Article Title

Name1 Surname[1,☯], Name2 Surname[2,☯], Name3 Surname[2,¤a], Name4 Surname[2,‡], Name5 Surname[2,‡], Name6 Surname[2], Name7 Surname[3,*], with the Lorem Ipsum Consortium[¶]

1 Affiliation Dept/Program/Center, Institution Name, City, State, Country
2 Affiliation Dept/Program/Center, Institution Name, City, State, Country
3 Affiliation Dept/Program/Center, Institution Name, City, State, Country

☯These authors contributed equally to this work.
‡These authors also contributed equally to this work.
¤a Insert current address of first author with an address update
†Deceased
¶Membership list can be found in the Acknowledgments section.
*** CorrespondingAuthor@institute.edu**

Abstract

The irreproducibility crisis in psychology is in part an interpretability crisis, as conclusions are drawn from inferences that ignore the limitations of the methodology. heteroscedasticity [1]

Author Summary

We take a critical look to these assumptions

Introduction

fMRI is one of the most common approaches to study brain function, in particular human cognition. Despite its pervasiveness and the remarkable scientific output a complete understanding of the link between the haemodynamic response and the neural underpinnings is missing. A fully satisfactory interpretation of the BOLD response has not been achieved yet, and conflicting candidates for the neurophysiological mechanisms that trigger the signal have been postulated. For example, in[?] it is found that excitatory neurons trigger vascular changes that are picked up by the BOLD signal. However, other studies show that the haemodynamic response is also present in the absence of neural firing[?] ,[?] . It is possible to bring some clarity here by reckoning that the haemodynamic response is primarily driven by changes in metabolic demands which are at its turn induced by neural activity of different sorts, including both excitatory and inhibitory inputs, neural spiking and neuromodulation[?] .

We might, however, want to distinguish between limitations that come from the technology being used and that can be eventually overcome or progressively improved, for example, the magnetic field strength, scan time reductions, high-density electrode

[1]The word comes from the Greek verb skedanime, which means to disperse or scatter

array to simultaneously stimulate the brain while imaging, from the limitations inherent in the statistical model used. In a recent paper[?] Eklund and colleagues argue that statistical methods used in fMRI, in particular, parametric cluster-wise inference, have been wrongly identifying regions of brain activity. For the last fifteen years, when the most common software tools (AFSL, FSL and SPM) were available, reports on brain activity across 400,000 published studies could be in reality the result of a careless treatment of false-positive rates.

At the core of the criticism elicited by the Eklund and colleagues is the poor understanding of physiological noise and more importantly on the assumptions made in theoretical framework used in the study of fMRI timer-series. The time series analysis of fMRI relies on a general linear model (GLM) approach, which as any other inferential model requires that a set of assumption about the data hold true. The goal of this paper is to review some of these assumptions, not as mere reminder of the obvious point that for a statistical model to be valid its assumptions need to be met but rather to show that resting state fMRI can benefit enormously if we study the stochastic properties of brain time series on their own. In particular we pay attention to the homoscedasticity assumption which despite the major consequences that has in the interpretability of fMRI results is poorly understood and very rarely investigated[?] .

GLM

The General Linear Model (GLM) is a massive-univariate approach that constructs a regression model to capture how the BOLD signal in a voxel should vary in response to a stimulus. GLM fits the model l independently to the time-course of each voxel. In the GLM applied to fMRI data, the BOLD response, y, is sampled n times (volumes), the intensity of the BOLD signal at each observation $y_i, i = 1..n$ is modeled as the sum of a number of known predictor variables $x_1, ..., x_p$ each scaled by a parameter β, $y_i = \sum_{j=1}^{p} \beta_j x_{i,j} + \epsilon_i$. A most compact notation of the GLM is $Y = \beta X + \epsilon$, where Y is a $n \times 1$ column vector representing the BOLD signal time series for one voxel, X is a $n \times p$ design matrix where each column is a predictor variable, β is a $p \times 1$ vector of unknown weights setting the magnitude between the predictor variable Y and ϵ is a $nx1$ vector containing the errors associated with each observation, not explained by the weighted sum of the predictor variables. The predictor variables or effects typically include regressors of interest related with the task and nuisance regressors. The regression problem of estimating the β parameters and therefore to assess whether a a predictor explains the variance observed in the voxel's uses ordinary squares methods, mainly OLS and GLS. The optimal parameters are those that minimize the residuals $\sum_{i=1}^{n}(Y_i - X_i\hat{\beta})$. The unknown parameter and its variance are directly estimated as: $\hat{\beta} = (X^TX)^{-1}X^TY$ and $var(\hat{\beta} = \sigma^2(X^TX)^{-1}$. According to the Markov-Gauss theorem, the OLS is the best linear unbiased estimator of the population parameters, provided the assumptions.

Heteroscedasticity

The standard GLM relies upon a set of assumptions: linearity, no regressor is a linear transformation of one or more regressors, the errors for different observations or time points are not correlated and follow a Gaussian distribution with 0 mean and constant variance or homoscedasticity. While some of the implicatiosn of these assumptions have been studied and ways to relax them have been proposed[?] ,[?] ,[?] , homocedasticity has been poorly studied. The homocedasticity assumption states that the variance of residuals is constant across observations (time points or volumes) and the covariances (off-diagonal elements in the variance-covariance matrix) are 0. If this assumption is not met, the parameter estimate $\hat{\beta}$ is still unbiased but not efficient (confidence intervals too long or too short), the variance covariance matrix is then the diagonal matrix scaled by the variance $I\sigma^2$. A time series that si not homoscedastic is heteroscedastic. Heteroscedasticity is time-varying variability or volatility in a time series.

Not coincidentally Eklund[?] has proposed to include heteroscedasticity in the vanilla GLM approach to fMRI data. In this model the noise variance is allowed to change over time, this allows to weight the scans based on some uncertainty measure, scans with large uncertainty compares to other scans are automatically remove when inferring connectivity or brain activity. The GLMH model proposed by Elklund shows that the homoscedastic model overestimates the constant variance term, the heteroscedastic model instead, as it models the variance increases is able to detect more brain activity. The GLMH can be used not only for brain activity estimation, but also for estimating functional connectivity, for example using covariates that affect the variance in the design matrix.

Heteroscedasticity has received very little attention in the fMRI literature, the few works that deal with this variance variability in the BOLD time series focus exclusively in head motion as the only relevant factor to explain variance of the errors in the regression model across observations (volumes)[?], [?], [?]. The GLMH arises out of the observation that the inclusion in the design matrix of head motion parameters and possibly their temporal derivatives remove motion related variance but not all of it. However, the heteroscedasticity found in the residual (after fitting the model with the design matrix) time series may not be entirely explained by motion correction. The idea that motion spikes is what makes the noise heteroscedastic is in need of a systematic reevaluation. A possible way to test this hypothesis is to study the BOLD time series in cadaveric brains, in case significant heterocedasticity were found the assumption that heterocesdaticity is uniquely correlated with head motion should be revised accordingly.

Financial time series in their level form are random walk, that is nonstationary, [2] on the other hand, on the first difference form they are generally stationary, I(1). Therefore, instead of modeling the levels of financial time series, why not model their first differences? But, these first differences often exhibit wide swings, or volatility, suggesting that the variance of financial time series varies over time. How can we model such "varying variance"? This is where the so-called autoregressive conditional heteroscedasticity (ARCH).

Heteroscedasticity, or unequal variance, may have an autoregressive structure in that heteroscedasticity observed over different periods may be autocorrelated.

Autocorrelation-stationarity

Several strategies for whitening the residuals such as autoregressive algorithms (AR) can effectively reduce non-whitened residuals. In[?] it is argued that the main source of autocorrelation in BOLD time-series could be model with an AR(2) model.

The autocorrelations considered as a function of lag k are called autocorrelation function (ACF). Thus, the ACF together with the process mean and variance characterize the stationary process that describes the evolution of y_t. The main tools for identification are the autocorrelation function (ACF) and the partial autocorrelation function (PACF). The autocorrelation (also know as serial correlation) function calculates the autocorrelation between a time series and the same time series lagged a certain period. The partial autocorrelation function measures correlation between time

[2] Why is a random walk not a stationary process? A random-walk series is, therefore, not weakly stationary, and we call it a unit-root nonstationary time series. For stationarity, the entire distribution has to be constant over time, not only its mean also the std deviation. To turn this into a stationary process, you would have to equally allow for all initial conditions which is impossible as there is no uniform distribution on the real numbers. If we treat the random walk model as a special AR(1) model, then the coefficient of $\alpha p_t - 1$ is 1 and therefore the moment depend on time t. If $|\alpha| < 1$, the moments do not depend on time and it is said to be weakly stationary.

series that are k time periods apart after controlling for correlations at intermediate lags. Thus, for a time series $\rho_t = \rho_1, \rho_2 .... \rho_t$, the autocorrelation function ACF(ρ_k) measures the correlation between the elements of original time series and the lagged time series $\rho_{t-k} = \rho_k, \rho_{k+1} ... \rho_{t-k}$ and the resulting elements of the lagged time series ρ_k. The PACF(ρ_k)) is the correlation between ρ_t and ρ_{t-k} after removing the effect of the intermediate $\rho_{t+1} ... \rho_{t-k-1}$. The correlgrams in autocorrelation in Figure can be inspected to see if there is a clear pattern in the residuals, the residuals are supposed to be white noise In other words, the correlograms of both autocorrelation and partial autocorrelation give the impression that the residuals estimated from the chosen ARIMA are purely random. Hence, there may not be any need to look for another ARIMA model. The lower the values of Akaike and Schwarz statistics, the better the model. From the estimated ACF and PACF correlograms in Figure we see any trend in the time series suggesting that the time series is stationary. The Dickey–Fuller unit root test is the analytical test to verify this point.

In the absence of stimuli, as it is the case in resting state time series, it makes more sense to follow the Box-Jenkins methodology for time series analysis. This methodology is standard in econometric analysis and is so far marginally employed in fMRI. In essence, the BJ methodology is an iterative method of four steps: identification (choosing tentative p,d,q), estimation (parameter estimation of the model), diagnostic checking (are the estimated residuals white noise, if so, do forecasting, if not go to identification) and forecasting.Once we have the model ARIMA(pd,q) and we have estimated the parameters we must ask if the model is a reasonable fir to the data, one simple diagnostic is to study the residuals up to some lag. We see that the estimated ACF and PACF up to lag 5 none of the autocorrelations (panel a) and partial autocorrelations (panel b) are individually statistically significant.

0.1 Multicollinearity

One assumption of CLRM is that there is no multicollinearity among the regressors included in the regression model. The degree of collinearity can be modeled using set theory. Why does the classical linear regression model assume that there is no multicollinearity among the X's? The reasoning is this: If multicollinearity is perfect, the regression coefficients of the X variables are indeterminate and their standard errors are infinite. If multicollinearity is less than perfect the regression coefficients, although determinate, possess large standard errors (in relation to the coefficients themselves), which means the coefficients cannot be estimated with great precision or accuracy. Multicollinearity, arises when the number of observations barely exceeds the number of parameters to be estimated. If the pair-wise or zero-order correlation coefficient between two regressors is high, say, in excess of 0.8, then multicollinearity is a serious problem. Multicollinearity is essentially a data deficiency problem There is collinearity when two or more predictor variables in a multiple regression model are highly correlated, meaning that one can be linearly predicted from the others with a substantial degree of accuracy. The regressors (features) should converge to 0 for big lags, that is, $a_i < a_j fori > j$.

Materials and Methods

In the absence of stimuli, as it is the case in resting state time series, it makes more sense to follow the Box-Jenkins methodology for time series analysis. This methodology is standard in econometric analysis and is so far marginally employed in fMRI. In essence, the BJ methodology is an iterative method of four steps: identification (choosing tentative p,d,q), estimation (parameter estimation of the model), diagnostic checking (are the estimated residuals white noise, if so, do forecasting, if not go to identification) and forecasting.

We study if there is a pattern in the residuals $y - \hat{y}$ ie. the observed minus the expected of the correlation. The time series deviates from pure noise if there is a correlation between one residual and the next. For example if the residual at time t was above expectation, then the residual at time $t + 1$ is likely to be above average as well $(e_t > 0, E(e_{t+1} > 0))$.

We perform a regression for each time series, in which the original time series is lagged from 1 to 5 seconds. Specifically, we fit the lagged model using OLS. If the OLS regression show correlation between the lagged values it indicates the existence of multicollinearity (see heatmap). Also for iid noise we would expect the coefficients β_i to gradually decline to zero (the further in time the less correlated), that is, $\beta_i > \beta_{i+1}$.

	ADF test
H0	A unit root is present in a time series sample
HA	Stationarity or trend-stationarity

Table 1. Augmented Dickey Fuller test. Null hypothesis H0: a unit root is present in a time series sample. The alternative hypothesis is usually stationarity or trend-stationarity..

Results

The assumption made thus far is that the prediction errors are independent random variables with a constant variance that is independent of the past. However, this assumption appears inconsistent with the heteroscedasticity often seen for time series in business and economics, in particular. For example, financial time series such as stock returns often exhibit periods when the volatility is high and periods when it is lower. This characteristic feature, or stylized fact, is commonly referred to as volatility clustering. The autoregressive conditional heteroscedastic (ARCH) model was introduced by Engle (1982) to describe time-varying variability in a series of inflation rates in Britain. An extension of this model called the generalized conditional heteroscedastic(GARCH) model was proposed by Bollerslev (1986). These models are capable of describing not only volatility clustering but also features such as heavy-tailed behavior that is common in many economic and financial time series. Still, there are other features related to volatility that are not captured by the basic ARCH and GARCH models.

Discussion

The objective of B–J [Box–Jenkins] is to identify and estimate a statistical model which can be interpreted as having generated the sample data. If this estimated model is then to be used for forecasting we must assume that the features of this model are constant through time, and particularly over future time periods. Thus the simple reason for requiring stationary data is that any model which is inferred from these data can itself be interpreted as stationary or stable, therefore providing a valid basis for forecasting.

Published research is based on implicit assumptions about the problem and data that are likely not correct and that produce estimates based on these assumptions.

Supporting Information

Acknowledgments