An introduction to the **getslmem** package

Genaro Sucarrat[1]
This version: 2nd August 2024

Contents:

1 Introduction

Let $y_t \geq 0$ be a non-negative variable. The multiplicative error class of models is given by

$$y_t = \mu_t \eta_t, \qquad \eta_t \geq 0, \qquad E(\eta_t) = 1, \qquad t = 1, 2, \ldots, \tag{1}$$

where $\mu_t > 0$ is a model or prediction of y_t and η_t is the innovation or error. Henceforth, specifications contained in (1) are referred to as Multiplicative Error Models (MEMs).

The package **getslmem** provides a toolkit for the estimation and modelling of logarithmic MEMs, the log-MEM-X model, which comprise both static and dynamic version. The 'X' in the abbreviation means covariates can be included in addition to Autoregressive (AR) terms. The inclusion of AR terms is optional.

The log-MEM-X model is given by

$$\ln \mu_t = \omega + \underbrace{\sum_{i=1}^{p} \alpha_i \ln y^*_{t-i}}_{\text{log-AR terms}} + \underbrace{\sum_{j \in Q} \beta_j \ln \mathrm{EqWMA}_{j,t-1}}_{\text{(log of) moving averages of } y_t} + \underbrace{\sum_{l=1}^{k} \delta_l x_{l,t}}_{\text{“X”, covariates}}, \tag{2}$$

where y^*_{t-i} is a zero-adjusted version of y_{t-i},

$$y^*_{t-i} = \begin{cases} y_{t-i} & \text{if } y_{t-i} > 0, \\ c & \text{if } y_{t-i} = 0,\ c > 0 \text{ a scalar}, \end{cases} \tag{3}$$

EqWMA is short for Equally Weighted Moving Averages of (lagged) y_t's and the x_t's are covariates. The lagged EqWMA terms, which are similar to the heterogeneous volatility proxies of Corsi (2009), can be interpreted as proxies of lagged μ_t's. The EqWMA terms are computed

[1]Department of Economics, BI Norwegian Business School. URL: https://www.sucarrat.net/

as

$$\mathrm{EqWMA}_{j,t-1} = \begin{cases} \frac{(y_{t-1}+\cdots+y_{t-j})}{j} & \text{if } (y_{t-1}+\cdots+y_{t-j}) > 0, \\ c & \text{if } (y_{t-1}+\cdots+y_{t-j}) = 0, \end{cases} \tag{4}$$

where $c >$ is the same scalar used for zero-adjustments of the y_{t-i}'s in (3).

2 Installation

The package **getslmem** requires version 0.39 of the package **gets**, which is currently under development (that is, version 0.39 is not on CRAN yet). The following code can be used to install version 0.39 from Github:

```
install.packages(
  "https://github.com/gsucarrat/gets/raw/master/gets_devel.tar.gz",
  repos = NULL, type = "source"
)
```

Remember to first execute `remove.packages("gets")` if a prior version of **gets** is already installed on your system. Next, to install the package **getslmem** from its Github page, the following code can be used:

```
install.packages(
  "https://github.com/gsucarrat/getslmem/raw/main/getslmem_1.0.tar.gz",
  repos = NULL, type = "source"
)
```

3 Estimation of log-MEM-X models

Consider the following simulated data:

```
set.seed(123)
y <- rchisq(40, df=1)
x <- matrix(rnorm(40*5), 40, 5)
```

To estimate a static log-MEM-X model, static in the sense that no AR terms are included, the following code can be used:

```
library(getslmem)
mymodel <- lmem(y, xreg=x)
print(mymodel)
```

Formally, the object named `mymodel` is of class `"lmem"`, and the `print()` method applied to objects of this class prints the estimation results. To estimate a new model that also contains logarithmic AR-terms at lags, say, 1 and 3, use:

```
mymodel <- lmem(y, ar=c(1,3), xreg=x)
```

If the autocorrelation of y_t is persistent, then it may be a good idea to include lagged Equally Weighted Moving Averages (EqWMAs) of past y_t's. To add the lag of the log of, say, a 5-period EqWMA and a 20-period EqWMA,[2] the argument `log.eqwma` can be used as follows:

```
mymodel <- lmem(y, ar=c(1,3), log.ewma=c(5,20), xreg=x)
```

General-to-Specific (GETS) modelling is a multi-path backwards elimination algorithm for variable selection. It is undertaken by applying the method `gets()` on a start model of class `"lmem"`. For example, the following undertakes GETS modelling of `mymodel` and prints the final, simplified model:

```
simplemodel <- gets(mymodel)
print(simplemodel)
```

The final, simplified model is also of class `"lmem"`. Note that, in GETS modelling of models of class `"lmem"`, the intercept is restricted from removal. To restrict additional variables to be kept, use the argument `keep`, see `help(gets.lmem)`.

4 Available methods (S3)

The following methods can be applied to objects of class `"lmem"`:

`coef()`	extract coefficient estimates
`fitted()`	extract fitted values ($\widehat{\mu}_t$'s)
`gets()`	do General-to-Specific (GETS) modelling
`hist()`	make histogram of residuals ($\widehat{\eta}_t$'s)
`logLik()`	extract log-likelihood
`model.matrix()`	extract regressors
`nobs()`	number of observations
`plot()`	plot fitted values ($\widehat{\mu}_t$'s), actual values (y_t's) and the residuals ($\widehat{\eta}_t$'s)
`predict()`	generate predictions
`print()`	print estimation results
`residuals()`	extract residuals ($\widehat{\eta}_t$'s)
`summary()`	list the named items of the object
`toLatex()`	make a LaTeX print of results (equation format)
`vcov()`	extract variance-covariance matrix

For help on these methods, type `help(coef.lmem)`, `help(gets.lmem)` and `help(predict.lmem)`.

5 Non-stationary log-MEM-X models

The underlying statistical theory that underpins the estimation methods of (2) relies on stationarity. Non-stationary log-MEM-X models can be specified by adding a deterministic component to the multiplicative decomposition. Let

$$y_t = g_t h_t \eta_t, \qquad \eta_t \geq 0, \qquad E(\eta_t) = 1, \qquad t = 1, 2, \ldots, \tag{5}$$

[2]In daily financial data, 5 periods usually corresponds to 5 trading days (i.e. a week) and 20 periods usually corresponds to 20 trading days (i.e. four weeks or about a month).

where $g_t > 0$ is a deterministic process, and $h_t > 0$ and $h_t \eta_t$ are similar to μ_t and y_t, respectively. In other words, writing $\phi_t \equiv h_t \eta_t$, the non-stationary log-MEM-X is given by

$$\ln h_t = \omega + \underbrace{\sum_{i=1}^{p} \alpha_i \ln + \phi_{t-i}^*}_{\text{log-AR terms}} + \underbrace{\sum_{j \in Q} \beta_j \ln \mathrm{EqWMA}_{j,t-1}}_{\text{(log of) moving averages of } \phi_t} + \underbrace{\sum_{l=1}^{k} \delta_l x_{l,t}}_{\text{“X”, covariates}} , \tag{6}$$

where ϕ_{t-i}^* is a zero-adjusted version of ϕ_{t-i} obtained in the same way as in (3), and where the EqWMA terms are computed as in (4) but with the y_t's replaced by ϕ_t's. Now, it is $g_t h_t$ which is the model or prediction of y_t. Note that (6) is also a MEM, but we now have $\mu_t = g_t h_t$. Estimation of (5) proceeds in two steps. First, estimate g_t by a suitable method to obtain estimates of ϕ_t via the relation $\phi_t = y_t / g_t$. Second, use the estimated ϕ_t's to estimate h_t.

References

Corsi, F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics 7*, 174–196.