## **Definir el problema**

* **Impacto en consultas por lesiones: disminución por lesiones y se seleccionó un radio de 1 a 3 kilómetros a plaza italia.**
* **Impacto en admisiones de urgencia**
* **ACC pensaba en un ITS. Actualmente hay un NBM que en base a datos de 2015 a 2018 sobre consultas**
* **El problema es que no es evidente el quiebtre**

## **Definir el problema- Variables**

**## Dataset for 15-64 Age Group ##= data15a64 <- data[data$age == "15-64", ]**

**## Dataset for 65+ Age Group ##= data65a <- data[data$age == "65+", ]**

**## Pre-Exposure Dataset Excluding 2019= datanew[datanew$txtime == 0 & datanew$tx == 0, ]**

**## Pre-Exposure Dataset 2019= datanew[datanew$txtime == 0 & datanew$tx == 1, ]**

**## Post-Exposure Excluding 2019= datanew[datanew$txtime == 1 & datanew$tx == 0, ]**

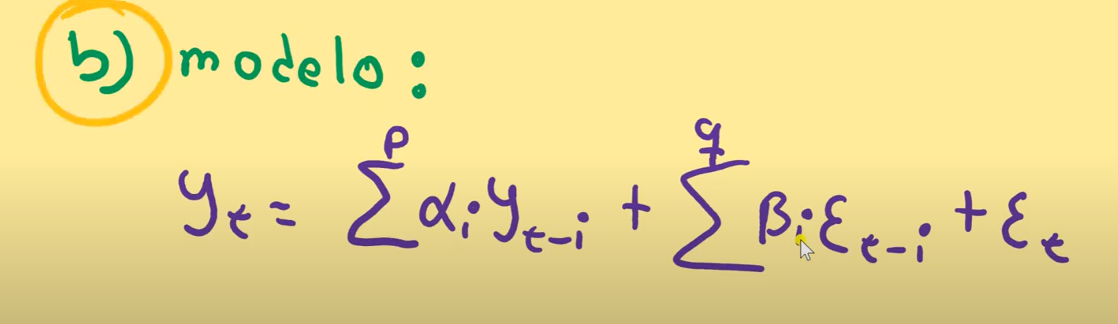
**## Post-Exposure 2019= datanew[datanew$txtime == 1 & datanew$tx == 1, ]**

**#### Model D with Negative Binomial Regression Modeling ####**

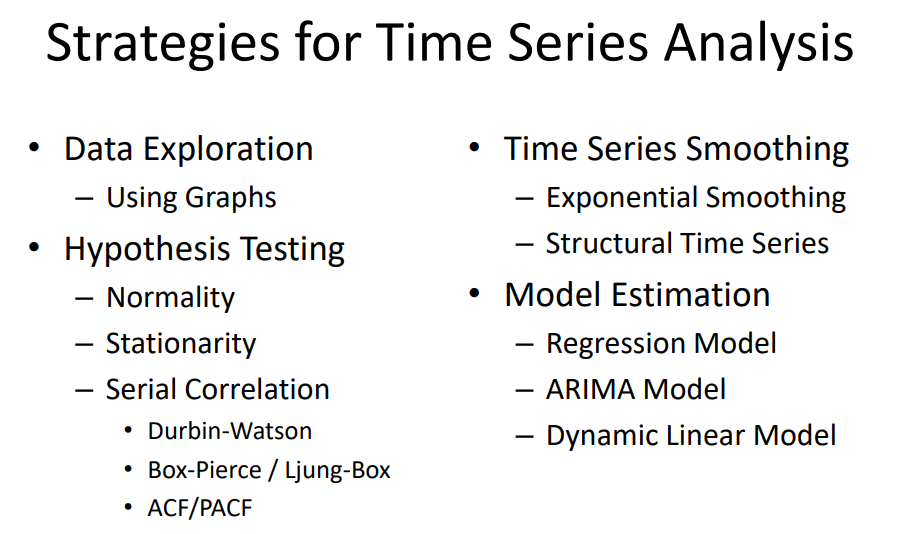
**modelD <-   
glm.nb(cons\_trauma ~ offset(log(offset)) + year + as.Date(date) + month + day + weekday + yearday + prevtrc + difftrc + hosp\_trauma,   
data15a64[data15a64$tx == 0 & data15a64$txtime == 1, ])**

## Series de Tiempo

* Proceso autorregresivo (AR)= cuando los valores de las series de tiempo pueden ser obtenidos usando los valores de la misma serie. Cuando tiene el valor de la serie rezagada. Constante\* su valor rezagado. Si alfa es 1, no es estacionario. Si es menor, es estacionario.
* ACF=
* PACF= residuos correlacionados, que quedan.
  + *If we see PACF plot there are many instances where correlation is above upper confidence band as PACF calculates correlations of lags of time series with residuals and our series itself is linear combination of residual and its lagged values.Hence we can get good correlation for near as well as past lags.*
* Moving Average (MA)= Un proceso en que el valor presente de una serie como una combinación lineal de errores pasados. Por el contrario, tendemos a asumir que los errores se encuentran distribuidos independientemente distribuido normal. SIEMPRE ES ESTACIONARIO. Se añade el error – error rezagado del anterior punto de tiempo.



* TODAS LAS SERIES QUE SON ESTACIONARIAS SE MODULAN A TRAVÉS DE UN ARMA
* # Autocorrelación= p d q para modelos ARIMA
* # Análisis espectral= comportamiento cíclico, separando componentes ruidosos
* # Estimación y descomposición= ajuste estacional, serie de c componentes con características
* #p= order of the autoregressive part;
* #d= degree of first differencing involved;
* #q= order of the moving average part.
* Estacionalidad en media: si no varía en el tiempo
* Estacionalidad en tendencia: si va creciendo en el tiempo
* No estacionaria: El valor de Yt= Yt-1 + Et(viene explicado por el valor pasado y residuo)
* Son equivalentes a los shocks(errores)
* Los no-estacionarios son resultado de una serie de acumulación de errores. Los shocks tienen memoria infinita y todos los errores pesan ene el tiempo. Si pz1 se puede ir extinguiendo. Los shocks en una estacionaria tienen memoria a corto plazo, por lo que los errores pasados tienen menor importancia. Son shocks aleatorios, los valores tienden a volver a la media.
* AR1, modelo autorregresivo de orden 1 o cadena de Markov. Depende de ella en el modelo anterior. Causal univariante. Rompe la hipótesis de no-autocorrelación, por sesgo de var. omitidas. Es habitual en las series. Servirá para predecir eso sí. AR1 es el más sencillo. Se recomienda trabajar con términos bajos (AR1, 2 o r), aunque hay un AR12, si hay relación con el pasado.
* Nuestros errores están correlacionados en una serie (AUTOCORRELACIÖN)
* Cambios imprevistos (shocks) que se encuentran condicionados
* El error anterior afecta el error siguiente
* Estacionaria: la meda y la desviación estándar son constantes y no hay estacionalidad (seasonality).
* Según la gráfica tenemos cuatro subplots:
* Observed, la serie original.
* Trend, los datos menos la estacionalidad.
* Seasonal, los datos con la estacionalidad (nuestro ejemplo semanal).
* Random, una serialidad temporal de manera aleatoria, muy útil para detectar anomalías.
* <https://www.paradigmadigital.com/dev/analitica-web-r-estadisticas-descriptivas-predictivas/>
* **ECONOMIA CON MANZANITAS:**  
  Proceso autorregresivo(AR)= cuando los valores de las series de tiempo puede ser obtenido usando los valores de la misma serie.
* Media móvil/Moving average(MA)= un proceso en que el valor presente de una serie como una combinación lineal de errores pasados. Por el contrario, tendemos a asumir que los errores se encuentran distribuidos independientemente distribuido normal.
* AF, ACF define el p= si decae lentamente es un modelo AR. Si tiene las primeras 2 barras sig. de ACF, es un AR(2), si tiene una es un AR(1).
* MA, PACF define el q= Es el mismo pero al revés. Tengo que ver el PACF sea decayente. El 1 o 2 o 3 se determina por valores siguientes en ACF.
* Tipos de estacionalidad:
  + En media: si no aría en el tiempo
  + En tendencia: Si va crecientdo en el tiempo
  + No estacionaria: El valor de Yt= Yt-1+Et (viene explicado por el de pasado + residuo)
  + Son equivalentes a los shocks (errores)
  + Los no-estacionarios son resultado de una serie de acumulación de errores hasta t. Los shocks tienen memoria infinita y todos los errores pesan en el tiempo.
  + Si p<1 se puede ir extinguiendo (rho). Los shocks en una estacionaria tienen memoria a corto plazo, por lo que los errores pasados tienen menor importancia. Sn shocks aleatorios, los valores tienden a volver a la media.
* AR(p)=1
  + Modelo autorregresivo de orden 1 o cadena de Markov. Depende de ella en el modelo anterior, es causal univariante. Rompe la hipótesis de no-autocorrelación, por sesgo de variables omitidas. Es habitual en las series. Servirá para predecir eso si. AR 1 es el más sencillo.
  + Se recomienda trabajar con términos bajos (AR1,2 o 3), aunque hay un AR12, si hay relación con el año pasado en una muestra mensual
* Nuestros errores están correlacionados en una serie (AUTOCORRELACION).
* Cambios imprevisor (shocks) que se encuentran condicionados
* El error anterior afect el error siguiente
* Estacionaria: la media y la desviación estándar son constantes y no hay estacionaridad (seasonallity)
* Series que no tienen tendencia pueden no ser estacionarias.
* Cuando no sólo los primeros rezagos, sino muchos van decreciendo lentamente en el ACF, puede significar que la serie es no-estacionaria
* Cuando caer más rápido el correlograma, puede ser estacionaria



* <http://web.pdx.edu/~crkl/WISE2017/TSAR-1.pdf>

## Modelo Teórico BSTS

The fundamental difference between a p-value and a posterior probability is that a p-value is a statement about the probability of observing data, while a posterior probability is a statement about the degree of belief of a particular parameter.

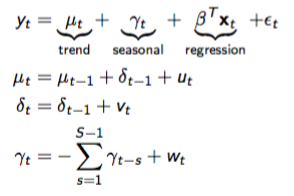
the negative binomial could be a more appropriate model as the variance of the negative binomial model is larger than the mean

1. Unconfoundedness: The Assignment is free from dependence on the potential outcome. That is the web activity of a scientist in 2016 will not be dependent on how many publications she/he is going to have in 2017.
2. Individualistic Assignment: The assignment mechanism is individualistic. The probability of sample unit i (i.e., a randomly chosen scientist) is a function of pre-treatment variables for the unit i only and free of dependence on the values of pre-treatment variables for other units. That is a web activity of a randomly chosen scientist is independent of the web activity of any other scientist of the community.
3. Probabilistic Assignment: The assignment mechanism is probabilistic so that the probability of receiving any level of the treatment is strictly between zero and one. In other words, the probability of the web activity of a scientist is purely random in nature.

* Natural experiments (and other quasi‐experimental designs) offer the opportunity to evaluate such interventions
* **compare temporal trends to their counterfactuals** which, because these are by definition unmeasurable, are estimated by using **synthetic control**
* The **synthetic controls are calculated using Bayesian structural time–series**, a methodology that was previously shown to be beneficial in the evaluations
* evaluate the impact of two licensing decisions anisd the introduction and subsequent defunding of new guideline documentation on health and crime in their immediate vicinity
* **Natural experiments** (and other **quasi‐experimental** designs) offer the opportunity to evaluate such interventions
* **Bayesian structural time–series models were implemented using the *bsts***[**18**](https://onlinelibrary.wiley.com/doi/full/10.1111/add.15002#add15002-bib-0018)**and CausalImpact**[**17**](https://onlinelibrary.wiley.com/doi/full/10.1111/add.15002#add15002-bib-0017)**packages in R**
* are state‐space models that can incorporate trends, seasonality and regression components. Two linked **equations describe (a) how the outcomes are related to an underlying latent state, and (b) how the latent state changes over time**. The errors of different state components are assumed to be normally distributed with a mean of zero and independent of each other. The **regression component of the model enables the inclusion of covariates**. To develop the synthetic controls, a spike‐and‐slab prior was placed on the regression coefficients, where the ‘**spike’ determines the probability that a covariate** (here, the time–series of the outcome in a control area) has a **non‐zero coefficient** based on independent Bernoulli distributions and the ‘slab’ is a weakly informative Gaussian prior with a large variance.
* We used Bayesian model **averaging across all combinations of predictors** to construct an outcome time–series from the covariate time–series
* the posterior distribution of the model parameters was estimated using a **sequence of Gibbs sampling steps from a Markov chain** with a stationary distribution
* The **95% Bayesian credible interval (CrI)** provides an estimate of the precision of average and cumulative effects over the forecasting time horizon and posterior tail area probabilities
  + differences in the data with respect to length of the time–series, number of control sets, temporal variability and other temporal patterns (such as seasonality). Each initial model was based on a random walk plus noise, or local‐level, model and sequentially other structures—autoregressive of order 1 [AR(1)] model, local‐linear time trends and additional seasonal components [**18**](https://onlinelibrary.wiley.com/doi/full/10.1111/add.15002#add15002-bib-0018) were assessed
  + **Especificación**: specifications: expected explained **variance was set to 90%**, and priors for the mean of each outcome and its standard deviation were defined as the **arithmetic mean of the outcome in the pre‐intervention period and 10%** of its standard deviation, respectively, and the **upper limit for the standard deviation was set to 200%** of that of the pre‐intervention time–series
  + **Model evaluation** was conducted graphically based on plots, and statistically using **root mean square errors for the pre‐intervention** (training) time‐period (partial) autocorrelation plots and Ljung–Box tests to assess **independent distribution of residuals** and precision of counterfactual estimates (standard deviation of forecasts). **Markov Chain Monte Carlo samples (10 000)** were used and convergence evaluated using Geweke and Raftery & Lewis diagnostics.
  + **Falsification tests** in which different implementation dates (**temporal falsification**) and replacement of case areas by control areas (**spatial falsification**) were also conducted to strengthen causal inference
* **RESULTADOS**: Evaluation of 12‐month follow‐up was modelled based on a **semi‐local linear trend model with two seasonal components (12 months and four seasons) and 4‐month follow‐up** (see below) based on a **local level with seasonal components**
* Time–series of **reported incidents of crime** (all offences) and antisocial behaviour specifically were available for 2015–17**. AR(1) models with linear local trend and monthly seasonal components were used to model the counterfactuals**
* Time–series of monthly reported incidents of drunk and disorderly behaviour, sexual offences, antisocial behaviour and domestic violence were available for the period 2008–14. **Different model specifications were required to best model the temporal patterns of each outcome: drunk and disorderly behaviour was modelled on an AR(1) autoregressive model with a monthly seasonal component, sexual offences on a local level model with a monthly seasonal component and antisocial behaviour and domestic violence both on a semi‐local linear time trend and a monthly seasonal component**. The results of the introduction of the LLG (3a) are shown in Table [**2**](https://onlinelibrary.wiley.com/doi/full/10.1111/add.15002#add15002-tbl-0002) and Fig. [**3**](https://onlinelibrary.wiley.com/doi/full/10.1111/add.15002#add15002-fig-0003), and suggest weak evidence for a reduction in the incidence of reported drunk and disorderly behaviour incidents [12‐month impact of −42% (95% CrI = –109%, +23%) compared to the counterfactual]. This, however, corresponds to fewer than one averted incident per month (95% CrI = –2, +0.3) on average. Falsification tests indicated that this weak evidence of effect was specific to the time‐point as well as to the area (with the exception of one neighbouring area, where a comparable effect was observed).

[**https://onlinelibrary.wiley.com/doi/full/10.1111/add.15002**](https://onlinelibrary.wiley.com/doi/full/10.1111/add.15002)

**de Vocht, F., McQuire, C., Brennan, A., Egan, M., Angus, C., Kaner, E., Beard, E., Brown, J., De Angelis, D., Carter, N., Murray, B., Dukes, R., Greenwood, E., Holden, S., Jago, R., and Hickman, M. (2020) Evaluating the causal impact of individual alcohol licensing decisions on local health and crime using natural experiments with synthetic controls. Addiction,** [**https://doi.org/10.1111/add.15002**](https://doi.org/10.1111/add.15002)**.**

* the number of hospitalizations for pneumonia was substantially lower than the counterfactual in the postvaccine period in Brazil [−22%; 95% credible interval (CI): −27%, −17%],
* To validate the synthetic control approach, we divided the prevaccine data from each country into different 48-mo training periods and 12-mo evaluation periods. The counterfactual estimates matched the observed
* Bayesian variable selection was used to weight the different control time series, effectively giving **more weight to those predictor variables that jointly explain the outcome variable best**. These weights are used to generate a **composite control variable** (the synthetic control), which **adjusts the counterfactual for changes in these other time series**.
* For the variable selection step, spike and slab priors were used with equal prior probability for inclusion for all covariates (18), and this probability was set so that the prior inclusion probability for each variable was π = 0.5.
* We used the bsts (30) and CausalImpact (18) packages in R for model fitting and formatting of output. The models ran for 10,000 Markov chain Monte Carlo (MCMC) iterations (burn-in of 1,000 iterations). Contrary to previous publications (18), we held the intercept and seasonality parameters constant over time.
* <https://github.com/weinbergerlab/synthetic-control/blob/master/ShinySC_Local_Deploy/synthetic_control_plot.R>
* <https://github.com/weinbergerlab/synthetic-control/tree/master/main%20analysis%20components>
* 
* **Y es la serie que se va a modelar**
* **El término de tendencia captura la tendencia a una serie en una particular dirección en el tiempo.**
* **Términos estacionales (seasonal terms) captura la asociación entre eventos periódicos (calendario, lunes, veranos, etc.)**
* **Los regresores son otras series de tiempo que son predictores de la otra serie de interés**

[**https://www.pnas.org/content/pnas/early/2017/01/31/1612833114.full.pdf**](https://www.pnas.org/content/pnas/early/2017/01/31/1612833114.full.pdf)

[**https://www.pnas.org/content/pnas/suppl/2017/02/01/1612833114.DCSupplemental/pnas.201612833SI.pdf?targetid=nameddest%3DSTXT**](https://www.pnas.org/content/pnas/suppl/2017/02/01/1612833114.DCSupplemental/pnas.201612833SI.pdf?targetid=nameddest%3DSTXT)

* The local level term defines how the latent state evolves over time and is often referred to as the unobserved trend. This could, for example, represent an underlying growth in the brand value of a company or external factors that are hard to pinpoint, but it can also soak up short term fluctuations that should be controlled for with explicit terms.
* Bayesian nature of the model, we can shrink the elements of β to promote sparsity or specify outside priors for the means
* we use a Bayesian structural time series model (BSTS) to investigate effectiveness of Seattle's GSI. This method estimates the impact of an intervention or a policy on an outcome by comparing the average value of the outcome variable after the intervention with its estimated average value in a hypothetical scenario in which the intervention does not take place. The difference between these two values is the estimated effect of intervention or policy on the outcome.
* BSTS method applied in this paper is superior to other commonly used approaches due to its flexibility and also makes causal inferences posible
* We use the monthly time series data of several available water quality parameters between 2004 and 2017, as well as a set of nine control time series, to create a counterfactual scenario
* the causal impact of intervention is estimated using the observed time series data of the outcome, along with several control variables which are correlated with the outcome, but are not affected by the intervention. Since the relationship between the outcome and control variables doesn't change over time, these variables can be used to estimate a counterfactual state for the outcome variable in absence of intervention in the ``after” period.
* A Structural time-series model can be described by a pair of equations relating yt to a vector of latent state variables α
* Three sources of information are used in this model: the time series data of the response variable prior to the intervention, the behavior of control time series that are predictive of the response variable before the intervention, and the prior knowledge about model parameters in a Bayesian setting
* The model allows for seasonal components in the data. Introducing monthly seasonality in the model improves the fit of the model in the pre-period, so we assume 12 seasons (for 12 months in each year) and draw 2000 MCMC samples.
* The model chooses the most appropriate variables for each WQI using a spike and slab method. However, our choice of these independent variables is limited due to data availability. We need monthly data on these variables over the entire time period of the study. Some variables might affect the water quality, but their data is not available at the monthly level, such as atmospheric nitrogen deposition and fertilizer use. The model also performs better with higher number of control variables
* This method also provides **higher flexibility** compared to other time series analysis methods due to applying a Bayesian approach and the synthetic control method enables us to estimate causal impacts.

**Jalali, P., & Rabotyagov, S. (2020). Quantifying cumulative effectiveness of green stormwater infrastructure in improving water quality. Science of The Total Environment, 138953. doi:10.1016/j.scitotenv.2020.138953**

* Obtener un componente estacional mensual con datos diarios es un poco difícil de hacer, pero puede hacer una temporada de 52 semanas AddSeasonal(..., nseasons = 52, season.duration = 7).
* El seasonal.durationargumento le dice al modelo cuántos puntos de tiempo debería durar cada temporada. El nseasonsargumento le dice cuántas estaciones hay en un ciclo. El número total de puntos de tiempo en un ciclo es season.duration \* nseasons.
* La segunda sugerencia es que quizás desee pensar en un modelo diferente para la tendencia. El LocalLinearTrendmodelo es muy flexible, pero esta flexibilidad puede aparecer como una variación no deseada en los pronósticos a largo plazo. Hay algunos otros modelos de tendencia que contienen un poco más de structura. GeneralizedLocalLinearTrend (perdón por el nombre no descriptivo) supone que el componente "pendiente" de la tendencia es un proceso AR1 en lugar de una caminata aleatoria. Es mi opción predeterminada si quiero pronosticar en el futuro. La mayor parte de la variación de su serie temporal parece provenir de la estacionalidad, por lo que puede intentarlo AddLocalLevelo incluso en AddArlugar de hacerlo AddLocalLinearTrend.
* Finalmente, en general, si está obteniendo pronósticos extraños y desea averiguar qué parte del modelo tiene la culpa, intente plot(model, "components")ver el modelo descompuesto en las piezas individuales que ha solicitado.

<https://qastack.mx/stats/209426/predictions-from-bsts-model-in-r-are-failing-completely>

* The model used in Berge, Sinha, and Smolyansky (2016) was a **probit regression, with Bayesian model averaging used to determine which predictors should be included**. The response variable was the the presence or absence of a recession (as [determined by NBER](http://www.nber.org/cycles.html)), plotted in Figure 9. The BMA done by Berge, Sinha, and Smolyansky (2016) is essentially the same as fitting a **logistic regression under a spike-and-slab prior** with the prior inclusion probability of each predictor set to 1/21/2. That analysis can be run using the **BoomSpikeSlab R package (**Scott 2010), which is similar to bsts, but with **only a regression component and no time series**. The marginal posterior inclusion probabilities produced by BoomSpikeSlab are shown in Figure 10(a). They largely replicate the findings of Berge, Sinha, and Smolyansky (2016), up to minor Monte Carlo error.
* **Time varying effects are available for arbitrary regressions with small numbers of predictor variables through a call to AddDynamicRegression.**

<http://www.unofficialgoogledatascience.com/2017/07/fitting-bayesian-structural-time-series.html>

## Contrafactual

la especificación de una hipótesis nula y la construcción de un universo alternativo y artificial que trataría de simular la evolución del sistema de ser cierta. De que los efectos lleven a suponerla más o menos creíble (hasta el punto de aceptarla o rechazarla, en la jerga del p-valor) a que exista una relación causal entre los unos y lo otro media mundo y mitad. Como de costumbre.

potential comparison time series are combined into a composite and are used to generate a counterfactual estimate, which can be compared with the time series of interest after the intervention

## CausalImpact

* An R package for causal inference using Bayesian structural time-series models
* estimating the causal effect of a designed intervention on a time series
* response time series (e.g., clicks) and a set of control time series (e.g., clicks in non-affected markets or clicks on other sites), the package constructs a Bayesian structural time-series model. This model is then used to try and predict the counterfactual
* **Supuestos**:
  + we assume that there is a set control time series that were *themselves not affected by the intervention*
  + relationship between covariates and treated time series, as established during the pre-period, remains stable throughout the post-period (see model.args$dynamic.regression for a way of relaxing this assumption)
  + Predictability. CausalImpact assumes that it is possible to model the outcome time series of interest as a linear combination of the set of control time series that were entered. For example, in the case of a marketing study, we'd assume that sales in one country can be predicted from sales in other countries. This assumption can be assessed by checking how well the predicted counterfactual (dotted line) matches the observed outcome time series in the pre-period (i.e., before the intervention started).
  + Unaffectedness. CausalImpact assumes that we have access to a set of control time series that were themselves not affected by the intervention. If they were, we might falsely under- or overestimate the true effect. Or we might falsely conclude that there was an effect even though in reality there wasn't. For example, in a marketing study, we'd assume that advertising in one country had no spill-over effect on our set of control countries.
  + Stability. CausalImpact assumes that the relationship between covariates and treated time series, as established during the pre-period, would have remained stable throughout the post-period if the intervention had not taken place.

**https://www.raybeam.com/focus/inferring-the-effect-of-an-event-using-causal-inference**

* Instead of using the default model constructed by the CausalImpact package, we can use the bsts package to specify our own model. This provides the greatest degree of flexibility.
* We do this by removing a post period after each ad spot from the original search volume time series, fitting the remaining curve, and imputing the missing data. We also require confidence (or credible) interval estimates so that we can make statistical significance claims. These considerations bring us to the Bayesian Structural Time Series (BSTS) model [12] that has the ability to fit time series with missing data and generate credible intervals.
* BSTS models are state space models for time series estimated using Bayesian methods. They have two components: a time series component and a regression component. An observation yt at time t (in our case minute-by-minute search volume time series) is linked to the state space through the observation equation
* control series are selected using a spike-and-slab prior. Specifically, the “spike” is a point mass at zero that shrinks a subset of coefficients to zeros, and the “slab” is a weakly informative distribution on the complementary set of nonzero coefficients. We set the values of the search time series within the post periods to missing when fitting the BSTS models.
* We fit a local-level BSTS model with the Poisson observation equation (1) on the original series with post-ad periods removed, plus the above four daily time series as controls in the regression component. This results in a baseline (counterfactual) time series with the missing values in the post periods imputed. Then the lift for a time period (e.g. a post period) is estimated as the difference between the observed number of searches during the post period and the sum of the counterfactual over that period. Figure 5 demonstrates the estimated counterfactuals from BSTS with non-contaminated data, with shaded areas representing post periods

**Brodersen KH, Gallusser F, Koehler J, Remy N, Scott SL. Inferring causal impact using Bayesian structural time-series models. Annals of Applied Statistics, 2015, Vol. 9, No. 1, 247-274. http://research.google.com/pubs/pub41854.html**

* impact$series
* in the point.pred column. The counter-factual is the part of the point predictions that occur in the post.period portion of that column. impact$series provides the data for all three graphs in plot(impact).
* **What does Posterior tail-area probability mean in Causal Impact?.** In the example (to be run with example("CausalImpact") in R), the summary shows a positive effect and reports a tail-area probability of 0.001. This means: **if the intervention had no effect on the response variable, there would be a chance of only 0.1% to see a positive effect** at least as large as the one observed.

[**https://stats.stackexchange.com/questions/263763/what-does-posterior-tail-area-probability-mean-in-causal-impact**](https://stats.stackexchange.com/questions/263763/what-does-posterior-tail-area-probability-mean-in-causal-impact)

* In a Bayesian framework, a third source of information for inferring the counterfactual is the available prior knowledge about the model parameters, as elicited, for example, by previous studies
* spike-and-slab prior on the set of regression coefficients and by allowing the model to average over the set of controls
* Subtracting the predicted from the observed response during the post-intervention period gives a semiparametric Bayesian posterior distribution for the causal effect
* They are subject to multiple seasonal variations, and they are often confounded by the effects of unobserved variables and their interactions
* Thus, while a large control group may be available, the treatment group often consists of no more than one region or a few regions with considerable heterogeneity among them.
* When fit to serially correlated data, static models yield overoptimistic inferences with too narrow uncertainty intervals
* when DD analyses are based on time series, previous studies have imposed restrictions on the way in which a synthetic control is constructed from a set of predictor variables, which is something we wish to avoid
* use classical variable-selection methods (such as the Lasso) to find a sparse set of predictors. This approach, however, ignores posterior uncertainty about both which predictors to use and their coefficients.
* State-space models distinguish between a state equation that describes the transition of a set of latent variables from one time point to the next and an observation equation that specifies how a given system state translates into measurements
  + it allows us to flexibly accommodate different kinds of assumptions about the latent state and emission processes underlying the observed data, including local trends and seasonality
  + inferring the temporal evolution of counterfactual activity and incremental impact
  + regression component that precludes a rigid commitment to a particular set of controls by integrating out our posterior uncertainty about the influence of each predictor as well as our uncertainty about which predictors to include in the first place, which avoids overfitting
* Contemporaneous covariates with dynamic coefficients. An alternative to the above is a regression component with dynamic regression coefficients to account for time-varying relationships
* When the relationship between controls and treated unit has been stable in the past, static coefficients are an attractive option
* Another option is latent thresholding regression, where one uses a dynamic version of the spike-and-slab prior
* A spike-and-slab prior combines point mass at zero (the “spike”), for an unknown subset of zero coefficients, with a weakly informative distribution on the complementary set of nonzero coefficients (the “slab”)
* However, the propensity-score approach requires that exposure can be measured at the individual level, and it, too, does not guarantee valid inferences, for example, in the presence of a specific type of selection bias recently termed “activity bias”

[**https://storage.googleapis.com/pub-tools-public-publication-data/pdf/41854.pdf**](https://storage.googleapis.com/pub-tools-public-publication-data/pdf/41854.pdf)

**BAYESIAN CAUSAL IMPACT ANALYSIS**

* As the CausalImpact methodology is based on a Bayesian approach, the CI are also not necessarily symmetrical (as seen in the example above).
* I´d say that the change you have performed does not have a direct influence on the development of the data series in post-period. In other words, if the change did not take place the development would be the same or just very very similar
* The default model in CausalImpact assumes Gaussian iid observation noise and a Gaussian random walk on the latent local level. There are no explicit switching variables for level shifts or outliers. You could replace Gaussian observation noise by t-distributed noise, for example, if your data suggests heavy tails.

<https://stats.stackexchange.com/questions/157606/causalimpact-on-single-time-series>

* Instead of a comparing single points in pre/post treatment, CausalImpact generates predictions at each time interval in the post-treatment period
* Each state of the post-treatment time series is associated with a probability distribution, allowing for better inference around treatment effects
* Post-treatment predictions incorporate trend and seasonality effects. The regression component incorporates behavior of control time series, resulting in a reduction in exposure to outliers in the training period

[**https://blogs.oracle.com/datascience/evaluating-the-effects-of-la-city-road-projects-using-causalimpact-v2**](https://blogs.oracle.com/datascience/evaluating-the-effects-of-la-city-road-projects-using-causalimpact-v2)

**There is a big push against “controlling for everything”. In my previous article about Bayesian statistics, I explained how applied statistics is moving away from mindless procedures to critical thinking.**

* The counter-factual predictor has three components: i) local trend, which works by sampling a normal distribution for noise between time points, based on the variance in the pre-event time period; ii) (optional) seasonal trend, which applies a repeating bias that sums to zero over its time period; iii) control trend, which applies coefficients to a set of data vectors from external unrelated objects, and serves to account for global variance in the model
* Each experiment then returned a p-value indicating the likelihood that the cumulative difference between the observed vector and the prediction could have occurred.

<http://www0.cs.ucl.ac.uk/staff/W.Martin/pubs/Martin_ACM_SRC_cameraReady.pdf>

## Spike and slab regression

* variable selection that is novel to most economists is spike-and-slab regression, a Bayesian technique
* initially we might think that all variables have an equally likely chance of being in the regression
* the “spike” is the probability of a coefficient being non-zero; the “slab” is the (diffuse) prior describing the values that the coefficient can take on.
* We combine these two draws with the likelihood in the usual way which gives us a draw from posterior distribution on both probability of inclusion and the coefficients. We repeat this process thousands of times using a Markov Chain Monte Carlo (MCMC) technique which give us a table summarizing the posterior distribution for γ (indicating variable inclusion), β (the coefficients), and the associated prediction of y. We can summarize this table in a variety of ways. For example, we can compute the average value of γp which shows the posterior probability that the variable p is included in the regressions.
* The “BMA” and “spike-slab” columns are posterior probabilities of inclusion

**http://people.ischool.berkeley.edu/~hal/Papers/2013/ml.pdf**

**Big Data: New Tricks for Econometrics Hal R. Varian**

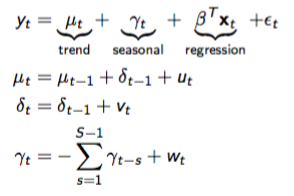
## BSTS

[**http://www.unofficialgoogledatascience.com/2017/07/fitting-bayesian-structural-time-series.html**](http://www.unofficialgoogledatascience.com/2017/07/fitting-bayesian-structural-time-series.html)

using a structural time series with three state components: a trend μtμt, a seasonal pattern τtτt and a regression component βTxtβTxt. The model is

yt=μt+τt+βTxt+ϵt

[**https://rstudio-pubs-static.s3.amazonaws.com/257314\_131e2c97e7e249448ca32e555c9247c6.html**](https://rstudio-pubs-static.s3.amazonaws.com/257314_131e2c97e7e249448ca32e555c9247c6.html)



* Y is the time series you are modeling (insurance claims)
* The trend term captures tendency of a time series to move in a particular direction over time.
* Seasonal terms capture association with periodic events (calendar seasons, holidays, e.g. Mondays)
* Regressors are other time series that are predictive of the time series of interest.

validation is more difficult for time series models than it is for classifiers and your audience may not be comfortable with the embedded uncertainty.

So, how does one navigate such treacherous waters? You need business acumen, luck, and *Bayesian structural time series models*. In my opinion, these models are more transparent than [ARIMA](https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average) – which still tends to be the go-to method. They also facilitate better handling of uncertainty

* Were compared with a counterfactual estimate of what those rates would have been absent reform. The counterfactual was estimated using a Bayesian structural time-series model based on mortality trends in similar states.
* Model -cheking: importante
* Cómo dividimos los conteos (¿promedio anual?)
* Aproximación analítica: Determinar cómo las tasas de mortalidad para distintas heridas hubiesen sido si no se hubiera adoptado el evento (en adelante, “intervención”)
* BSTS: Para estimar el comportamiento de chile en ausencia de la intervención. BSTS usan la flexibilidad del modelo bayesiano promediando para combinar el número de distintos modelos de series de tiempo en un solo forecast. En este análisis, promediamos 2 modelos simples para el comportamiento de chile en el periodo postintervención.
  + 1ero es un modelo estacional para estimar la tasa de mortalidad usando una variable dummy para modelos por cada 3 meses.
  + 2do es un “spike-and-stab” modelo de regresión lineal, en que la tasa de mortalidad en cada mes en el periodo de intervención fue regigresado en la tasa de mortalidad en los estados de comparación. El spike-and-stab usa “shrinkage” para ponderar covariabes que hacen menos para mejorar la precisión predictiva, reduciendo la varianza y mejorar la precisión. Se configuro con un valor 1. El a priori de 1 funcionó como forma de constreñir para reducir la varianza. Es una opción conservadora porque no presume que algún estado sea más o menos predcictivo, yy es el default para modelar en el paquete.
* Estimar el cambio postintervención, usando el BSTS (contrafactual). Para cada mes en el periodo postintervención, convirtió la tasa predictiva en conteos predichos y tomó la diferencia absoluta entre las muertes observadas y el conteo estimado de cambio de muertes que sigue.
  + Segundo, para cada mes del periodo de post-intervención, añadimos el conteo estimado mensual en el modelo, para sacar la diferencia absoluta y relativa (cambio porcentual) entre el valor estimado por el modelo y el observado. Este es un estimativo del cambio acumulativo, para ver hasta el mes en mortalidad que siguió la introducción de reformas.
* Presentación de resultados: Los resultados los presenta en forma acumulativa a la izquierda de la figura 2, y a la derecha el acumulativo.
* **Model-checking** e inferencia: Chequeamos los supuestos de que los cambios se atribuyen a la intervención y mal-especificaciones.
  + Primero, usamos una versión simplificada de la prueba propuesta por abadie en que se reemplaza el mismo análisis para cada uno de los estados comparados. Dado que los estados comparados no adoptaron intervenciones, las asociaciones debiesen ser mucho menores que las observadas en chile. EN NUESTRO ESTUDIO, SERÍA OCUPANDO OTROS AÑOS EN VEZ DEL 2019. Si las asociaciones estimadas fuesen comparables en magnitud, pudiese existir una mala especificación o por error aleatorio en el modelo que se eligió.
  + Segundo, repetimos el análisis completo para 2 causas de mortalidad que no debiesen estar afectadas porla reforma de Florida.
* Análisis de sensibilidad: DiD
* Credible Interval (CrI) 95%.
* Cuando algo no funciona se puede decir “fluctuated over the period” “essentially unchanged”

**Kenneth A Feder, Ramin Mojtabai, Elizabeth A Stuart, Rashelle Musci, Elizabeth J Letourneau, Florida’s Opioid Crackdown and Mortality From Drug Overdose, Motor Vehicle Crashes, and Suicide: A Bayesian Interrupted Time-Series Analysis, American Journal of Epidemiology, Volume 189, Issue 9, September 2020, Pages 885–893, https://doi.org/10.1093/aje/kwaa015**

**de Vocht, F., McQuire, C., Brennan, A., Egan, M., Angus, C., Kaner, E., Beard, E., Brown, J., De Angelis, D., Carter, N., Murray, B., Dukes, R., Greenwood, E., Holden, S., Jago, R., and Hickman, M. (2020) Evaluating the causal impact of individual alcohol licensing decisions on local health and crime using natural experiments with synthetic controls. *Addiction*,**[**https://doi.org/10.1111/add.15002**](https://doi.org/10.1111/add.15002)**.**

StanCon 2019

[**http://htmlpreview.github.io/?https://github.com/lauken13/Beginners\_Bayes\_Workshop/blob/master/stancon2019-intro.html**](http://htmlpreview.github.io/?https://github.com/lauken13/Beginners_Bayes_Workshop/blob/master/stancon2019-intro.html)

* recently developed a Bayesian method based on the structural time series model, and applied the method to nowcast unemployment initial claims with Google search data. Their method took the search data as regressors, and used a spike-and-slab prior for variable selection.
* When dealing with seasonality, most time series models rely on state space models, where the latent components capture the trend and seasonality

ny ARIMA model can be recast as a structural model.

Generally, we can write a Bayesian structural model like this:

Yt=μt+xtβ+St+et,et∼N(0,σ2e)

μt+1=μt+νt,νt∼N(0,σ2ν).

Here xt

* Trend and seasonality
* Forecast created by averaging across the MCMC draws
* Credible interval generated from the distribution of the MCMC draws
* Discarding the first MCMC iterations (burn-in)
* Using a log transformation to make the model multiplicative

### Run the bsts model

ss <- AddLocalLinearTrend(list(), y)

ss <- AddSeasonal(ss, y, nseasons = 12)

bsts.model <- bsts(y, state.specification = ss, niter = 500, ping=0, seed=2016)

Suggest the size of an MCMC burn in sample as a proportion of the total run.

### Predict

p <- predict.bsts(bsts.model, horizon = 12, burn = burn, quantiles = c(.025, .975))

### Actual versus predicted

d2 <- data.frame(

# fitted values and predictions

c(10^as.numeric(-colMeans(bsts.model$one.step.prediction.errors[-(1:burn),])+y),

10^as.numeric(p$mean)),

# actual data and dates

as.numeric(AirPassengers),

as.Date(time(AirPassengers)))

names(d2) <- c("Fitted", "Actual", "Date")

### MAPE (mean absolute percentage error)

MAPE <- filter(d2, year(Date)>1959) %>% summarise(MAPE=mean(abs(Actual-Fitted)/Actual))

### 95% forecast credible interval

posterior.interval <- cbind.data.frame(

10^as.numeric(p$interval[1,]),

10^as.numeric(p$interval[2,]),

subset(d2, year(Date)>1959)$Date)

names(posterior.interval) <- c("LL", "UL", "Date")

### Join intervals to the forecast

d3 <- left\_join(d2, posterior.interval, by="Date")

### Plot actual versus predicted with credible intervals for the holdout period

ggplot(data=d3, aes(x=Date)) +

geom\_line(aes(y=Actual, colour = "Actual"), size=1.2) +

geom\_line(aes(y=Fitted, colour = "Fitted"), size=1.2, linetype=2) +

theme\_bw() + theme(legend.title = element\_blank()) + ylab("") + xlab("") +

geom\_vline(xintercept=as.numeric(as.Date("1959-12-01")), linetype=2) +

geom\_ribbon(aes(ymin=LL, ymax=UL), fill="grey", alpha=0.5) +

ggtitle(paste0("BSTS -- Holdout MAPE = ", round(100\*MAPE,2), "%")) +

theme(axis.text.x=element\_text(angle = -90, hjust = 0))

predict.bsts function automatically supplies the upper and lower limits for a credible interval (95% in our case). We can also access the distribution for all MCMC draws by grabbing the distribution matrix (instead of interval). Each row in this matrix is one MCMC draw. Here’s an example of how to calculate percentiles from the posterior distribution:

credible.interval <- cbind.data.frame(

10^as.numeric(apply(p$distribution, 2,function(f){quantile(f,0.75)})),

10^as.numeric(apply(p$distribution, 2,function(f){median(f)})),

10^as.numeric(apply(p$distribution, 2,function(f){quantile(f,0.25)})),

subset(d3, year(Date)>1959)$Date)

names(credible.interval) <- c("p75", "Median", "p25", "Date")

Although the holdout MAPE (mean absolute percentage error) is larger than the ARIMA model for this specific dataset (and default settings), the bsts model does a great job of capturing the growth and seasonality of the air passengers time series. Moreover, one of the big advantages of the Bayesian structural model is that we can visualize the underlying components.

**Sorry ARIMA, but I’m Going Bayesian**

[**https://www.predictiveanalyticsworld.com/machinelearningtimes/sorry-arima-but-im-going-bayesian/7565/**](https://www.predictiveanalyticsworld.com/machinelearningtimes/sorry-arima-but-im-going-bayesian/7565/)

[**https://stats.stackexchange.com/questions/284204/q-r-causalimpact-order-of-control-time-series/285002#285002**](https://stats.stackexchange.com/questions/284204/q-r-causalimpact-order-of-control-time-series/285002#285002)

 My understanding is that the output provided are "credible limits", not "confidence intervals" in a frequentist sense, so I am not sure if there is a formula. Unfortunately, I am not clear as to whether the SD provided in the output is the standard error

[**https://stats.stackexchange.com/questions/309698/meta-analysis-on-effect-sizes-with-95-bayesian-ci-from-causalimpact-r-package?rq=1**](https://stats.stackexchange.com/questions/309698/meta-analysis-on-effect-sizes-with-95-bayesian-ci-from-causalimpact-r-package?rq=1)

[**Meta-Analysis on Effect Sizes with 95% Bayesian CI from CausalImpact R package**](https://stats.stackexchange.com/questions/309698/meta-analysis-on-effect-sizes-with-95-bayesian-ci-from-causalimpact-r-package)

#Posterior inference for bsts models.

[**https://github.com/google/CausalImpact/blob/master/R/impact\_inference.R**](https://github.com/google/CausalImpact/blob/master/R/impact_inference.R)

From the Bayesian structural time-series model, the difference between observed data and counterfactual predictions is the inferred causal impact of the intervention

The Bayesian structural time-series model is used to determine the extent to which the price change affects a consumer’s electricity consumption (i.e., the probability of the price having a causal effect on consumption) [33]. This causal analysis is used for comparison and validation of the results obtained from the Robin g-model. This package performs causal inference through counterfactual predictions using a Bayesian structural time-series model. The model requires a set of control time series (the time before the intervention—pre-period) and a set of response time series (the time from the beginning of intervention until it ends—post-period). With these two sets of time defined datasets, the package constructs a Bayesian structural time-series model and predicts how the response would have evolved if the intervention had never occurred (counterfactual prediction). As this is a non-experimental approach to estimate causal effects, conclusions require assumptions to facilitate accuracy. The model assumes that:

* There is a set control time series (pre-period) that is itself not affected by the intervention. If they were, it might falsely under or overestimate the true effect. Or it might falsely conclude that there was an effect even though in reality there was no effect.
* The relationship between covariates and the treated time series, as established during the pre-period, remains stable throughout the post-period.
* The relationship between the covariates (the time components) and the treated time series (electricity consumption) remains stable throughout the post-period.

This probability is obtained by calculating the posterior distribution of the response variable (Y—consumer electricity consumption) that would be expected in the absence of an intervention. The actual observed response is then compared to this posterior distribution. The tail-area probability is the probability under the calculated posterior that the response is at least as extreme (away from the expected value) as the observed one. The Bayesian structural time-series model is implemented using the CausalImpact R package

**On the Use of Causality Inference in Designing Tariffs to Implement More Effective Behavioral Demand Response Programs**

impact$summary$AbsEffect

To get the average effects using log transformed data:

**average effects using log transformed data** 🡪 mean(exp(impactLog$series$response[7:25,]) - exp(impactLog$series$point.pred[7:25,]))

**cumulative effect:** 🡪sum(exp(impactLog$series$response[7:25,]) - exp(impactLog$series$point.pred[7:25,]))

**For the credible interval for the absolute average effect**: 🡪

upper

mean(exp(impactLog$series$response[7:25,]) - exp(impactLog$series$point.pred.upper[7:25,]))

lower

mean(exp(impactLog$series$response[7:25,]) - exp(impactLog$series$point.pred.lower[7:25,]))

**summary(impact, "report")**

**How to impose restrictions on predictions made using a Bayesian structural time series model as implemented in the R package CausalImpact?**

[**https://stackoverflow.com/questions/30303680/how-to-impose-restrictions-on-predictions-made-using-a-bayesian-structural-time**](https://stackoverflow.com/questions/30303680/how-to-impose-restrictions-on-predictions-made-using-a-bayesian-structural-time)

**series <- ts(data, start=2012+342/365.25, frequency = 365.25/7)**

**kk <- 178**

**seas <- 365.25/7**

**st <- tsp(series)[1] + (1/seas)\*(kk-1)**

**training <- window(series, end = st)**

**testing <- window(series, start = st + 1/52.17857, end = st+14/52.17857)**

**train1 <- training[,"units"]**

**test1 <- testing[,"units"]**

##Bayesian Structural time series

ss <- AddLocalLinearTrend(list(), train1)

ss <- AddSeasonal(ss, train1, nseasons = 52, season.duration = 7)

model2 <- bsts(train1, state.specification = ss, niter = 500)

fbsts <- predict(model2, horizon = 14, burn = 100)

acc\_bsts <- accuracy(fbsts$mean,test1)

**Forecasting in R - ARIMA, TBATS, UCM, Bayesian Structural time series etc**

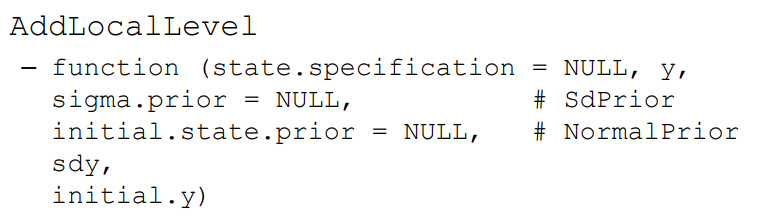
[**https://stackoverflow.com/questions/39237254/forecasting-in-r-arima-tbats-ucm-bayesian-structural-time-series-etc?rq=1**](https://stackoverflow.com/questions/39237254/forecasting-in-r-arima-tbats-ucm-bayesian-structural-time-series-etc?rq=1)

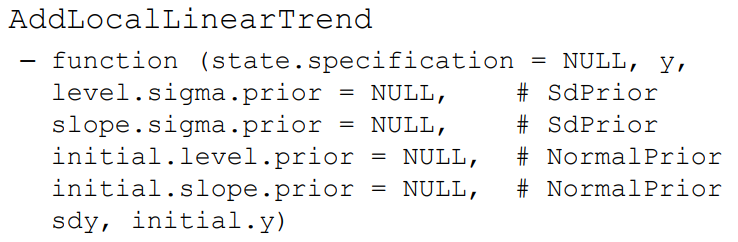
[Kalman filtering](https://en.wikipedia.org/wiki/Kalman_filter) and an [MCMC algorithm](https://en.wikipedia.org/wiki/Markov_chain_Monte_Carlo" \t "_blank) are used to fit the model. Forecasts are then calculated from the posterior predictive distribution.

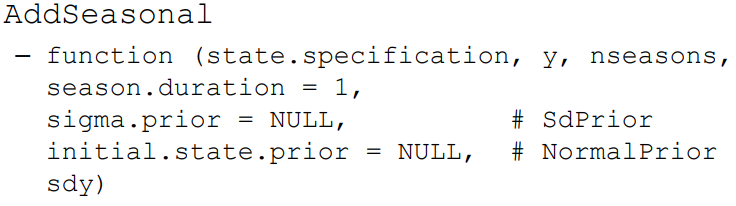
We didn't mention the question of stationarity, which is usually discussed when dealing with ARMA and ARIMA models. Stationarity is not discussed in decomposition-based methods, but it is dealt with implicitly by modeling the trend and seasonal components separately from the remainder series.

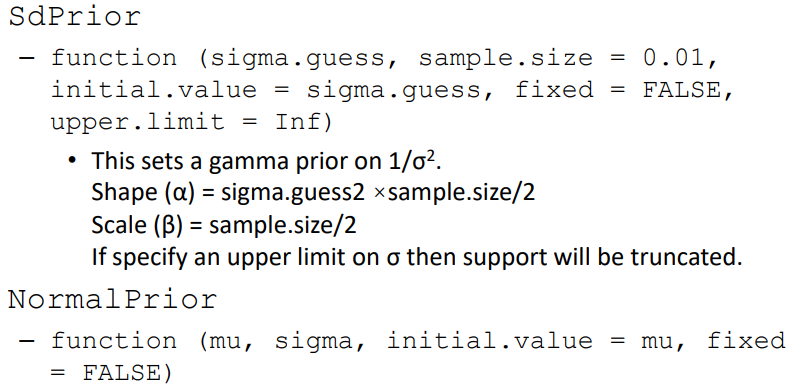
Let's fit a BSTS model on our data and then plot the components:

In the air passengers data set, there is only one yearly seasonal component. But several business time series can have multiple seasonalities. For example, a restaurant customer time series will have daily seasonality (with peaks at lunch time and dinner time) and weekly seasonality (less customers on weekends). STL cannot handle more than one seasonal component. Prophet can handle multiple seasonalities although, as mentioned above, it requires specific date formats. BSTS can handle multiple seasonalities by using a Fourier series (the same way as Prophet) for representing the seasonal component (call AddTrig instead of AddSeasonal).







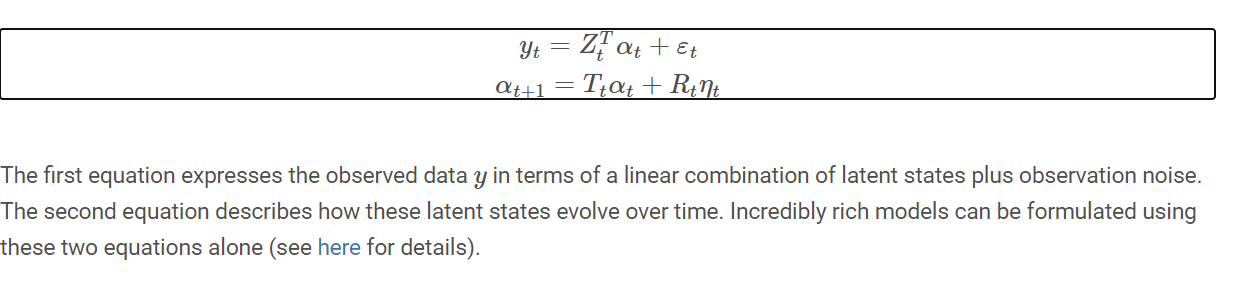


**http://web.pdx.edu/~crkl/WISE2017/TSAR-1.pdf**

fully Bayesian approach to inferring the temporal evolution of counterfactual activity and incremental impact. One advantage of this is the flexibility with which posterior inferences can be summarised. Third, we use a regression component that precludes a rigid commitment to a particular set of controls by integrating out our posterior uncertainty about the influence of each predictor as well as our uncertainty about which predictors to include in the first place, which avoids overfitting

**https://arxiv.org/pdf/1506.00356.pdf**

**INFERRING CAUSAL IMPACT USING BAYESIAN STRUCTURAL TIME-SERIES MODELS**



We estimated the impact of the strike using two separate approaches, interrupted [time series](https://www.sciencedirect.com/topics/medicine-and-dentistry/time-series-analysis) and Bayesian structural time series models

we used a recently developed Bayesian structural time series model implemented using the CausalImpact R package ([Brodersen et al., 2015](https://www.sciencedirect.com/science/article/pii/S221414051830553X" \l "bib2)). This model estimates a counterfactual using a Bayesian approach and accounts for autocorrelation inherent in time series data. Results are presented for the pooled data and stratified by membership type. We provide all of the data and analysis R code on GitHub

**Impact of a public transit strike on public bicycle share use: An interrupted time series natural experiment study**

[**https://www.sciencedirect.com/science/article/pii/S221414051830553X**](https://www.sciencedirect.com/science/article/pii/S221414051830553X)

|  |
| --- |
| annotate("rect", xmin = as.Date("2016-11-01", "%Y-%m-%d"), xmax = as.Date("2016-11-07", "%Y-%m-%d"), |
|  | ymin = 0, ymax = 20000, alpha = 0.3) + |

## Because 'y' is 0/1 and the state is on the logit scale the default prior

## assumed by AddLocalLevel won't work here, so we need to explicitly set the

## priors for the variance of the state innovation errors and the initial value

## of the state at time 0. The 'SdPrior' and 'NormalPrior' functions used to

## define these priors are part of the Boom package. See R help for

## documentation. Note the truncated support for the standard deviation of the

## random walk increments in the local level model.

ss <- AddLocalLevel(list(),

sigma.prior = SdPrior(sigma.guess = .1,

sample.size = 1,

upper.limit = 1),

initial.state.prior = NormalPrior(0, 5))

## Tell bsts that the observation equation should be a logistic regression by

## passing the 'family = "logit"' argument.

* According to Scott (2017), the forecast errors from a local linear trend model are wider than a semilocal linear trend model for long-term forecasting. He explained that the variance of a local linear trend model continuously grew with time and he built the hybrid model, the semilocal linear trend model, which replaces the random walk with a stationary AR process
* **Scott, S. L., (2017). Fitting Bayesian structural time series with the bsts R package. [Online] Available at: http://www.unofficialgoogledatascience.com/2017/07/fittingbayesian-structural-time-series.html [Accessed 14 8 2018].**

Deciding among the 3 trend components available in the library, in increasing order of complexity: local level, linear local, and semi-local linear. More complex components should perform better in a prediction with long horizon, but the cross-validation setup finds the most suitable

[**http://sisifospage.tech/docs/MineThatData\_ForecastChallenge\_whitepaper.pdf**](http://sisifospage.tech/docs/MineThatData_ForecastChallenge_whitepaper.pdf)

ARIMA (0,1,0), in example yt = yt-1 + ei is a unique case and known as the Random Walk model

he second suggestion is that you might want to think about a different model for trend. The LocalLinearTrend model is very flexible, but this flexibility can show up as undesired variance in long term forecasts. There are some other trend models that contain a bit more structure. GeneralizedLocalLinearTrend (sorry about the nondescriptive name) assumes the "slope" component of trend is an AR1 process instead of a random walk. It is my default option if I want to forecast far into the future. Most of your time series variation seems to come from seasonality, so you might try AddLocalLevel or even AddAr instead of AddLocalLinearTrend.

The local level model assumes the trend is a random walk (AddLocalLevel):

The local linear trend is a popular choice for modelling trends because it quickly adapts to local variation, which is desirable when making short-term predictions. However, this degree of flexibility may not be desired when making long-term predictions, as such predictions often come with implausibly wide uncertainty intervals

The semi-local linear trend is similar to the local linear trend, but more useful for long-term forecasting. A stationary AR(1) process is less variable than a random walk so it often gives more reasonable uncertainty estimates when making long term forecasts.

<http://oliviayu.github.io/post/2019-03-21-bsts/>

State-space models: modelling methodology in which the system is described as composed of a state vector and an observation vector, both time series. The relation between state and observation is described by the state-space model; the objective is to infer the properties of the state, which is hidden, from the observations available in the past. Forecasts are then produced from the estimated future states. Kalman filter: designed in the 60’s, and famously used in the Apollo 11 guidance system, is actually the first state-space model ever. That’s why some concepts are still called like Kalman did coin them: Kalman filter, Kalman gain.

**QUÉ HAGO PARA PREVENIR PREDECIR VALORES NEGATIVOS**

Thanks for this! Is there a way to prevent the forecast from forecasting negative values? I don't think it's possible for unemployment claims to be negative

sure, you can apply a Box-Cox Transformation, e.g. lambda = 0 (Log-Transform), on the time series before fitting the model. After which you need the apply the inverse Box-Cox transformation on predictions. In that way you ensure that the forecasts and confidence intervals are positive

S. Scott and H. Varian, “Predicting the present with bayesian structural time series,” International Journal of Mathematical Modeling and Numerical Optimization, June 2013. [Online]. Available: http://people.ischool.berkeley.edu/ hal/Papers/2013/pred-present-withbsts.pdf

his package uses spike and slab prior for the regression component of the models and Kalman filter for the time series component. The usual Bayesian dynamic linear models require variables to be Gaussian. The BSTS models have the capacity to incorporate the non-Gaussian variables. One additional advantage of the BSTS models is that we can visualize the trend, seasonal and regression components of the model [[20]](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7420989/#bib0020). In addition, the causal impacts of lifting/relaxing the lockdowns have been analyzed using intervention analysis under BSTS models. The numerical results for the causal impacts have been obtained using CausalImpact package in

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7420989/>

The author exploits the flexibility of Local Linear Trend, Seasonality, Contemporaneous covariates of dynamic coefficients in the Bayesian structural time series models. In addition, Causal Impact function in R programming is applied to analyze the model and read report of model. The results of the model show that the total confirmed cases who infected COVID-19 will be still most likely to increase straightly, the total numbers infected COVID-19 would be broken through 600,000 in the United States in near future (in the subsequent months). And then arrive at the peak around mid-May 2020. The cumulative prediction values are ± 1e+05 of cumulative trend. Also, the model suggests that the probability of variable Recovered cases daily is 0.07.

The author thinks that the Causal Impact function in R programming is the feature selection in BSTS model that may perform causal inference by counterfactual predictions. It has good characteristics such as trend, seasonal, regression, holiday. It is one of good default model [12]. It is supposed that the time series of the treated unit is explained with a set of covariates while do not allow themselves to be impacted by the intervention.

For most models of BSTS we know that “niter” is number of Markov Chain Monte Carlo (MCMC) sample to draw. The model can generate time series model for short- and long-term forecasting [13]. If there is higher number, it is more accurate inferences. “standardized. data” allows all columns of the data with moments estimated for the pre-intervention period prior to fitting the model to be standardized. It means that empirical Bayes accessing setting the priors so that the results will be linear transformations of the data. “priot.level.sd.” denotes in terms of data standard deviations that does have the Gaussian random walk of the local level. It can choose and let dataset with low residual volatility when regressing out known predictors such as total confirmed or fatality in this data. “nseasons” is period of the seasonal components. In general, it contains a seasonal component, set this to entire number larger than 1. For example, when there are daily observations, e.g. 1 for a day from a component of data, it interfaces currently only supports up to one seasonal component. In addition, it can let observations specify a number of seasonal components. Therefore, BSTS. model defaults to 1 that means no used seasonal component. “season. duration” is a kind of duration of each season. For example, if adding a day-of-week component to data with daily granularity, and offers the arguments model.args=list(nseasons=7. Season, duration=1), etc. “dynamic. Regression” includes the coefficients of time varying regression. It may combine local trend or local level that effects one of over specification.

<https://www.onomyscience.com/admin/uploads/20200910121447.pdf>

**Ejemplo BSTS Predict con un buen train y validation**

library(ggplot2, bsts, ggplotly, data.table)

# generate some data

set.seed(1)

n = 20

train\_size = 10

A = seq(1,n) + rnorm(n)

B = seq(1,n) + rnorm(n)

# this variable is not like the others

C = rnorm(n) + 5\*sin(seq(1,n))

D = seq(1,n) + rnorm(n)

X = data.table(A, B, C, D)

# transform the data for ggplot

long\_data = melt(X)

m[, t := seq\_len(.N), by = variable]

g1 = ggplot(data=m, aes(x=t, y=value, colour=variable)) + geom\_line() + labs(title="Evolution of Parameters over Time")

ggplotly(g1)

#break the data into training/testing data

train\_ind = seq(1,train\_size)

train\_X = X[train\_ind,]

test\_X = X[-train\_ind,]

ss <- AddLocalLinearTrend(list(), train\_X$D)

model4 <- bsts(D ~ .,

state.specification = ss,

niter = 1000,

data = train\_X,

expected.model.size = 3)

plot(model4, "components")

# observe that the model can tell that C isn't strongly related to D, but A and B are.

plot(model4, "coef")

pred4 <- predict(model4, newdata = test\_X, horizon = 24)

# plot predictions, vs actual (in red)

plot(pred4, ylim=c(0,50))

lines((max(train\_ind)+1):nrow(X), test\_X$D, col="red")

[**https://stats.stackexchange.com/questions/420186/incorporating-prior-information-into-time-series-prediction**](https://stats.stackexchange.com/questions/420186/incorporating-prior-information-into-time-series-prediction)

.**2 State Components** The following sections list the state components that are implemented in the BSTS-package and will be used for the empirical analysis of this thesis. The descriptions are based on the accompanying manual of the package.

2.2.1 Local Level

2.2.2 Local Linear Trend

2.2.3 Robust Local Linear Trend

2.2.4 Generalized Local Linear Trend

2.2.5 AR(p) State Component

2.2.6 Seasonal State Component

2.2.7 Static Regression

[**http://othes.univie.ac.at/36771/1/2015-03-23\_0808398.pdf**](http://othes.univie.ac.at/36771/1/2015-03-23_0808398.pdf)

Because of the limited length of each time–series, we used a static framework in which the regression coefficients were fixed and did not include an additional linear trend component.

[**https://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC6563459&blobtype=pdf**](https://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC6563459&blobtype=pdf)

t-distributed noise (if there is a high variability)

You could replace Gaussian observation noise by t-distributed noise, for example, if your data suggests heavy tails

—for example, assuming a random-walk, a semi-local linear trend or a local linear trend

**The posterior distribution of the time–series can then be projected forward, based on the known time–series data and the projected time–series, using the regression component of the model.**

In relative terms, the response variable showed an increase of 6.70% [95% CI: 4.10, 9.47]. The probability of obtaining this effect by chance was very small (Bayesian one‐sided tail‐area probability p ¼ <0.0001)

these results confirm that the 95% CI of the intervention impact is different from zero for the both period from January 2016 to June 2018 and from July to December 2018.

[**https://jaumepuigjunoy.cat/wp-content/uploads/2020/10/HEC4161.pdf#page=9&zoom=100,0,0**](https://jaumepuigjunoy.cat/wp-content/uploads/2020/10/HEC4161.pdf#page=9&zoom=100,0,0)

Advantages:

* Within a BSTS framework, one is able to turn all of the knobs and levers, if you will. By literally write down your model explicitly, you can insert any/all of the moving parts that you would like under the hood (suitable priors, seasonal/trend terms, etc.). This offers an added layer of interpretability often missing from more traditional approaches.
* Tying into the last point, two particular strengths unique to BSTS are (1) the control that one has over uncertainty and (2) the ability to incorporate feature selection via spike-slab priors. For anomaly detection, (1) is particularly relevant, as what is classified as an anomaly is often dependent upon how far predicted values stray from the observed values.
* All of the components of the underlying model are modeled simultaneously, and the user has the ability to explore these components independent of each other by appropriately marginalizing over the posterior distribution.

Disadvantages:

* Arguably, the only downside to this approach is the bit of extra thought/effort that crafting an appropriate model takes. Other popular approaches can be “blackboxed” and used to generate predictions fairly quickly; this, however, requires one to (1) specify an appropriate model, (2) sample the posterior distribution, and (3) use the posterior samples to get what you need to tell a story.

[**https://medium.com/@chbonfield/2-000-words-on-time-series-forecasting-1839714dcce**](https://medium.com/@chbonfield/2-000-words-on-time-series-forecasting-1839714dcce)

You can also add a static intercept with this kind of forecast:

ss <- AddStaticIntercept(ss,pre.period.response)

The result of the call is the plot shown in Figure 6. The bottom panel shows the original series. The top panel shows the cumulative total of the mean absolute one step prediction errors for each model. The final time point in the top plot is proportional to the mean absolute prediction error for each model, but plotting the errors as a cumulative total lets you see particular spots where each model encountered trouble, rather than just giving a single number describing each model’s predictive accuracy. Figure 6 shows that the Google data help explain the large spike near 2009, where model 1 accumulates errors at an accelerated rate, but models 2 and 3 continue accumulating errors at about the same rate they had been before. The fact that the lines for models 2 and 3 overlap in Figure 6 means that the additional predictors allowed by the relaxed prior used to fit model 3 do not yield additional predictive accuracy.

[**http://www.unofficialgoogledatascience.com/2017/07/fitting-bayesian-structural-time-series.html**](http://www.unofficialgoogledatascience.com/2017/07/fitting-bayesian-structural-time-series.html)

The one step prediction errors are a helpful diagnostic for evaluating a number of bsts fashions which have been match to the identical knowledge. They’re used to implement the operate CompareBstsModels, which is named as proven beneath.

The results of the decision is the plot proven in Determine 6. The underside panel reveals the unique collection. The highest panel reveals the cumulative complete of the imply absolute one step prediction errors for every mannequin. The ultimate time level within the prime plot is proportional to the imply absolute prediction error for every mannequin, however plotting the errors as a cumulative complete permits you to see explicit spots the place every mannequin encountered hassle, slightly than simply giving a single quantity describing every mannequin’s predictive accuracy. Determine 6 reveals that the Google knowledge assist clarify the massive spike close to 2009, the place mannequin 1 accumulates errors at an accelerated fee, however fashions 2 and three proceed accumulating errors at about the identical fee that they had been earlier than. The truth that the traces for fashions 2 and three overlap in Determine 6 implies that the extra predictors allowed by the relaxed prior used to suit mannequin Three don’t yield further predictive accuracy.

**https://openbootcamps.com/fitting-bayesian-structural-time-series-with-the-bsts-r-package/**

**Como puede ver, los modelos que contienen una parte de regresión muestran errores de previsión de un solo paso más pequeños que el primer modelo, que se refiere únicamente a un componente de tendencia y estacional. Del modelo dos al tres, también habrá una mejora en la adaptación a la serie temporal real a partir de 2012. Se reconoce 81 cf. Stephens-Davidowitz S.S. Varian H.: S.1ff Forecastmethods 57 muy bien que la diferencia de nivel, que surge de 2012, se puede modelar con mayor precisión con el aumento del tamaño del modelo. La hora de finalización en la Figura 19 es proporcional al error de predicción absoluto y medio para cada modelo.**

outliers  When in doubt, a safer option is to use **0.1**, as validated on synthetic data, although this may sometimes give rise to unrealistically wide prediction intervals.

family = "student" argument to the bsts function call. Allowing for heavy tailed errors in the observation equation makes the model robust against individual outliers, while heavy tails in the state model provides robustness against sudden persistent shifts in level or slope. This can lead to tighter prediction limits than Gaussian models when modeling data that have been polluted by outliers. The observation equation can also be set to a Poisson model for small count data if desired.

We estimated the difference in post treatment and pre-treatment periods while controlling for covariates and a lagged dependent variable through a Type-2 Sum Squares ANCOVA Lagged Dependent Variable model. The model also accounts for baseline levels and trends present in the data, allowing us to attribute significant changes to the interruption. The F-statistics were calculated using a bootstrap model of 5,000 replications.