# An Introduction to Robust Inference for Sequential Data with the R Package `fixedCV`

Rebecca Kurtz-Garcia,...

## Table of contents

# 1 Introduction

```
setwd("~/Documents/GitHub/fixedCV")
```

The R package `fixedCV` provides a set of functions and critical values to conduct a robust inference procedure in the presence of dependent data sequences. Our emphasis in data sets with sequential dependent structure that occurs with time series data sets. When presented with a data set with an unknown sequential structure care must be taken when conducting an inference procedure, such as hypothesis testing or constructing confidence intervals. In particular is recommended to both use a robust estimator for calculating the standard error of the parameters of interest, and to use an alternative limiting distribution of pivotal quantities and test statistics. The `fixedCV` package provides support for both objectives.

The audience of the document is familiar with basic R syntax and functions, and also has some familiarity with linear regression and time series. Furthermore, the goal is to conduct inference procedures when the generating model is *unknown* but the data set is presumed to be stationary. If either of these parts is not met, then the reader may wish to pursue alternative methods.

For a careful exploration of the packages utlitiy please see sections 3-5, which use the simulated examples in section 2 to showcase various options. To see a start to finish brief example, please skip to section 6 which contains a non-simulated data set.

# 2 Examples

We start by simulating two data sets to help illustrate the ultility of the `fixedCV` package. The first data set is generated via an ARMA(2, 1) process, and the second is classical econometrics model.

## 2.1 ARMA(2, 1) Model

Let $\{e_t\}_{t=1}^T$ be a white noise process with mean 0 and variance $\sigma_e^2$, which we denote as $e_t \sim WN(0, \sigma_e^2)$ for short. Let $\phi_1, \phi_2, \theta_1$ all be fixed but unknown constants. We define the sequence of random variables $\{Y_t\}_{t=1}^T$,

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} = \theta_1 w_{t-1} + w_t.$$

Which can be equivalently written as

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \theta_1 e_{t-1} + e_t. \tag{1}$$

The process $\{Y_t\}_{t=1}^T$ is an ARMA(2, 1) model.

## 2.2 Regression Example

We again let $e_t \sim WN(0, \sigma_e^2)$, $w_{i,t} \sim WN(0, \sigma_w^2)$, and let $\theta_1, \theta_2$ be a set of two fixed unknown random variables. We define the process $\{X_{i,t}\}_{t=1}^T$ as

$$X_{i,1} = \theta_i X_{i,t-1} + w_{i,t}$$

for $i = (1, 2)$. Thus, each sequence $X_{1,t}$ and $X_{2,t}$ is an AR(1) model with autoregressive coefficients $\theta_1$ and $\theta_2$. Let $\beta_0, \beta_1$ and $\beta_2$ be fixed unknown values. We further define the process $\{Y_t\}_{t=1}^T$ as

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + e_t \tag{2}$$

# 3 Estimating Standard Error For Model Parameters

In each of the regression examples for 1 and 2 the solutions can easily be derived. Using maximum likelihood estimation methods the coefficients for our ARMA(2, 1) can be expressed as

## 3.1 ARMA(2, 1) Example

## 3.2 Regression Example

## 3.3 Long Run Variance (LRV) Estimators

In situations where the underlying correlation structure of the process is unknown but stationary a long run variance estimator may be used to estimate the correlation process. Consider in examples 1 and 2, except we are not confidence the model is correctly specified. In this case, a robust hypothesis testing procedure maybe considered.

In these procedures we rely on an estimate for the long run variance (LRV) which we denote $\Omega$. We consider the widely popular spectral variance (SV) estimator for $\Omega$ which has two tuning parameters: a kernel function $k(\cdot)$ and a bandwidth parameter $b \in [0, 1]$. Let $\hat{\Gamma}(h) = \frac{1}{T} \sum_{t=1}^{T-h} \hat{u}_t \hat{u}_{t-h}$ sample auto-covariances. Our SV estimator is defined as,

$$\hat{\Omega}_T = \sum_{h=-(T-1)}^{T-1} \kappa \left( \frac{h}{bT} \right) \hat{\Gamma}(h). \tag{3}$$

3

### 3.3.1 Mother Kernels

There are a large selection of choices for the kernel function. A classical set of kernels, hereby referred to as the *mother kenels*, must be piece wise continuous, symmetric about 0, and its Fourier transformation must be non-negative for all frequencies greater than 0. For more details on technical details see, CITE-FixedCV-paper, and more. The `fixedCV` package supports three papular kernels that meet these criteria: Bartlett, Parzen, Tukey-Hanning (TH), Quadratic Spectral (QS).

### 3.3.2 Lugsail Kernels

### 3.3.3 Bandwidth Selection

### 3.4 LRV Estimation with the `fixedCV` package

# 4 Critical Values for Robust Estimation

## 4.1 Fixed-$b$ Critical Values

When conducting a robust inference procedure it has been well documented that the while the standard $\chi^2$ limiting distribution is asymptotically applicable to our test statistics in REF???, the sample size must typically be exceptionally large for the test statistic calculated to be reasonably represented by it. Instead, an alternative limiting distribution introduced by REF??? was introduced to combat this problem. This alternative limiting distribution captures the variability of the kernel and bandwidth of the test statistic, as opposed to the $\chi^2$ limiting distribution which does not account for these aspects.

Unfortunately, the fixed-$b$ limiting distribution does not have a closed for expression that can be used to calculate the critical values analytically. Instead, practitioners have been left to approximate these critical values necessary for inference using various approximation methods. The three approximation methods are:

- **Simulation Approximation**: For each Test statistics generated
- Fitted Approximation
- Analytical Approximation

### 4.2 Obtaining Critical Values with the `fixedCV` package

The fixed-$b$ critical values can be obtained via the `get_cv` function as part of the `fixedCV` package using all three approximation methods. The function has

```r
lm(1:5~ sample(1:100, 5))
```

```
Call:
lm(formula = 1:5 ~ sample(1:100, 5))

Coefficients:
    (Intercept)  sample(1:100, 5)
        4.72580          -0.04741
```

### 4.3 Generating New Critical Values with the `fixedCV` package

## 5 The `robust_lm` function in the `fixedCV` package

### 5.1 Options for Critical Values

### 5.2 Options for Bandwidth Selection

### 5.3 Options for Kernel Selection

### 5.4 Confidence Interval Estimates

### 5.5 Graphical Diagnositics

## 6 An Example with Ocean Sea Levels