Supplementary Material for "The Fractionally Cointegrated Vector Autoregression Model in R"

Lealand Morin
Department of Economics, University of Central Florida

Morten Ørregaard Nielsen
Department of Economics, Queen's University and CREATES

Michał Ksawery Popiel
Analysis Group

May 23, 2021

1 Review of Existing Software

An exhaustive listing of the R packages (R Core Team, 2017) available for time series analysis were compiled in Hyndman (2020). Among these, the packages most relevant to the FCVAR model are those related to cointegration within the family of vector error correction models or those related to long memory, often under the $ARFIMA$ framework.

In R, contegration analysis can be conducted using a variety of packages. One such package is **aTSA** for *Alternative Time Series Analysis* (Qiu, 2015). In this package, the `coint.test` function performs Engle-Granger tests, as in Engle and Granger (1987), for the null hypothesis that two or more time series, each of which is I(1), are not cointegrated. This package is designed to focus on a particular response variable, restricting attention to relationships with a one-dimensional cointegrating space. That is, this framework can detect a single equilibrium equation. Another package following the Engle and Granger (1987) approach is the **egcm** package in (Clegg, 2017). The **egcm** package restricts to a simplified form of cointegration. It is designed for bivariate analysis, with a concentration on applications to the prices of financial assets.

Other packages have implemeted the cointegration tests in Phillips and Ouliaris (1990). This amounts to running a regression of the response variable on a set of regressors and testing the residuals for a unit root following Phillips and Perron (1988). The `po.test` from the **tseries** package (Trapletti et al., 2019) implements this test, as well as the `ca.po` function in the **urca** package (Pfaff et al., 2016).

The **cointReg** package in Aschersleben and Wagner (2016) follows a different approach, using modified ordinary least squares (OLS) approaches to the analysis of cointegration. One such method is the fully modified OLS (FM-OLS) approach of Phillips and Hansen (1990) in the `cointRegFM` function. Another option is the dynamic OLS (D-OLS) approach (see Phillips and Loretan (1991), Saikkonen (1991) and Stock and Watson (1993)) implemented in `cointRegD`. It also implements, in `cointRegIM`, a variant called integrated modified OLS (IM-OLS) of Vogelsang and Wagner (2014), which is based on an augmented integration transformation of the regression model.

Following another approach, Johansen (1995) analyzes the cointegrated VAR model in a more holistic fashion. In this framework, the time series are treated as an endogenous system of equations and permits the estimation of a higher-dimensional cointegrating relationship, i.e. several equilibrium relationships. It also allows for the joint estimation of parameters relating to the system of equations, permitting likelihood ratio tests for a wide variety of hypotheses. The `VECM` function in the **tsDyn** package allows for the application of either the Engle and Granger (1987) or the Johansen (1995) MLE method. However, this package is designed primarily with nonlinear time series models in mind. The **urca** package (Pfaff et al., 2016), which is designed to perform unit root tests and cointegration analysis, follows the Johansen (1995) approach in the function `ca.jo`. It also provides options for testing restrictions on the parameters in the model. Of the packages designed for the CVAR model, this is perhaps the closest available to the **FCVAR** package, in terms of the testing opportunities available.

While the packages that follow the framework of Johansen (1995) are most closely related to the **FCVAR** package, these are not suited to the analysis of series with a fractional degree of integration. That is to say that these packages allow for only a discrete form of cointegration between the series. For example, the series are all integrated, i.e. $I(1)$, and the residuals from a regression are stationary and $I(0)$. The fractionally cointegrated VAR model allows for the possibility that variables can be integrated of order d and cointegrated of order $d - b$, where d and $b > 0$ can be real numbers. Analysis using the above packages typically involves a preliminary analysis of the form of non-stationarity of the variables, using a number of unit root tests, i.e. to test whether the series are $I(1)$. With fractionally integrated variables, the first stage of the analysis is to determine the order of fractional integration, i.e. the parameter d. This has been the focus of much of the available software to analyze series with the characteristics of so-called long memory.

In another section of the *CRAN Task View: Time Series Analysis* (Hyndman, 2020), several packages are listed for the estimation of models for series with features of fractional integration or long memory. A number of these pacakages are focused on estimation of $ARFIMA$ models, known as autoregressive fractionally integrated moving average models. The **fracdiff** package (Maechler et al., 2020) includes functions for fitting $ARFIMA(p, d, q)$ models, including the step of estimating the long memory parameter d. The namesake function `fracdiff` calculates the maximum likelihood estimators of the parameters of a fractionally-differenced $ARIMA(p, d, q)$ model. A few notable functions in this package estimate the long memory parameter d within this model[1]. The **arfima** package Veenstra and McLeod (2018) fits a wider variety of $ARFIMA$ models. Also, the **nsarfima** (Groebe, 2019) package provides methods for fitting and simulating non-stationary $ARFIMA$ models. This package is more innovative in terms of the types of optimization problems built on the $ARFIMA$ model, including both maximum likelihood (as in Beran (1995)) and minimum distance (as in Mayoral (2007)) estimators. Overall, $ARFIMA$ models treat the data by fractional differencing to transform data to a form suitable for an $ARMA$ model, similar to ordinary first differencing for variables that have unit roots as in $ARIMA$ models. This transformation precludes the use of models that study cointegration relationships.

The package **LongMemoryTS** is in a class of its own, in that it uses a wide variety of methods to investigate both fractional integration and cointegrat-

[1] Note that the `diffseries` function in **fracdiff** is based on the same algorithm in Jensen and Nielsen (2014) as `FracDiff` in **FCVAR**, except that `diffseries` demeans the data first. Specifically, `fracdiff::diffseries(x, d) - FCVAR::FracDiff(x - mean(x), d)` is numerically very small. The demeaning step is not required to estimate the FCVAR model, as the mean parameters are estimated jointly with the others while optimizing the likelihood function.

ing relationships.[2] For estimating the order of the fractional integration in a series, there are several options including the log-periodogram estimators of Geweke and Porter-Hudak (1983) and Robinson (1995b). Other options include the semiparametric local Whittle estimator of Robinson (1995a), the exact local Whittle estimator of Shimotsu and Phillips (2005), and a version for series with unknown mean and time trend in Shimotsu (2010). They also implement more recent approaches, such as the local polynomial Whittle plus noise estimator of Frederiksen et al. (2012) and the modified local Whittle estimator of Hou and Perron (2014).

For determining the cointegrating rank, i.e. the dimension of the cointegrating relationship, this package also provides several options. These inclide a semiparametric method in Chen and Hurvich (2003). Robinson and Yajima (2002) propose a model selection procedure to estimate the cointegrating rank , which includes a test for equality of all memory parameters simultaneously and is further explored in Nielsen and Shimotsu (2007). Following another approach, the package also provides a method of identifying cointegration by eigenanalysis (Zhang et al., 2018).

Finally, for estimation the cointegrating relationship itself, the **LongMemoryTS** package implements a number of approaches. It implements semiparametric approaches for estimating the cointegrating vector, including that of Robinson (1994) and later Robinson and Marinucci (2003) and Christensen and Nielsen (2006). They implement a semiparametric residual-based test for fractional cointegration from Chen and Hurvich (2006), and

[2] Morten: You would be best suited to comment on the references in this package. There are several citations to your papers and papers that I'm sure you know better and I want to make sure that we are honest about the difference between what we do and what they do. It seems to me that this package is a who's who of analyzing the cointegration of fractional systems, except for Johansen's MLE framework, which is what we implement. I like to think that our approach following Johansen (1995) is more holistic, in that all the parameters are estimated jointly, aside from the rank and lag selection. It captures all of the pieces in one maximum likelihood framework and this approach has the added benefit of allowing for a wide range of restrictions to test.

Before I forget, one important point to note is that I did not succeed in installing this package. It requires dependencies that would not install on either my local machine or the ones in Dunning 211 or the equivalent at UCF. I'm a patient guy but if I can't get it to work in 10 minutes, their user base is going to be very small. That's too bad, because they have also implemented Jensen and Nielsen (2014), with your permission, in the function `fdiff`. I would have liked to test it for myself. In my opinion, they have too much going on in one package and the level of complexity can lead to dependency problems like this. As a user, I would move on to the next package that works, which is a good reason to implement this function ourselves, without these problems. Actually, now that I have looked at it more closely, the documentation is incomplete as well. Our closest competitor is a work in progress – but there is no need to call them out on it.

other semiparametric tests in Marmol and Velasco (2004)and Wang et al. (2015) and a semiparametric Hausmann-type test for fractional cointegration by Robinson (2008). Also within the semiparametric family is the fully modified narrow band least squares (FMNBLS) approach in Nielsen and Frederiksen (2011). They also implement a nonparametric approach to test for fractional cointegration and rank estimation by Nielsen (2010). Following another framework, the frequency-domain test for fractional cointegration in Souza et al. (2018) is also available in this package.

In contrast, our approach is to use a fully parametric model in the maximum likelihood framework. The **FCVAR** package, introduced here, is closest to a cross between the Johansen (1995) cointegration model in **urca** and the models involving fractionally integrated variables discussed above. In particular, the model estimated in the **urca** packages is the special case of **FCVAR** in which the fractional integration parameters d and b are both equal to one.[3] The R package **FCVAR** closely follows a companion package **FCVARmodel.m**, written in MATLAB. The MATLAB package is documented in Nielsen and Popiel (2016), an expanded version of the package documented in Nielsen and Morin (2014).

2 Extended Description of Estimation Options

This material appears in the package documentation. The description in the paper includes only the chosen options relating to the model that is estimated.

Once the data is imported, the user sets the program options. The script contains two sets of options: variables set for function arguments in the script itself and model/estimation related options. The first of set of options is as follows.

```
p            <- ncol(x1)
kmax         <- 3
order        <- 12
printWNtest  <-  1
```

Listing 1: Initialization of local variables

[3] They also use the Danish data, so it might be worthwhile to include in the documentation an example with and without the restriction $d = b = 1$ to compare. It may not fit in this paper but could be a short vignette (short pdf with code and descriptions) that would go on the CRAN webpage for the package.

The variable `kmax` determines the highest lag order for the sequential testing that is performed in the lag selection, whereas `p` is the dimension of the system. The `order` specifies the number of lags used for the white noise test in lag selection, while `printWNtest` indicates whether to print results of white noise tests post-estimation.

The next set of initialization commands assign values to the variables contained in the object `opt` defined by the function `FCVARoptions`.

```
# Define variable to store estimation options.
opt                <- FCVARoptions()
opt$dbMin          <- c(0.01, 0.01) # lower bound for d, b.
opt$dbMax          <- c(2.00, 2.00) # upper bound for d, b.
opt$unrConstant    <- 0   # include an unrestricted constant?
opt$rConstant      <- 0   # include a restricted constant?
opt$levelParam     <- 1   # include level parameter?
opt$constrained    <- 0   # impose restriction dbMax >= d >= b >= dbMin?
opt$restrictDB     <- 1   # impose restriction d = b ?
opt$db0            <- c(0.80, 0.80) # set starting values for optimization.
opt$N              <- 0   # number of initial values to condition upon.
opt$print2screen   <- 1   # print output.
opt$printRoots     <- 1   # do not print roots of characteristic polynomial.
opt$plotRoots      <- 1   # do not plot roots of characteristic polynomial.
opt$gridSearch     <- 1   # For more accurate estimation, perform a grid search.
                          # This will make estimation take longer.
opt$plotLike       <- 0   # Plot the likelihood (if gridSearch <- 1).
opt$progress       <- 0   # Show grid search progress indicator waitbar.
opt$updateTime     <- 0.5 # How often progress is updated (seconds).

# Store the options to reset them in between hypothesis tests.
DefaultOpt <- opt
% \end{Code}
```

Listing 2: Choosing estimation options

The first line initializes the object `opt` and assigns all of the default options set in `FCVARoptions`. The user can see the full set of options by typing `DefaultOpt` (or `opt` after initialization) in the command line. The code block above shows how to easily change any of the default options.

Defining the program options in this way allows the user to create and store several option objects with different attributes. This can be very convenient when, for example, performing the same hypothesis tests on different data sets.

The set of available options can be broken into several categories: numerical optimization, model deterministics and restrictions, output and grid search. We recommend that only advanced users make changes to the numerical optimization options. Adding deterministics requires setting the variable corresponding to the type of deterministic component to 1. For instance, in the present example, a model estimated with options `opt` will include the level parameter μ but no restricted or unrestricted constant. Output variables refer to either printing or plotting various information post-estimation and usually take values 1 or 0 (on or off). For example, if the user is not interested in the estimates of Γ, they can be suppressed by setting `opt$printGammas <- 0`.

The bounds on the parameter space for d and b are specified in `opt$dbMin` and `opt$dbMax`. In this example, these are both specified as 2-dimensional column vectors, in which case the first element specifies the bound on d and the second element the bound on b. Alternatively, one can set `opt$dbMin` and `opt$dbMax` as scalars, which imposes the same bounds on d and b.

An important feature in this package is the ability to pre-estimate by using a grid search. If the user selects this option, they can view progress by setting `opt$progress` to 1 (waitbar) or 2 (output in command line). The minimum frequency of these updates is set by `opt$updateTime`. The user also has the option (`opt$plotLike`) to view a plot of the likelihood over d and/or b after the grid search completes. The output of the grid search is a preliminary estimate of the fractional parameters. These are used as starting values in the subsequent numerical optimization, and the bounds on d and b are set to these starting values plus/minus 0.1 but still within the original `dbMin` and `dbMax` settings.

Warning: The current package does not have the following functionality. It could be implemented but I imagine there already exists something in R that can serve this role. However, I think I don't quite understand the problem without it. Can you provide an example?

As of v.1.4.0, the new option `opt$LocalMax` allows more control over the grid search. If `opt$LocalMax <- 0`, the function `FCVARlikeGrid` returns the parameter values corresponding to the global maximum of the likelihood on the grid. If `opt$LocalMax <- 1`, then `FCVARlikeGrid` returns the parameter values for the local maximum corresponding to the highest value of b. This is meant to alleviate the identification problem discussed in Johansen and Nielsen (2010, Section 2.3) and Carlini and de Magistris (2019). As of v.1.4.0, the default setting is `opt$LocalMax <- 1`.

Another option is the addition of a line search to the switching algorithm for estimation of models with restrictions on α and/or β. This is added via the option `opt$LineSearch <- 1` and is the default. See Doornik (2018, Section 2.2) for details.

After all options have been set, the last line stores them in `DefaultOpt` so that the user can recall them at any point in the estimation. This is particularly useful if the user wants to change only a few options in between estimations.

References

Aschersleben, Philipp, and Martin Wagner (2016) **cointReg**: *Parameter Estimation and Inference in a Cointegrating Regression.* R package version 0.2.0

Beran, J. (1995) 'Maximum likelihood estimation of the differencing parameter for short and long memory autoregressive integrated moving average models.' *Journal of the Royal Statistical Society. Series B (Methodological)* 57(4), 659–672

Carlini, F., and P. S. de Magistris (2019) 'On the identification of fractionally cointegrated VAR models with the F(d) condition.' *Journal of Business & Economic Statistics* 37(1), 134–146

Chen, W. W., and C. M. Hurvich (2003) 'Semiparametric estimation of multivariate fractional cointegration.' *Journal of the American Statistical Association* 98(463), 629–642

_ (2006) 'Semiparametric estimation of fractional cointegrating subspaces.' *The Annals of Statistics* 34(6), 2939–2979

Christensen, B. J., and M. Ø. Nielsen (2006) 'Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting.' *Journal of Econometrics* 133, 343–371

Clegg, Matthew (2017) **egcm**: *Engle-Granger Cointegration Models.* R package version 1.0.12

Doornik, J. A. (2018) ‘Accelerated estimation of switching algorithms: The cointegrated VAR model and other applications.’ *Scandinavian Journal of Statistics* 45(2), 283–300

Engle, Robert F., and Clive W. J Granger (1987) ‘Co-integration and error correction: Representation, estimation and testing.’ *Econometrica* 55(2), 251–276

Frederiksen, P. S., F. Nielsen, and M. Ø. Nielsen (2012) ‘Local polynomial whittle estimation of perturbed fractional processes.’ *Journal of Econometrics* 167(2), 426–447

Geweke, J., and S. Porter-Hudak (1983) ‘The estimation and application of long memory time series models.’ *Journal of Time Series Analysis* 4, 221–238

Groebe, Benjamin (2019) **nsarfima***: Methods for Fitting and Simulating Non-Stationary ARFIMA Models.* R package version 0.1.0.0

Hou, J., and P. Perron (2014) ‘Modified local whittle estimator for long memory processes in the presence of low frequency (and other) contaminations.’ *Journal of Econometrics* 182(2), 309–328

Hyndman, Rob J (2020) *CRAN Task View: Time Series Analysis*

Jensen, Andreas Noack, and Morten Ørregaard Nielsen (2014) ‘A fast fractional difference algorithm.’ *Journal of Time Series Analysis* 35, 428–436

Johansen, Søren (1995) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models* (New York: Oxford University Press)

Johansen, Søren, and Morten Ørregaard Nielsen (2010) ‘Likelihood inference for a nonstationary fractional autoregressive model.’ *Journal of Econometrics* 158, 51–66

Maechler, Martin, Chris Fraley, Friedrich Leisch, Valderio Reisen, Artur Lemonte, and Rob Hyndman (2020) **fracdiff***: Fractionally Differenced ARIMA aka ARFIMA(P,d,q) Models.* R package version 1.5-1

Marmol, F., and C. Velasco (2004) ‘Consistent testing of cointegrating relationships.’ *Econometrica* 72(6), 1809–1844

Mayoral, L. (2007) ‘Minimum distance estimation of stationary and nonstationary arfima processes.’ *The Econometrics Journal* 10, 124–148

Nielsen, M. Ø. (2010) 'Nonparametric cointegration analysis of fractional systems with unknown integration orders.' *Journal of Econometrics* 155(2), 170–187

Nielsen, M. Ø., and K. Shimotsu (2007) 'Determining the cointegrating rank in nonstationary fractional systems by the exact local whittle approach.' *Journal of Econometrics* 141(2), 574–596

Nielsen, M. Ø., and P. S. Frederiksen (2011) 'Fully modified narrow-band least squares estimation of weak fractional cointegration.' *The Econometrics Journal* 14, 77–120

Nielsen, Morten Ørregaard, and Lealand Morin (2014) 'FCVARmodel.m: a Matlab software package for estimation and testing in the fractionally cointegrated VAR model.' QED working paper 1273, Queen's University

Nielsen, Morten Ørregaard, and Michał Ksawery Popiel (2016) 'A Matlab program and user's guide for the fractionally cointegrated VAR model.' QED working paper 1330, Queen's University

Pfaff, Bernhard, Eric Zivot, and Matthieu Stigler (2016) **urca***: Unit Root and Cointegration Tests for Time Series Data.* R package version 1.3-0

Phillips, P.C.B., and B. Hansen (1990) 'Statistical inference in instrumental variables regression with i(1) processes.' *Review of Economic Studies* 57, 99–125

Phillips, P.C.B., and M. Loretan (1991) 'Estimating long run economic equilibria.' *Review of Economic Studies* 58, 407–436

Phillips, P.C.B., and P. Perron (1988) 'Testing for a unit root in time series regression.' *Biometrika* 75(2), 335–346

Phillips, P.C.B., and S. Ouliaris (1990) 'Asymptotic properties of residual based tests for cointegration.' *Econometrica* 58(1), 165–193

Qiu, Debin (2015) **aTSA***: Alternative Time Series Analysis.* R package version 3.1.2
R Core Team

R Core Team (2017) *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing (Vienna, Austria)

Robinson, P. M. (1994) 'Semiparametric analysis of long-memory time series.' *Annals of Statistics* 22, 515–539

Robinson, P. M. (1995a) 'Gaussian semiparametric estimation of long range dependence.' *The Annals of Statistics* 23(5), 1630–1661

_ (1995b) 'Log-periodogram regression of time series with long range dependence.' *The Annals of Statistics* 23(5), 1048–1072

Robinson, P. M., and D. Marinucci (2003) 'Likelihood-based inference in cointegrated vector autoregressive models.' In *Time Series with Long Memory,* ed. P. M. Robinson (Oxford: Oxford University Press) pp. 334–373

Robinson, P. M., and Y. Yajima (2002) 'Determination of cointegrating rank in fractional systems.' *Journal of Econometrics* 106(2), 217–241

Robinson, P.M. (2008) 'Diagnostic testing for cointegration.' *Journal of Econometrics* 143(1), 206–225

Saikkonen, P. (1991) 'Asymptotically efficient estimation of cointegrating regressions.' *Econometric Theory* 7, 1–21

Shimotsu, K. (2010) 'Exact local whittle estimation of fractional integration with unknown mean and time trend.' *Econometric Theory* 26, 501–540

Shimotsu, K., and P. C. B. Phillips (2005) 'Exact local whittle estimation of fractional integration.' *The Annals of Statistics* 33(4), 1890–1933

Souza, I. V. M., V. A. Reisen, G. d. C. Franco, and P. Bondon (2018) 'The estimation and testing of the cointegration order based on the frequency domain.' *Journal of Business and Economic Statistics* 36(4), 695–704

Stock, J.H., and M.W. Watson (1993) 'A simple estimator of cointegrating vectors in higher order integrated systems.' *Econometrica* 61, 783–820

Trapletti, Adrian, Kurt Hornik, and Blake LeBaron (2019) **tseries***: Time Series Analysis and Computational Finance.* R package version 0.10-47

Veenstra, Justin Q., and A.I. McLeod (2018) **arfima***: Fractional ARIMA (and Other Long Memory) Time Series Modeling.* R package version 1.7-0

Wang, B., M. Wang, and N. H. Chan (2015) 'Residual-based test for fractional cointegration.' *Economics Letters* 126, 43–46

Zhang, R., P. Robinson, and Q. Yao (2018) 'Identifying cointegration by eigenanalysis.' *Journal of the American Statistical Association* 114(526), 916–927