic control method. However, this method has important and related drawbacks. Perhaps most importantly, the synthetic control method does not lend well to estimates of uncertainty, making traditional hypothesis testing ad hoc. I develop a new method, Gaussian process projection (GPP), that circumvents the issues associated with the synthetic control method. By comparing the projected unit to the true unit, we can estimate the causal effect of an intervention, and by using the measure of uncertainty, we can determine how unlikely the difference would be due to chance. After demonstrating the usefulness of the method in traditional settings, I generalize the method to allow for multiple treated units, treated at different times, and allow for link functions for applications with discrete outcomes.

1. INTRODUCTION

Predicting a counter-factual in which a treatment did not occur allows political scientists to estimate a treatment effect in observational studies. However, this has long-eluded practitioners of the discipline. A major step forward was the introduction of the synthetic control method (Abadie, Diamond, and Hainmueller 2015), though the method has several drawbacks.

First, the synthetic control method does not lend well to estimates of uncertainty, making the testing of hypotheses difficult. Second, by disregarding several untreated units, data is being wasted, and the conclusions are highly sensitive to the underlying assumptions. For example, if an untreated unit experiences a similarly timed change in the outcome variable, the synthetic unit will be a biased estimate of the counter-factual. Further, the method is based on the assumptions of a linear model, specifically the assumptions of conditional independence, homoskedasticity, stationarity, and linearity. Finally, a degree of precision is lost in the estimator by (correctly) not accounting for post-treatment control variables.

This paper develops a new method, Gaussian process projection (GPP), that circumvents these issues. As discussed in Carlson (2018), Gaussian process regression (GPR) is very well-adept at predicting missing data in time-series cross-sectional (TSCS) studies. Gaussian processes have also long been used in projecting time-series data (e.g. Rasmussen and Williams 2006; Kirk and Stumpf 2009; Qian, Zhou, and Rudin 2011; Gibson et al. 2012). Combining these two applications intuitively leads to an estimator that can predict a counter-factual in TSCS applications, by taking into account the temporal trends of the predictors and the outcome of interest across and within units of analysis.

GPP is a non-parametric Bayesian method that leads to full posteriors of the counter-factual outcome, allowing estimation of the treatment effect, and estimation of the uncertainty, allowing for hypothesis testing. It is less sensitive to variations in the outcome of the "donor pool," relaxes the assumption of conditional independence, utilizes all of the available data, and does not ignore post-treatment predictors. Of course, post-treatment predictors

for the outcome of interest are not used, but instead the counter-factual predictors had there been no treatment are jointly estimated with the outcome. This naturally also allows for imputation of missing variables in a manner consistent with TSCS data. The uncertainty caused by this imputation is a part of the joint modeling process and thus is included in the uncertainty of the outcome estimation.

The following section explains the spirit and intuition of the method, with contrasts and comparisons to extant methods. I then fully specify the model for the identity link. There are then two applications in the same vein as synthetic control method applications. The first replicates the study in the synthetic control method introduction (Abadie, Diamond, and Hainmueller 2015), finding that there is insufficient evidence that German reunification had a substantial impact on West German GDP, but there is suggestive evidence that it hurt growth in the long-term. The second application analyzes FDI outflows in The United Kingdom following the Brexit vote to demonstrate that there was an almost-immediate capital flight despite there being no immediate policy change. I then generalize the method to allow for multiple treated units at different times, and demonstrate that there is again insufficient evidence that joining the European Union affects financial stability. The final application demonstrates how to generalize the method to include a link function, and shows strong evidence that the US war in Afghanistan increased terrorist attacks globally by Afghan nationals, using a negative-binomial link. The final section concludes. All models are run in Stan (Carpenter et al. 2016) through `R`.

2. THE UNDERLYING PRINCIPLES OF GPP

GPR, as demonstrated in Carlson (2018), is arguably the best default alternative for analyzing TSCS data. The method relaxes the conditional independence assumption by smoothing over the covariate space, meaning that points close in the hyperplane of covariates share a correlation, while those further away share less correlation (see also Monogan and Gill 2016). This correlation need not be assumed or strictly based only on the distance. In other words,

time one and two need not share the same correlation as location one and location two, or time two and time three. Instead, the degree of correlation is learned from the data through a machine learning algorithm. The prior is known as a Gaussian process, and it outperforms fixed effects, random effects, a lagged dependent variable model, and more, across a range of commonly encountered situations in TSCS data.

Carlson (2018) also demonstrates that GPR is exceptionally good at predicting observations that are missing. Not only is the correlation of missing outcomes and predictions about 0.90, but the 0.95 credible interval coverage probablities are also approximately 0.95. Of course this is ideal, and impressive given the nature of messy political science TSCS data. However, predicting within the time horizon is a slightly distinct problem from predicting beyond the time horizon. Gaussian processes, in other disciplines, have long been used to predict future time-series data (e.g. Rasmussen and Williams 2006; Kirk and Stumpf 2009; Qian, Zhou, and Rudin 2011; Gibson et al. 2012). Perhaps their main rival is long short term memory (LSTM) recurrent neural networks (Hochreiter and Schmidhuber 1997). These LSTM models do not lend well to estimates of uncertainty, making GPP perhaps the best modeling strategy for the outlined goals.

While projecting outcomes into the future is not a new application, estimating a treatment effect, borrowing information from untreated units, projecting predictors into the future to increase precision, and imputing missing values are all novel additions to the Gaussian process literature. All of these are accomplished via simple extensions of standard GPR. Because the projection into the future has generally been literally into the future, borrowing the information from untreated units has not been possible. However, in this case the projection is a counter-factual, meaning we can account for global trends by including in the analysis untreated units beyond treatment time. Further, these temporal trends within and between units can allow us to simulateously impute missing predictors, project predictors, and ultimately project the counter-factual outcome while incorporating all of the uncertainty from these estimations into the final projection.

2.1. *Assumptions*

The assumptions for GPP to estimate a counter-factual are similar to the assumptions of the synthetic control method, though perhaps less strict. For example, it is assumed in the synthetic control method that no other units in the study are affected by the treatment, or the outcome variable of other units did not change due to a confounder (Abadie, Diamond, and Hainmueller 2015). While this assumption is still needed in GPP, because the projection is based on a much larger set of donor units, it is much less sensitive to deviations from this assumption. As I will discuss in more detail below in the application section, at the same time as German reunification, Norway experienced a spike in GDP growth due to their oil reserves. Despite this, Norway need not be excluded from the study, because its inclusion or exclusion do not substantially, or even moderately, impact the conclusions.

3. GPP SPECIFICATION (IDENTITY LINK)

This section fully specifies the identity link GPP model. The first subsection explains the model's ability to impute missing predictors and project post-treatment predictors. The following section discusses projecting the counter-factual outcome. Finally, I specify the prior distributions for the parameters.

3.1. *Imputing Missing Predictors and Projecting Post-Treatment Predictors*

Consider a single explanatory variable $\mathbf{x}$. There may be missingness in the variable, and we should not include post-treatment values of the variable in the analysis. We therefore model the missingness as a Gaussian process, accounting for temporal trends within and between units of analysis. This is a different approach but similar in spirit to Honaker and King (2010).

Let $\mathbf{x}^*$ indicate the explanatory variable with missingness modeled. In other words, it

can be considered a latent variable in the modeling process. The $\mathbf{x}^*$ is modeled with an intercept (β_0^x), a year random effect ($\boldsymbol{\gamma}^x$), a unit random effect ($\boldsymbol{\lambda}^x$), and a GP term ($\boldsymbol{\eta}^x$), smoothing between and within units over time. As I will discuss in the subsection on priors, I give this GP term a seperable exponential kernel (Rasmussen and Williams 2006). This is a stationary GP kernel, and can be changed based on the properties of the variable of interest.

Let t index time, and i index observational unit. The latent variable $\mathbf{x}^*$ will simply take the value of $\mathbf{x}$ where it is not missing or post-treatment. The imputation of missing values takes the following form:

$$x^*_{t,i} \sim \mathcal{N}(\beta_0^x + \gamma_t^x + \lambda_i^x + \eta_{t,i}^x, \sigma_x). \tag{1}$$

To test the modeling of missing values, I both randomly and systematically remove values from the first application, and the coverage is almost exactly 0.95, while the correlation with the true values is about 0.98. This is encouraging that both the estimates are accurate, and the degree of uncertainty is ideal in the modeling decision. The modeling of the projected outcome is quite similar, as I will discuss in the following subsection.

3.2. *Projecting the Counter-Factual Outcome*

Similar to modeling missing explanatory variables, projecting the counter-factual can be thought of as modeling a missing variable. Let $\mathbf{y}$ indicate the outcome variable with missingness, and let $\mathbf{y}^*$ indicate the modeled outcome variable with imputed values. The goal is to model the $\mathbf{y}^*$ to make inferences about what the outcome would be absent treatment.

Let all of the explanatory variables, with joint imputation, be indicated by $\mathbf{X}^*$ as a matrix of (estimated) predictors. I model the mean of the outcome with an intercept (β_0^y), a vector of parameters multiplied by the predictors ($\boldsymbol{X^*\beta^y}$), a time random effect ($\boldsymbol{\gamma}^y$), a unit random effect ($\boldsymbol{\lambda}^y$), and a GP term ($\boldsymbol{\eta}^y$) temporally smoothing within and across units.

Because the outcome variable in the two applications is likely non-stationary, I use a dot product non-stationary kernel for this GP term. This can again be changed based on the specific application. Where $\mathbf{y}$ is observed or not post-treatment, $\mathbf{y}^*$ takes its value. The missing or post-treatment outcome values are modeled in the following way:

$$y^*_{t,i} \sim \mathcal{N}(\beta^y_0 + \boldsymbol{X^*\beta^y} + \gamma^y_t + \lambda^y_i + \eta^y_{t,i}, \sigma_y). \tag{2}$$

These are modeled jointly with the missing predictors such that uncertainty in the outcome contains the necessary uncertainty deriving from the projection and imputation of the predictors.

However, due to the peculiarities of any given application, and GPR's tendency to overfit, noise is added to model the final outcome levels, $\hat{\boldsymbol{y}}$:

$$\hat{y}_{t,i} \sim \mathcal{N}(y^*_{t,i}, \psi). \tag{3}$$

The value of ψ is determined using placebo tests in order to attain 0.95 coverage of the 0.95 reliable intervals. I discuss the specifics of the process below. I will now fully specify the model by discussing the prior distributions, including the Gaussian process prior specification.

3.3. *Priors*

Before I discuss the prior specification, it is important to note that all variables, including time, are scaled so as to make the prior specification insensitive to the scale of the data. The results are then rescaled back to their original scale to make inferences about the counterfactual outcome. As I mention above, Gaussian processes (GPs) smooth over the parameter space, in this case over time within and between units of observation. I use two different

GPs to specify the model: a stationary GP using the seperable-exponential kernel, and a non-stationary GP using the dot product kernel. These kernels are used as priors for a multivariate normal distribution, which models all the outcomes jointly as a process. This modeling strategy allows and estimates a correlation between units and over time.

The explanatory variables in the applications below use the seperable exponential kernel. In order to smooth over our covariate space, we want observations that are close in the covariate space to share a correlation, while points further in the covariate space to share less correlation. Therefore, we sum the squared distances across input columns. We then take the exponential of the negative sum to capture the correlation. The variance-covariance of the multivariate normal distribution is indicated by $\mathbf{\Omega}$. Let m indicate the number of columns of the input matrix to the kernel, and let p index these columns in the summation. Also let j and i indicate two observations. Recall that x^*'s are modeled explanatory variables. Let z be inputs to impute the x^* values (time and an indicator for location). We start with setting the correlations according to:

$$\mathbf{\Omega}(\mathbf{z}_j, \mathbf{z}_i) = \exp\left\{-\sum_{p=1}^{m}(z_{pj} - z_{pi})^2\right\}. \tag{4}$$

However, we want time one and time two to perhaps have a different correlation than location one and location two, for example. For this purpose, we introduce another parameter to the kernel, $\boldsymbol{\zeta}$, a vector learned from the data that estimates the degree of correlation relative to distance for each dimension. The kernel then becomes:

$$\mathbf{\Omega}(\mathbf{z}_j, \mathbf{z}_i|\boldsymbol{\zeta}) = \exp\left\{-\sum_{p=1}^{m}\frac{(z_{pj} - z_{pi})^2}{\zeta_p}\right\}, \tag{5}$$

which is a symmetric matrix with all diagonal elements equal to one, because the sum of the distance of the covariate space will be zero for elements that are identical, and all off-diagonal elements represent a correlation. However, to ensure the matrix is non-singular, and

to separate signal from noise, we add a small estimated constant to the diagonal elements. This is necessary because points very close in the covariate space will have almost one in the kernel matrix, and this leads to numerical instabilities. This is called the nugget or jitter.

To distinguish between inputs to the mean and inputs to the kernel, allow $\tilde{\mathbf{Z}}$ to indicate inputs to the mean and $\mathbf{Z}^*$ to indicate inputs to the kernel. The fully specified model is therefore:

$$\mathbf{x}^* \sim \mathcal{MVN}(\tilde{\mathbf{Z}}\boldsymbol{\beta}, \sigma^2\boldsymbol{\Omega}), \tag{6}$$

$$\boldsymbol{\Omega}(\mathbf{z}_j^*, \mathbf{z}_i^*) = \boldsymbol{\Omega}^\dagger(\mathbf{z}_j^*, \mathbf{z}_i^*) + \delta g_{j,i}, \tag{7}$$

$$g_{i,j} = \begin{cases} 1 & \text{if } j = i, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

$$\boldsymbol{\Omega}^\dagger(\mathbf{z}_j^*, \mathbf{z}_i^* | \boldsymbol{\zeta}) = \exp\left\{ -\sum_{p=1}^{m} \frac{(z_{pj} - z_{pi})^2}{\zeta_p} \right\}. \tag{9}$$

The outcome (recall $\hat{\mathbf{y}}$ indicates the outcome and predicted outcome for the counterfactual) is modeled with a non-stationary kernel, known as the dot product kernel. The intuition is very similar to the seperable exponential kernel. Points similar in the covariate space share a greater correlation in the prior than do points further away. However, the form of the prior takes a different form to allow for the non-stationarity.

Again, let $\mathbf{z}$ indicate the inputs to the outcome kernel. We take the dot product of the kernel inputs for two observations and add an estimated parameter θ in order to place a prior on the variance-covariance for the two observations. More formally, the variance-covariance prior between the observations is given as follows, again letting $\boldsymbol{\Omega}$ indicate the full variance-covariance prior matrix:

$$\boldsymbol{\Omega} = \theta + \mathbf{z} \cdot \mathbf{z}'. \tag{10}$$

Now that the GP priors are specified, I will specify the remaining priors for all the parameters and hyperparameters discussed thusfar.

The variance that pre-multiplies the variance-covariance priors for both the outcome and explanatory variables is inverse-gamma distributed: $\sigma \sim \mathcal{IG}(1, 1)$. The length-scale of both GPs (the prior on the degree of correlation, $\boldsymbol{\zeta}$) has a Weibull prior, $\boldsymbol{\zeta} \sim \mathcal{W}(30, 8)$. This is an appropriate prior for the number of units studied in the applications, but may need to be adjusted given the number of units. The standard deviation of the GPs have a normal prior, $\sqrt{\delta} \sim \mathcal{N}(0, 1)$.

All of the random effects, for unit and time and for the outcome and explanatory variables, follow a normal distribution with mean zero and a hyperprior for the variance, which follows a gamma distribution $\mathcal{G}(2, \frac{1}{2})$. The β's, including intercepts, all have a normal distribution centered at zero with variance three. Finally, the variance of the outcome is distributed inverse-gamma, $\mathcal{IG}(1, 1)$. Having specified the identity link model, I will now apply it to estimate the impact of two treatments.

4. APPLYING THE IDENTITY LINK GPP

In this section, I apply the identity link model to test two hypotheses. The first is a replication of the synthetic control introductory paper (Abadie, Diamond, and Hainmueller 2015). That is, did German reunification affect the West German GDP? The second application tests a novel hypothesis. Specifically, did the signal of the Brexit vote cause capital flight out of the UK despite there being no change in policy?

4.1. *West German GDP and Reunification*

In Abadie, Diamond, and Hainmueller (2015), West German GDP after reunification is compared to a counter-factual using the synthetic control method. While the method does not lend well to estimates of uncertainty and therefore traditional hypothesis testing is ad-

hoc, they find that GDP was hurt in the longer-term relative to the counter-factual.

There is reason to re-investigate their findings. First, methodologically, the model is highly sensitive to post-treatment donor units' changes in the explanatory variable. As an example, Norway is included in the analysis, but due to an oil boom, the GDP of Norway increased beyond expectations in the period under study. GPP is much less sensitive to these deviations from the underlying assumptions. Second, I will reiterate that the synthetic control method does not lend well to hypothesis testing. Therefore, the study warrants additional attention.

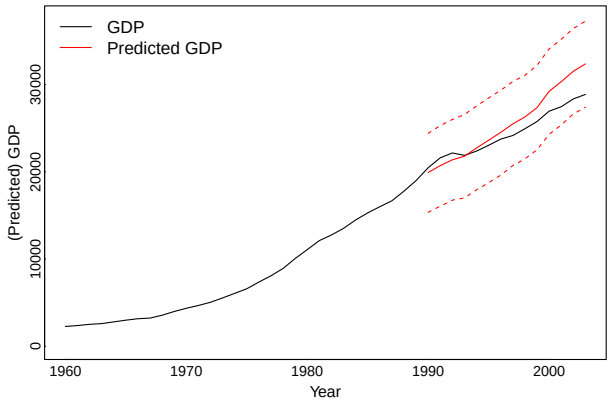
Theoretically, reunification provides a larger market, but integrating the two very different German sides was a challenge, with the two operating under very different market structures. With this in mind, I still expect GDP growth to slow in the long-term, but I do not necessarily expect there to be sufficient evidence that this slow was not due to chance.

I maintain all controls as the original analysis. These are standard predictors of growth rate, and include per capita GDP, inflation rate, industry share of value added, investment rate, schooling, and a measure of trade openness. 16 OECD countries are used for the counter-factual estimation. To test my claim that GPP is insensitive to deviations from the assumptions, I run the model both including and excluding Norway. Because of the insensitivity of the modeling strategy to outliers, however, conclusions are not affected. All models are run in Stan (Carpenter et al. 2016) through `R`.

The model results can be seen in Figure 1. It shows West German GDP before and after reunification in black. The red solid line is the predicted point estimates of the post-treatment counter-factual. The dashed red lines represent the 0.95 posterior. As can be seen, GDP growth slowed down slightly in the long-term after reunification, but there is insufficient evidence to reject the null hypothesis of no effect at traditional reliability levels.

To test the model's validity, and determine the appropriate level of noise to add, ψ, I run placebo models. That is, I eliminate the outcome and explanatory variables post-1989 for one country, and run the model. I repeat this for all remaining countries. Ideally, the point estimates will be close to the true GDP since they experienced no treatment, and the

Figure 1: West German GDP and Predicted Counter-Factual GDP



Note: The figure shows West German GDP before and after reunification in black. The red solid line is the predicted point estimates of the post-treatment counter-factual. The dashed red lines represent the 0.95 posterior. As can be seen, GDP growth slowed down slightly in the long-term after reunification, but not at traditionally reliable levels.

0.95 posterior should capture, across models, approximately 0.95 of the true GDP. In these results, the coverage probability is roughly 0.95 with a noise level of $\psi = 0.25$. The results of this exercise are shown in Figure 2. It shows actual GDP before and after reunification in black across the countries in the study. The red solid line is the predicted point estimates of the post-treatment counter-factual. The dashed red lines represent the 0.95 posterior. As can be seen, with the exception of Norway, the model is quite accurate with its predictions.

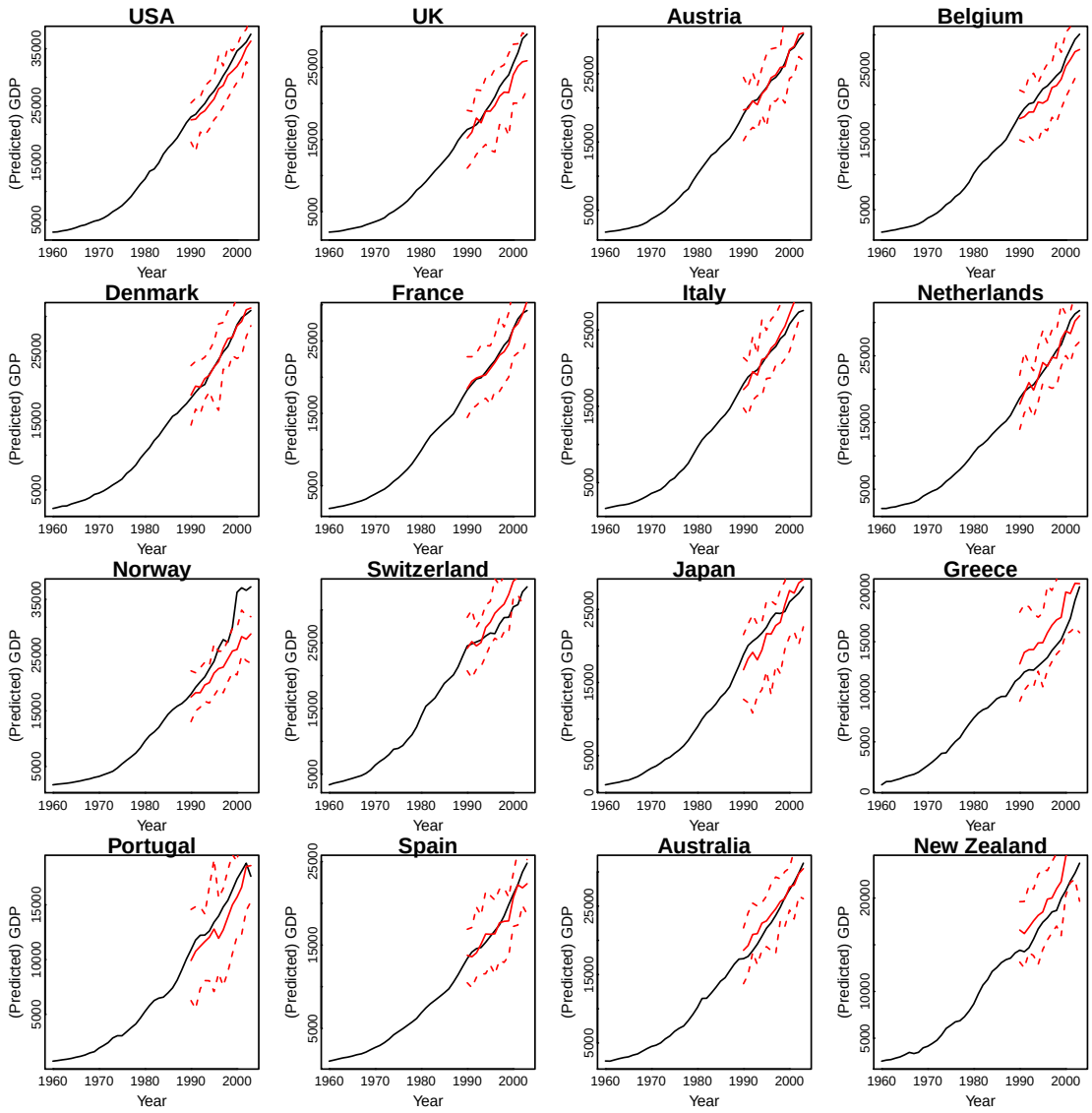
These findings are interesting, and should cast doubt on the original findings in Abadie, Diamond, and Hainmueller (2015). I will now move on to the next application regarding capital flight following the Brexit vote.

4.2. *UK Capital Flight Following the Brexit Vote*

There is speculation that Brexit will increase capital flight out of the UK (Rosefielde 2016). However, the mere vote, despite having no immediate impact on policy, has not been thoroughly examined. It is sensible that the signal of the vote caused capital flight on its own.

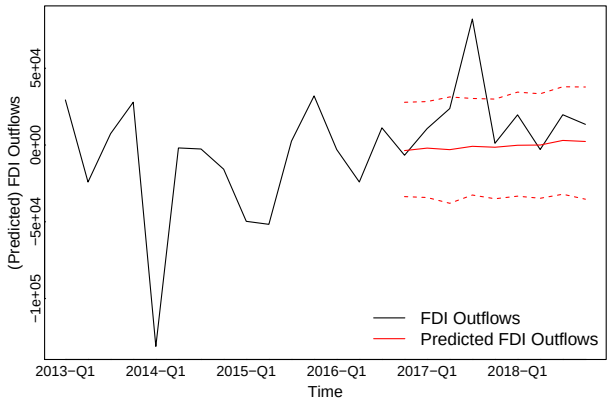
To test this proposition, I use all OECD countries to estimate a counter-factual UK, to demonstrate that FDI outflows increased following the vote. As controls I include GDP,

Figure 2: Placebo Results Across Countries in the Study



Note: The figure shows actual GDP before and after reunification in black across the countries in the study. The red solid line is the predicted point estimates of the post-treatment counter-factual. The dashed red lines represent the 0.95 posterior. As can be seen, with the exception of Norway, the model is quite accurate with its predictions.

Figure 3: UK FDI Outflows and the Predicted Outflows if the Brexit Vote did not Occur



Note: The figure shows FDI outflows in the UK in black. The red solid line is the predicted point estimates of the post-treatment counter-factual. The dashed red lines represent the 0.95 posterior. As can be seen, FDI outflows increased following the vote.

population, and unemployment. The results of the model are shown in Figure 3. It shows FDI outflows in the UK in black. The red solid line is the predicted point estimates of the post-treatment counter-factual. The dashed red lines represent the 0.95 posterior. As can be seen, FDI outflows increased shortly after the vote.

To again test the model's performance and determine the level of noise, I run placebo analyses on all remaining countries. Because there are so many countries, I do not show the results graphically. However, the 0.95 coverage is almost exactly 0.95 at a noise level of $\psi = 0.23$, suggesting the model is performing as expected.

5. GENERALIZING GPP

In this section, I generalize GPP in two important ways. First, I demonstrate how it can apply to multiple treated units, treated at different times. To aid in the intuition, I apply this model to test the impact of joining the European Union (EU) on financial stability. Second, I show how GPP can incorporate a link function for discrete outcomes of interest. Again, to aid in the understanding, I apply a negative-binomial link and analyze the effect of the United States' war in Afghanistan on the number of global terrorist attacks by Afghan

nationals.

5.1. *Allowing for Multiple Treated Units at Different Times*

GPP is essentially an imputation task taking into account temporal trends within and between units on variables of interest. It is therefore straight-forward to apply the model to multiple treated units, treated at perhaps different times. Without any change to the above model specification, I simply remove post-treatment variables across units that were treated, run the model, and compare the projected outcomes in the same manner as above. To motivate this potential, I apply the model to understand the impact of joining the EU on financial stability.

Applying GPP to Multiple Treated Units: Joining the EU and Financial Stability Joining the EU offers many potential benefits to member-states. One such benefit, as a result of opening financial and other economic markets, is financial stability. However, the transition arguably may decrease stability in the short-term, while increasing stability in the long-term. This expectation makes GPP a well-suited model to test both the short-term and long-term effects of joining the EU.

I analyze 61 countries over the time period 1996 — 2017. In this period, 10 countries in the data joined the EU at various times. Cyprus, Czechia, Hungary, Latvia, Lithuania, the Slovak Republic, and Slovenia joined in early 2004. Bulgaria and Romania joined in early 2007. Finally, Croatia joined in early 2013.

To test the proposition that financial stability will be harmed in the short-term after joining but improve in the long-term, I use data from, and theoretically motivated by, Bandaranayake (2019). Specifically, the outcome is captured by a country's banking z-score. I control for the Lerner index, a measure of banking competition, annual GDP growth, and real interest rate. These controls are all important determinants of a country's level of financial stability.

The results are shown in Figure 4. As can be clearly seen, there is insufficient evidence

that joining the EU had an impact on financial stability at conventional levels. There is not enough evidence to reject the null in any country at any time period. There is also no evidence of a homogeneous treatment effect. This goes against my expectation.

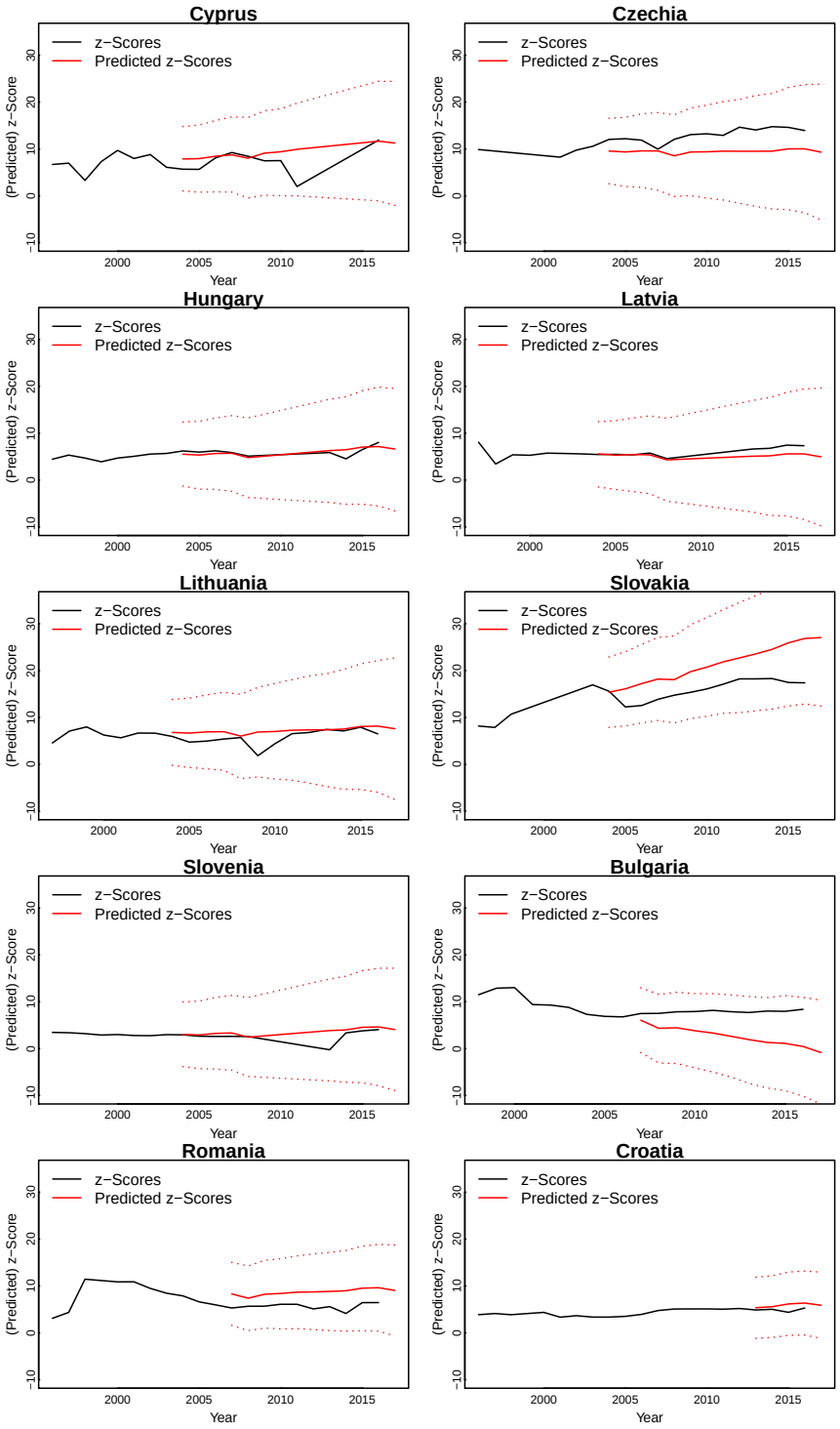
It is possible that there is something similar to an anticipatory effect. In anticipation of joining the EU, and in order to gain membership, the countries likely have to demonstrate that they are already financially stable. If this is the case, then the counter-factuals are not going to pick up any effect at the time of treatment. While my hypothesis was not supported by the evidence, this application still serves to demonstrate how the model can be applied to multiple treated units. Further, it demonstrates the importance of providing estimates of uncertainty in order to be able to determine the strength of the evidence.

While the (lack of) substantive conclusions of this exercise are likely unchanged by the nature of the placebo tests, it is important to note that the tests are not as straight-forward as the earlier applications. I chose to test one country at-a-time, starting from 2004 across all models, which led to a noise level of $\psi = 0.01$. I see no reason for this to be problematic, however conceptually this is somewhat trickier. There are multiple treated units, and they are treated at various times. It could be argued, then, that the placebo tests should reflect this multiplicity and various timings. This is an alternative approach that may lead to different noise levels to maintain appropriate coverage. This decision is ultimately up to the researcher, but may warrant further investigation. Having generalized the model for multiple treated units, I will now discuss how to generalize GPP to use link functions.

5.2. *Allowing a Link Function for Discrete Outcomes*

In political science research, we are frequently interested in discrete outcomes. To model such an outcome, we need to employ a link function, particularly when making predictions, such that the projections are in the support of the (potential) data. To demonstrate how to apply a link function, I generalize GPP to allow for a negative binomial outcome. I then apply it to global terrorist attacks by Afghan nationals following the invasion of Afghanistan

Figure 4: Banking Stability in Newly-Joined EU-States and Predicted Stability Had They Not Joined the EU



Note: The figure shows banking z-scores (financial stability) across the 10 countries in the data that recently joined the EU. The solid red lines represent the point estimates of the counter-factual had they not joined the EU. The dashed red lines represent the 0.95 posterior. As can be seen, there is insufficient evidence that there is a homogeneous treatment at conventional levels of reliability.

by the United States in late 2001.

Changes to the Specification The first parameter of the negative binomial model has essentially the same parameterization as the mean of the normal for the identity link. However, rather than add noise to the outcome for appropriate coverage, I add noise to this latent mean before putting it through the negative binomial link, such that the projected outcomes are discrete. I place a Cauchy prior on the reciprocal of the second parameter of the negative binomial distribution. Everything else in the specification remains the same.

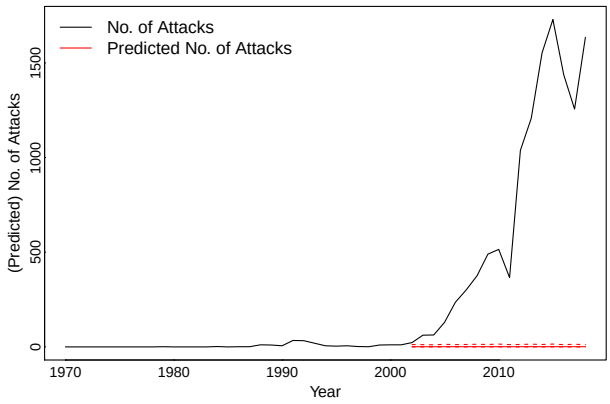
Applying GPP to a Negative Binomial Outcome: The Effect of the US Invasion of Afghanistan on Global Terror Attacks by Afghan Nationals The US waged a war in Afghanistan in late 2001 with the explicit intention of reducing global terrorism. However, I hypothesize that the war had the opposite effect. Specifically, I expect that terrorist attacks by Afghan nationals increased as a result of the conflict. War creates hostilities and opportunities for the spread of terrorism. I argue that this war in particular had and continues to have a large positive effect on the incidents of terror by Afghans.

To test this claim, I analyze incidents of terror from 1970 — 2018. These data come from the Global Terrorism Database (LaFree 2010). I include almost every nationality in the world, 188 specifically, dropping only a few small countries due to other data unavailability (e.g., Barbados). The unit of interest is therefore the number of attacks by a given nationality over time.

The first control is whether or not there is a war within the borders of a country, taken from the UCDP/PRIO Armed Conflict Dataset (Themnér and Wallensteen 2012). A nationality is coded as one if there is a conflict within the country's borders in a given year, zero otherwise. I do not impute this data, but instead instead Afghanistan is coded as one for all post-treatment (post-US-intervention) years. This allows me to isolate the impact of this particular war, which was aimed at reducing terrorism, instead of just any war within a country's borders.

I also control for unemployment and social spending on education, taken from the World

Figure 5: Number of Global Terror Attacks by Afghan Nationals and Predicted Number Had There Been No Conflict



Note: The figure shows the number of global terror attacks by Afghan nationals and the predicted number had the US not waged a war. The black line is the actual number of attacks. The solid red lines represent the point estimates of the counter-factual had the US not intervened. The dashed red lines represent the 0.95 posterior. As can be seen, there is a drastic surge of attacks, well above conventional levels of reliability, following the US invasion.

Development Indicators of the World Bank, and extracted using the `R` package `WDI`. There is substantial missingness in these data, but these data are easily imputed in the GPP framework. However, I do drop some small countries that have no data at all, as I mention above.

Finally, I control for ethnic fractionalization (Drazanova 2019). These data are only available through 2015, so imputation in the GPP is necessary here as well. It should be noted that this extensive imputation will necessarily increase uncertainty, making it more difficult to find an effect. That said, the imputation highlights the benefit of GPP of jointly modeling missing data and the counter-factual outcomes.

The results of the negative binomial GPP are shown in Figure 5. There is very strong evidence that the number of attacks by Afghan nationals increased dramatically following the US invasion, contrary to the war's intent. This is in line with my expectation.

The placebo tests called for noise at the level of $\psi = 0.01$ to be added to the latent mean. The coverage is relatively straight-forward to assess using this over-dispersed count outcome. The 0.95 interval can be used to determine how many cases fall within the interval.

However, note that this may become more difficult as the outcome becomes less spread, such as a binary outcome. This may be particularly problematic for rare events, such as war, in which always predicting a zero would achieve at least 0.95 coverage. This raises a challenge to the researcher, as testing for coverage of a latent variable or testing for coverage of the actual outcome may be impossible. My advice to the researcher is to add slightly more noise than theoretically justified to ensure the test does not result in a false positive.

6. CONCLUDING REMARKS

This paper develops a new method, Gaussian process projection (GPP), that estimates a counter-factual with uncertainty. This allows for traditional hypothesis testing of a treatment effect, and is therefore widely applicable to political science questions. Further, by projecting not just the outcome but jointly modeling the post-treatment covariates in a similar fashion, a degree of precision is gained while not controlling for post-treatment variables. The uncertainty in the projected covariates carries over to the uncertainty in the outcome through this joint modeling. This also naturally allows for imputation of missing data.

While the proposed model is applicable to a wide variety of questions, there are drawbacks. For example, testing the placebo coverage of the generalized GPP is not as straightforward, depending on the application. Second, the presentation does not make suggestions for continuous treatments. This is certainly a possible extension, but is left for future work. Third, the method is computationally expensive. While one run is generally quite fast, using two cores and taking, depending on the size of the data, anywhere from 15 minutes to a few hours, the problem lies in the placebo tests. The final application had an N of approximately 10,000, and took a few hours for each run. I used multiple cores for each nationality in the application such that the placebo tests were not prohibitively time-consuming, but on a standard laptop the placebo tests may be almost impossible. This is less of a concern as computers become increasingly powerful, but my suggestion would be to add more noise than theoretically justified to ensure false positives are less common.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2015. "Comparative politics and the synthetic control method". *American Journal of Political Science* 59 (2): 495–510.

Bandaranayake, Samangi. 2019. "How Does Bank Competition Affect Systemic Banking Crises?" *Colombo Business Journal* 10 (2).

Carlson, David. 2018. "Modeling Without Conditional Independence: Gaussian Process Regression for Time-Series Cross-Sectional Analyses".

Carpenter, Bob, et al. 2016. "Stan: A probabilistic programming language". In press, *Journal of Statistical Software*.

Gibson, NP, et al. 2012. "A Gaussian process framework for modelling instrumental systematics: application to transmission spectroscopy". *Monthly notices of the royal astronomical society* 419 (3): 2683–2694.

Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. "Long short-term memory". *Neural computation* 9 (8): 1735–1780.

Honaker, James, and Gary King. 2010. "What to do about missing values in time-series cross-section data". *American Journal of Political Science* 54 (2): 561–581.

Kirk, Paul DW, and Michael PH Stumpf. 2009. "Gaussian process regression bootstrapping: exploring the effects of uncertainty in time course data". *Bioinformatics* 25 (10): 1300–1306.

LaFree, G. 2010. "The global terrorism database: accomplishments and challenges. Perspect". *Terrorism* 4 (1): 2003–2016.

Monogan, James E, and Jeff Gill. 2016. "Measuring State and District Ideology with Spatial Realignment". *Political Science Research and Methods* 4 (01): 97–121.

Qian, P, Y Zhou, and C Rudin. 2011. "Using Gaussian processes to monitor diabetes development". In *Proceedings of the 6th INFORMS Workshop on Data Mining and Health Informatics (DM-HI 2011) P. Qian, Y. Zhou, C. Rudin, eds.*

Rasmussen, Carl Edward, and Christopher KI Williams. 2006. *Gaussian processes for machine learning*. Vol. 1. MIT press Cambridge.

Rosefielde, Steven. 2016. "Grexit and Brexit: Rational choice, compatibility, and coercive adaptation". *Acta Oeconomica* 66 (s1): 77–91.

Themnér, Lotta, and Peter Wallensteen. 2012. "UCDP/PRIO armed conflict dataset". *Journal of Peace Research* 49 (4): 565–75.