

# Weekly Progress: 30 June - 6 July 2014

July 4, 2014

All of the source code for this work can be found on GitHub at:

<http://github.com/cbuntain/ChangePointDetection>

## 1 Achievements

- **Found an existing Python library for modeling VAR time series data** – Python’s Statsmodels package ([statsmodels.sourceforge.net](http://statsmodels.sourceforge.net)) provides some nice VAR modeling implementations for maximum likelihood estimates of  $\Phi$  matrices. It does not currently support VARIMA models, however, so we are limited in that regard.
- **Implemented a Data Simulation Script** – Can simulate three-dimensional time series data with an arbitrary  $\Phi$  matrix, an arbitrary number of change points, and an arbitrary number of data points.
- **Implemented Two Versions of Galeano and Peña’s LRT Statistic** – The first implemented version of the LRT-based covariance change detection algorithm uses Galeano and Peña’s method for finding the maximum likelihood ratio statistic (which corresponds to a change point) in the given time series, recursively splits the series at the change point, and rather than computing  $W$ , it re-applies Statsmodels’ VAR model fitting functionality to estimate a new model on the smaller series. This operation continues until no more change points are detected. The second implementation follows the paper’s version more closely by estimating  $W$  when a change point occurs. Both implementations seem to perform consistently, by the paper’s version is faster in that it does not require a round of model fitting for every change point.
- **Implemented Two Versions of Galeano and Peña’s CUSUM Statistic** These two implementations of CUSUM were developed separately. One makes use of binary segmentation of the time series when a change point is found, and the other implementation follows the procedure provided in the paper (with several additions to obtain a working algorithm). As with the LRT implementations, the version following the paper demonstrate better accuracy in identifying change points in simulated data.
- **Authored a Preliminary Research Abstract** – This abstract sets forth, at a high level, our plans for the remainder of the project.

## 2 Plans for the Upcoming Week

- **Simulate Data for Analysis** – We need to simulate more data of varying dimension and underlying process type.
- **Evaluate Implementations on Other Model Types** – Since our current implementations of LRT and CUSUM statistics rely on a library that can only perform MLEs for VAR-based series, we need to figure out how well (if at all) this works for VARMA or VARIMA and models of different  $(p, q)$  orders.
- **Capture Data for Analysis** – We already have the bridge data, but we need to grab the Bitcoin market data and Twitter data.
- **Begin Implementing Online Algorithms** – We need to start implementing the SVM and Gaussian process algorithms.