Short horizon forecast vs. long horizon projection

A Time Series Approach to estimate quantiles of the distribution of the relevant quantity at the projection horizon. An application to DTCC estimation of its Raw Liquidity Need.

Roberto Strepparava[[1]](#footnote-1) and Zhuying Xu[[2]](#footnote-2)

This version: October 27, 2017

# Abstract

We develop a model that, given the time series of past realizations of a quantity subject to periodic jumps of random magnitude, estimates the long term (asymptotic) mean and long term (asymptotic) variance-covariance matrix from which, under normality approximation, asymptotic quantiles can be inferred. We realize this first by producing a short term-term forecast by means of a discrete process that produces an accurate point estimate but only over a short horizon and then by embedding the process in a natural way in a continuous Ornstein-Uhlenbeck (Vasicek) process, from which the asymptotic behavior is extracted. The instability associated with the periodic jumps of random magnitude is suitably kept under control in the error term. An application to DTCC raw liquidity need is described in detail.

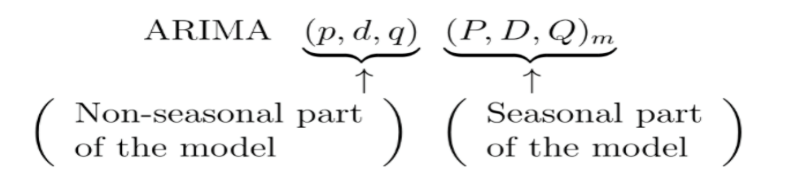
**Keywords**: SARIMA models, ARIMA models, multivariate Ornstein-Uhlenbeck, Vasicek model, auto-regressive (AR) process, vector-autoregression (VAR), Principal Component Analysis (PCA), filtered EWMA (floored EWMA).

# Short Term forecast of a quasi-stationary process via SARIMA

The starting point and, in fact, the only input to the model is a time series where is positive (in our case study measured in ) and is time (again in our case measured in days).

We will call the process quasi-stationary, meaning that it is overall stationary (as measured for example by the Augmented Dickey-Fuller test), with the exception of periodic spikes of random magnitude (in our case study example monthly spikes up to 80% of the average size of the underlying stationary process).

In order to capture these properties of the process, we decided to employ a Seasonal Autoregressive Integrated Moving Average () model to estimate (the Raw Liquidity Need RLN, see last section [Case Study: DTCC Raw Liquidity Need](#_Case_Study:_DTCC)). The (autoregressive) part of this model indicates that the evolving variable of interest is regressed on its own lagged (i.e., prior) values. The (moving average) part of this model indicates that the regression error is actually a linear combination of error terms whose values occurred contemporaneously and at various time in the past. The (Integrated) part indicates that the data values have been replaced with the difference between their values and the previous values and the seasonal part will explain the seasonal patterns in the data. It is written as follows:

****

where number of periods per season. We use uppercase notation for the seasonal parts of the model, and the lowercase notation for the non-seasonal parts of the model.

The final resulting model is a . It is very important for the next steps of the model to note that the true model should be with and , that is we are close to stationarity and we chose to employ an order of integration simply because the software package R does not handle fractional integration.

We divided our dataset into a training set (from 2012 to 2014) and a test set (2015 to 2017). We conducted an out-of-sample 2-day prediction and the results are shown below:

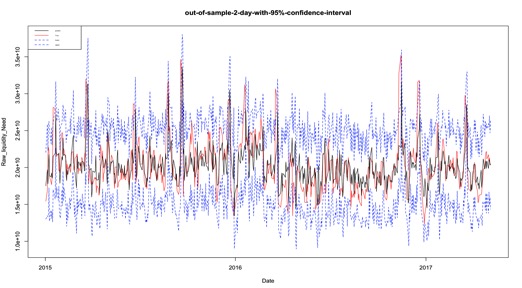


Figure 1 Out-of-sample 2-day RLN prediction with 95% confidence intervals

Plot above shows the out of sample 2-day forecasting of RLN (i.e. ) with 95% confidence intervals where blue, dashed lines are confidence intervals, red line is the true/observed RLN and black line is the predicted RLN. The model successfully captured the shape of RLN. Furthermore, among 582 business days from 2015 to 2017, there are only 19 days for which RLN have not been covered by the predicted 95% upper confidence interval.

However, while the model is well suited for a short time horizon (2-3 days) forecasting, our (and the client’s) objective is to have a model that can be used to project over a much longer horizon (6 months to a year).

In order to accomplish this objective, we expanded the model into a Multivariate Ornstein-Uhlenbeck () with extrapolation of asymptotic moments. The with embedded is capable to project long term RLN .

# Embedding a SARIMA into a Multivariate Ornstein-Uhlenbeck process

The idea behind this part of the model is the following: a forecast is a process that produces an accurate, point estimate but which deteriorates quickly with time, so it is naturally short horizon. If we give up a precise point estimate and we are content with only average quantities (mean, covariance, and quantiles) that is a projection and not a forecast, as a trade-off we can hope to have a long horizon estimate of these average quantities.

This is exactly the case and the short-term, quasi-stationary forecast can be embedded into a continuous, stationary, Gaussian, Markovian process. There is essentially only one nontrivial process with all the previous attributes and it is a Multivariate Ornstein-Uhlenbeck () process. It can be proved that these properties imply mean-reversion, therefore an asymptotic mean and an asymptotic covariance from which we can estimate quantiles of the distribution of the quantity

The transformation (embedding) of a short term forecast into a continuous process occurs in steps, here they are:

### Filter out the seasonal component from and be left with

Typically you can filter out a deterministic, seasonal component in an additive way as follows

Or, more explicitly, our input process can be written as

Where

is the covariate process, i.e. the time series of our input quantity

is the deterministic, seasonal component, separated in an additive way

is the convolution operator

is a linear causal filter of order for the part

is a linear causal filter of order for the Integrated part

is the white noise process

is a linear causal filter for the part

(There are APIs in R, Python or MATLAB that separate additively this component via FFT Fast Fourier Transform).

**Note**: The error term contains the stochastic part of the seasonal component (difference between true stochastic spike and periodic magnitude obtained via FFT of the time series)

It will propagate and plague during all subsequent transformations, hence the need for an explicit control of the error term in the next paragraphs “[Control of the error term through rolling window analysis](#_Control_of_the)” and “[Control of the error term through Filtered EWMA (floored EWMA)](#_Control_of_the_1)”.

### An can be approximated by an process with large

Recall that our process is quasi-stationary, so effectively our is in fact an process, since we cannot deal with fractionally integrated processes ().

We can convert our into an process with , the number of lags, large, thanks to the following

**Proposition**: a , process is equivalent to an process

*Proof*: see [Appendix A.1](#_A.1_Proof_that)

Since we don’t deal with infinite lags, we are happy to deal with with large (10 lags or so). Also we can exploit the R built-in function which, by the symmetric roles played by the and portions of the process can be used to convert to equivalent representation as follows

Where and are vectors of coefficients obtained from the autocorrelation function plot (acf) and partial autocorrelation function plot (pacf) respectively.

### An process is a process

Our original process now transformed into (approximated by) an process reads

Where is the deterministic, seasonal portion of the and is the error term which no longer is simple white noise since it contains the difference between the periodic jumps with random magnitude and the periodic component with constant magnitude.

We can represent the as a vector auto regression process, doing something quite similar to what is being carried out when we represent a scalar Ordinary Differential Equation (ODE) of order as a vector ODE of order .

which we can write in compact form as

Where

, , and

### A discrete process is an integrated process evaluated at

In what follows we will follow [Meucci 2010], [Neumaier et al. 1997] and [Behme et al. 2013].

The multivariate Ornstein-Uhlenbeck (also called Vasicek) process is defined by the following stochastic differential equation

In this expression is the transition matrix, namely a fully generic square matrix that defines the deterministic portion of the evolution of the process; is a fully generic vector, which represents the unconditional expectation; is the scatter generator, namely a fully generic square matrix that induces the dispersion of the process and is typically assumed to be a vector of independent Brownian motions, although more in general it is a vector of independent Levy processes [Meucci 2010], so that it includes the case of random jumps as in our situation.

**Proposition**: the integrated process reads

where

*Proof*: see [Appendix A.2](#_A.2_Construction_of)

Now let’s write side by side our process evaluated between time steps and our integrated process corresponding to a time increment

It is clear from that the two processes are equivalent, so that a process is a process sampled at discrete unit time. We can also express all the quantities of the stochastic differential equation defining the in terms of the coefficients .

* **Proposition**: let ,

then

where is the operator that stacks the columns of a matrix, is the operator that reshapes a vector back into a square matrix and is the Kronecker sum.

*Proof*: see [Meucci 2010] formula with and [Van der Werf 2007]

### From compute the asymptotic mean and the asymptotic covariance

Now that we have established the equivalence of our initial process with its corresponding , we can focus on the latter. Its main characteristic is that is Gaussian on the short time scale (e.g. square root law for the propagation of standard deviation with time, for small times), while it admits asymptotic finite moments at the (infinite) horizon. This property can be summarized as follows:

**Proposition**: let a defined by and characterized by parameters (transition matrix),  **(**unconditional expectation vector) and (scatter generator); define

Then

Where is the operator that stacks the columns of a matrix and is the Kronecker sum.

*Proof*: see [Meucci 2010] formulas and

Remark: Note the similarity with the 1-dimensional case (Vasicek model), where

### Via PCA compute the leading eigenvalue of and estimate quantiles of

The final step of our model is to wrap up everything and estimate a quantile of the asymptotic distribution of our quantity .

First we fit to a model, then we embed it into a process from which we extract asymptotic mean and asymptotic covariance . The first component of will represent our long term mean (see equation ).

We can perform PCA (aka eigenvalue decomposition) on and obtain

Finally, under normality assumption – which at this stage can be considered pretty reasonable— we can express any quantile of the distribution of as

Where is the quantile level and is the cumulative of the standard normal distribution function . To be precise care should be taken when looking at the eigenvalues of , since complex conjugate eigenvalues (i.e. rotations + dilations) could lead to a variety of behavior of the asymptotic system [Meucci 2010], [Neumaier et al. 1997], but we can be sure that the approximation done by taking the leading positive eigenvalue leads to satisfactory results.

# Control of the error term through rolling window analysis

As pointed out in the previous sections, the error term in **inherits the instability of the system**. In other words the error term contains the difference between the deterministic seasonal component obtained from the time series via Fast Fourier Transform and the periodic random spike. This instability propagates directly in the estimation of the asymptotic covariance and leads to poor, extremely volatile, estimates of the quantiles .

The approach to tame the instability is a rather holistic one and consists of collecting enough historical samples of the error term and pick a suitable percentile: large enough to account for large deviations between the random spike and its constant magnitude counterpart, but not too large to avoid outliers. Our choice was to select the 95th percentile of a rolling window for the error term

The preference goes to the higher quantile because, as it will be discussed in the case study below, we need to ensure that the chosen quantile is greater than any level of the quantity within a certain confidence level; in other words .

We perform the analysis in the following steps:

* Center the RLN so that it has mean 0.
* Choose a rolling window size, m, i.e., the number of consecutive observation per rolling window. The size of the rolling window will depend on the sample size , and periodicity of the data. In our case, we choose
* The number of increments between successive rolling windows is 1 day, then partition the entire data set into subsamples. The first rolling window contains observations for day through , the second rolling window contains observations for day through , and so on.
* Estimate the error using each rolling window subsamples.
* Select the maximum or 95th percentile of the errors over the rolling window as our final estimate of the error term.

Plots below show the true RLN data (centered) from 2012 to 2017 with 95% confidence intervals using maximum of the errors over the rolling window and 95 percentile of the errors over the rolling window, respectively. Both plots show that our estimated long-term distributions cover the true RLN data well during 2012 and 2017.

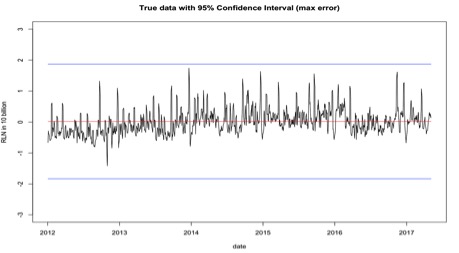


Figure 2 Estimated long-term distributions with 95% Confidence Intervals using maximum of the rolling window errors

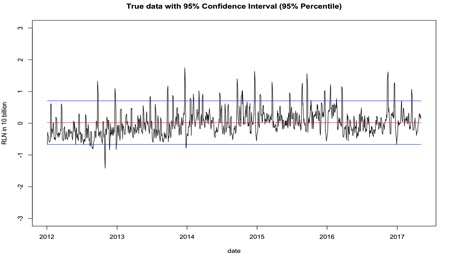


Figure 3 Estimated long-term distributions with 95% Confidence Intervals using 95 percentile of the rolling window errors

Plot below shows the in-sample back testing from 2012 to 2016, where blue lines using the error term estimated as the 95 percentile of the historical error term distribution and green lines using the error term estimated as the maximum historical error term. All estimations show our long-term asymptotic estimation successfully cover the RLN during the test period.

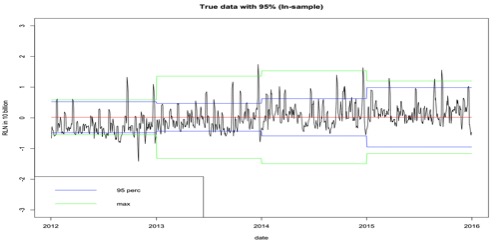


Figure 4 In-sample back-testing using rolling window estimation

Plot below shows the out-of-sample back testing from 2012 to 2016, where blue lines are our 95 percentile error estimations using rolling window estimation and green lines are maximum error estimations. All estimations show our long-term out-of-sample asymptotic estimation successfully cover the RLN during the test period.

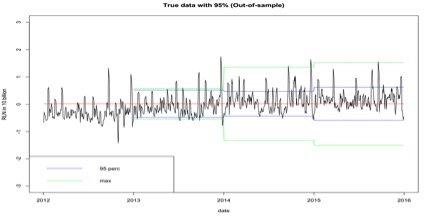


Figure 5 Out-of-sample back-testing using rolling window estimation

\

# Implicit dependence on Volatility and Volatility Regime Change Analysis

It is important to note that in our model the dependence is present but implicit:

* It is plain because, logically, volatility is the main driver of spikes of random (large) magnitude in the sequence of the quantity and it is also verified by comparison of the corresponding time series.
* It is implicit because the scatter matrix of the process is not directly estimated from the data, but rather inferred from the autoregressive coefficients and error term of the fit. Therefore we cannot apply directly GARCH techniques to stabilize the process as a function of volatility of the time series.

We therefore also used in-sample and out-of-sample back testing from 2012 to 2017 to show there is a volatility regime change.

Plot below shows the in-sample back testing from 2012 to 2017, where blue lines are our 95th percentile error estimations using rolling window estimation and green lines are 70th percentile error estimations. Bandwidths for the 95th percentile error estimations from 2016 are significantly larger than the previous years. The volatility within each year (volatility for a particular date has been calculated as the standard deviation using the data from the first day of a year to that date) is shown in Figure 7. It can be seen that volatility during 2016 is much higher than all previous years resulting in much larger bandwidths for error term estimation. This might be due to the ‘Chinese Scare’ during the last quarter of 2015 and propagating in early 2016 and the Presidential election in late 2016.



Figure 6 In-sample back-testing using rolling window estimation from 2012 to 2017



Figure 7 Volatility within each year

# Control of the error term through Filtered EWMA (floored EWMA)

In order to better model the periodic random spikes captured by the error term, instead of estimating it by taking a large quantile of its historical distribution, we calculate it using a floored EWMA () on the historical error terms, a technique well known both in Risk Management (filtered VaR), volatility modelling [Murphy et al. 2014] and forecasting [Grillenzoni 2013]. Essentially it’s a plain EWMA process on the error term series but with a floor that guarantees an upward bias, which is necessary since we want a sufficiently large quantile that covers all with a high confidence over the long horizon.

In addition it is known that this methodology works very well in presence of volatility regime changes that we described in the previous section. In the case of these market turbulences or significant volatility changes, as shown in Figure 8 in this section, the filtered EWMA with a floor methodology for estimating the error term successfully captures a few spikes during the same back testing period, which outperforms the rolling window estimation.

The formula for reads

where has to be suitably chosen, for example

1. A hardcoded value that represents a prior view on the error term
2. The initial period error estimate
3. A long term estimation, for instance from the previous period is 95 percentile of previous error .

We choose to use the long-term average as the floor and set the decay factor as (we also tried different decay factors such as 0.93 and the results are similar in this case). The resulting estimated asymptotic long-term mean and variance is shown in the plot below. Consistent with the rolling window in section 4.3, we set rolling window as m=500 and use the first 250 days (1-year) to obtain the long-term average . Plot below shows that the error estimation by filtered EWMA (blue line) performs better than that by rolling window (orange line) as the error estimation by filtered EWMA successfully captures a few spikes.

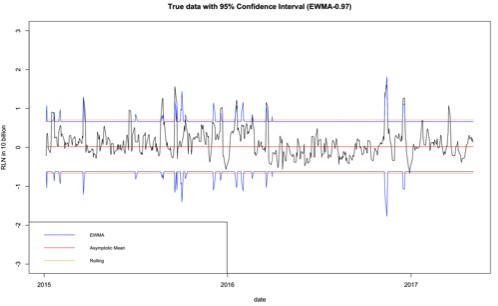


Figure 8 Estimated long-term distributions with 95% Confidence Intervals using EWMA approach

# Case Study: DTCC Raw Liquidity Need

The model we have just described in the previous paragraphs originated from a request for model development from DTCC –Deposit and Trust Clearing Corporation—. Its Enterprise Risk Management needs a model for its raw liquidity need.

This quantity represents the buffer to be allocated by DTCC, a major Clearing House, in case of default of a clearing member.

For each clearing member (currently 183) the long only positions (from the point of view of DTCC) are computed: each of these quantities represents the obligation that the clearing member has with the Clearing House. The net positions (netting occurring) within three days are the liabilities, the obligations that the clearing member has with DTCC. Of these 183 members we identify five families, that is five clearing members that alone sum up roughly 90% of these liabilities and for each of these families its own **raw liquidity need** is the time series of these long (w.r.t. DTCC) positions.

We will call **cover1** the time series given by the max of each of these five raw liquidity needs day by day. Typically, the dominant family stays dominant for a long time and crossovers happen seldom (transition Goldman -> Morgan Stanley only once). And even in case of a crossover, the cover1 path displays continuity and no sudden jump or regime change, meaning that the transition from one family to another occurs smoothly after a steady approach to each other.

The cover1 time series displays unique features that require an ad-hoc model to handle it: the process is stationary for the majority of dates, with a pronounced mean-reversion to an average level, but on each third Friday and subsequent Monday, Tuesday and Wednesday of each month the raw liquidity need spikes because of the exercise of in-the-money options. During these **special dates** the raw liquidity need process spikes up to 80% mode than the average non special date spike. The magnitude of these spikes can be approximated by a jump process, since any other estimation would be formidable since it would involve knowledge of the portfolios of the clearing members’ family.

Therefore cover1 can be very well described by the initial hypotheses stated for this model: quasi-stationarity, periodicity of the jumps, random magnitude of these jumps.

The business requirement from DTCC Enterprise Risk Management was to provide a model for cover1 so that DTCC could be safe in case of default of its major family (i.e. its cover1).

The initial model developed was the described above, which is well suited for a short time horizon: its predictive power as a point estimate can be measured in 2—3 days before deterioration.

After input from DTCC Enterprise Risk Management for a model that could predict quantiles over a long horizon (at least 180 days), we twisted the model into its final form, i.e. its embedding into a with extrapolation of asymptotic moments.

A nice feature of the model is that it can be recalibrated within minutes (given it takes only one relatively simple input), meaning that, once a regime change is detected, a brand new model with different asymptotic behavior can be computed.

Another nice feature of the model is that it is *generic*, meaning that it can be applied to different settings within or outside DTCC. For example in case of a quasi-stationary time series with periodic jumps *without* randomness, the model can be applied almost verbatim, basically in a simplified version since there is no need to control any instability of the error term

# Conclusions and future work

Let us reiterate that this model is *generic*, which is good: it can be applied to virtually any time series of quantities that satisfy the same properties as the raw liquidity needs, which are quasi-stationary with periodic spikes of deterministic or random magnitude.

Also let us recall that this model is a two-step process: first forecast with and then project over an asymptotic horizon via embedding in a . This has been done on purpose to leverage the good results of the short term forecast, **but certainly the process could be calibrated directly on the data itself**. This represents a future direction of research, keeping in mind that the seasonal component of random magnitude presents a problem for a direct fitting of a on the data (it fits well on stationary data only, without the seasonal random component)

In fact a potentially more robust and more mathematically sound approach would be a process subject to periodic jumps of random magnitude and **a novel mathematical development of its embedding into a continuous process, which would probably lead outside the simple setting of a Multivariate Ornstein-Uhlenbeck**, presumably into Lévy processes and Bates models.

The limitations due to certain assumptions on the process, e.g. Gaussianity and Markov property, do not seem to be critical as they are very common assumptions across all modeling of most time series and any modeling would be extremely difficulty without them.

# References

Meucci A., 2010, Review of Statistical Arbitrage, Cointegration and Multivariate Ornstein-Uhlenbeck, symmys.com > Research > Working Papers, 1—19

Neumaier A., Schneider T., 1997, Multivariate Autoregressive and Ornstein-Uhlenbeck Processes: Estimates for Order, Parameters, Spectral Information and Confidence Regions, researchgate.net, 1—33

Behme A., Lindner A., 2013, Multivariate Generalized Ornstein-Uhlenbeck Processes, 7th International Conference on Lévy Processes: Theory and Applications, 1—56

Murphy D., Vasios M., Vause N., 2014, An investigation into the procyclicality of risk-based initial margin models, Bank of England Financial Stability Paper, No. **29**, 1—19

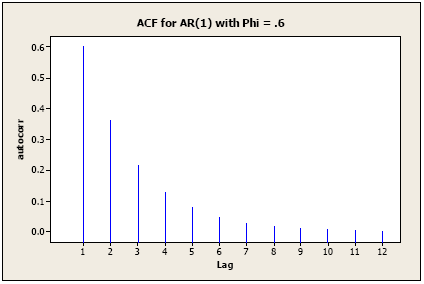
Grillenzoni C., 2012, Sequential Estimation and Control of Time-Varying Unit Root Processes with an Application to S&P Stock Price, Sequential Analysis - Taylor & Francis, 1—24

Van der Werf, K. W., 2007, Covariance of the Ornstein-Uhlenbeck process, Personal Communication.

# Appendix

### A.1 Proof that a , process is equivalent to an process

1. * From the autocorrelation function plot (acf) or from the equation by repeated substitution you obtain



1. Similarly
   * From the equation you obtain , that is
   * The proof is almost identical to the case and involves the sum of geometric series

### A.2 Construction of the integrated process

To integrate introduce the integrator

Using Ito’s Lemma we obtain

Integrating both sides of the preceding relation and substituting the def. of we obtain and

1. KPMG US LLP Advisory, 345 Park Avenue, New York, NY 10154. [rstrepparava@kpmg.com](mailto:rstrepparava@kpmg.com) [↑](#footnote-ref-1)
2. KPMG US LLP Data Science, Lighthouse, 560 Lexington Avenue, New York, NY 10022, [zhuyingxu@kpmg.com](mailto:zhuyingxu@kpmg.com) [↑](#footnote-ref-2)