Tomas Bata University in Zlín
Faculty of Management and Economics

DRIVE
YOUR
KNOWLEDGE
BE
A
SCIENTIST

28/4/2016

CONFERENCE PROCEEDINGS

12th Annual International Bata Conference
for Ph.D. Students and Young Researchers

Conference Proceedings

DOKBAT

12th Annual International Bata Conference

for Ph.D. Students and Young Researchers

The publication was released within the DOKBAT conference, supported by the IGA project.

Many thanks to the reviewers who helped ensure the quality of the papers.

Edited by: Ing. Lukáš Danko, Ing. Markéta Slováková

ISBN: 978-80-7454-592-4

Expert guarantor of the conference:

Ing. Jana Matošková, Ph.D.

Manager and coordinator of the conference:

Ing. Filip Kučera

Members of the organizing team:

Ing. Zuzana Crhová

Ing. Lukáš Danko

Ing. Martin Horák

Ing. Tomáš Janů

Ing. et Ing. Karel Kolman

Ing. Viera Pechancová

Ing. Markéta Slováková

Ing. Blanka Vytrhlíková

CONTENT

THE APPLICATION OF EXPECTATION MAXIMIZATION TO MANAGE MISSING DATA, BIASES VALUE-AT-RISK AND VOLATILITY MODELS IN FINANCIAL TIME SERIES?

Gábor Dávid Kiss, Marianna Sávai

Abstract

Multivariate time series analysis requires synchronized and continuous data for its models. However, there can be special occasions when one or some data is missing due to lack of trading activity. This paper focuses on the impact of different missing data handling methods on GARCH and Value-at-Risk model parameters, namely the volatility persistence and asymmetry and the fat-tailness of the corrected data. The main added value of current paper is the comparison of the impact of different methods (like listwise deletion, mean-substitution and maximum-likelihood-based Expectation Maximization) on daily financial time series, because this subject has insufficient literature. Current study tested daily closing data of floating currencies from Kenya (KES), Ghana (GHS), South Africa (ZAR), Tanzania (TZS), Uganda (UGX), Gambia (GMD), Madagascar (MGA) and Mozambique (MZN) in USD denomination against EUR/USD rate between March 8 2000 and March 6 2015 acquired from Bloomberg database. Current paper suggest the usage of mean substitution or listwise deletion for daily financial time series due to their tendency to have a close-to-zero first momentum.

Keywords: missing data, EM method, Value at Risk, GARCH

1 INTRODUCTION

Multivariate time series analysis requires synchronized and continuous data for its models. However, there can be special occasions when one or some data is missing due to lack of trading activity. This paper focuses on the impact of different missing data handling methods on GARCH and Value-at-Risk model parameters, namely the volatility persistence and asymmetry and the fat-tailness of the corrected data.

Missing data (or missing values) is defined "as the data value that is not stored for a variable in the observation of interest" (Kang 2013), where time series can be affected by wave nonresponse cases as the suspension of data generating process is only a temporary issue (Graham 2012). There is a strong supposition about time series: they should not contain missing observations without biasing the ARIMA and GARCH parameters, reduction of representativeness or statistical power due to their impact on the mean, variance, and auto-correlation (Juan Carlos et al. 2010, Kang 2013).

The main added value of current paper is the comparison of the impact of different methods (like listwise deletion, mean-substitution and maximum-likelihood-based Expectation Maximization) on daily financial time series, because this subject has insufficient literature.

Current study tested daily closing data of floating currencies from Kenya (KES), Ghana (GHS), South Africa (ZAR), Tanzania (TZS), Uganda (UGX), Gambia (GMD), Madagascar (MGA) and Mozambique (MZN) in USD denomination against EUR/USD rate between March 8 2000 and March 6 2015 acquired from Bloomberg database.

The paper has the following structure: Theoretical Background section summarizes the main assumptions behind missingness of data, pointing on some differences between query and time series data. Then listwise deletion, mean substitution and Expectation Maximization approaches were introduced in Methods section as well as GARCH and DCC-GARCH models and a basic Value-at-Risk application. The Results and Data section presents the statistical properties of raw, unsynchronized time series and compares it with the three approaches to identify their tendencies for bias.

2 THEORETICAL BACKGROUND

Financial time series, like daily closing currency data can be missing due to the lack of trading activity on the specific data – while other markets are open. Therefore the phenomena has a multivariable-dimension. This temporary suspension of market data can be originated by national differences like holydays and weekends, or by market forces like illiquid situations (in small-cap shares mostly) or when trading activity is suspended due to a sudden collapse in pricing. There is a huge literature how pricing and market efficiency is affected by such brakes as the most cited "weekend effect" appears (Keim – Stambaugh 1984, Robins – Smith 2015, Shahid – Mehmood 2015).

The literature distinguishes among three forms of mechanism behind missingness (Graham 2012, Junger – Leon 2015): one can assume that data is missing completely at random (MCAR), when missingness does not depend on the values of the data or other observed particular variable and their exclusion do not bias our estimations due to their homogeneity (Enders 2010, Junger – Leon 2015, Kang 2013). Missing at random (MAR) happens when dropout conditionally independent of the variable (Kang 2013), but we can assume some sort of mechanism what is behind the missingness (Graham 2012). Their exclusion may corrupt temporal structures such as autocorrelation, trends, and seasonality (Junger – Leon 2015). Missing no at random (MNAR) case occurs when it is possible to make and unbiased estimation to model the missing data. When missingness is beyond researcher's control (their distribution is unknown), MAR is only an assumption (Graham 2012).

Following Baraldi et al. (2015), there are three different approaches to assess the missing data problem. First we can remove those time intervals, when there is at least one missing data for a specific date. Listwise deletion or last observation carried forward scheme can make time series more fragmented or may introduce bias in the estimation of the parameters unless there is a chance that our missingness is MCAR (Kang 2013). Second approach substitutes the missing data by the unconditional mean value or the median (for skewed data, suggested by Junger – Leon 2015) of the available historical data. It has a similar impact like the last observation carried forward scheme for the calculated logarithmic returns for time series with zero mean and mode[5]. This solution is not recommended by Graham (2012) due to its distortions to make a higher concentration around the mean and underestimate errors and varience at MCAR states (Junger – Leon 2015, Enders 2010). Third, there are the modern, computation based approaches to reconstruct missing data trough minimization of an error function, derived from mean variance or a likelihood ratio (Baraldi et al. 2015, Ceylan et al. 2013, Juan Carlos 2010). Expectation maximization (EM) models applying maximum likelihoods to estimate variance, covariance matrixes of the data, while there are also neural networks-based and genetic structure-based approaches as well (Ceylan et al. 2013, Juan Carlos 2010). The expectation maximization takes more computation time, because EM algorithm may be as difficult to compute as the likelihood function itself (Ruud 1991) and

[5] Assuming that P_t and P_{T-1} prices are equal, their logarithmic return will be zero: $r_T = \log(P_t / P_{T-1}) = (P_t - P_{T-1}) = 0$.

they require more specification of a data generation model (Houari et al. 2013), but they do not rely on the MCAR requirement is a feature that remains to be fully exploited. Unbiasedness under MAR and higher efficiency under MCAR make maximum likelihood the method of choice in situation with incomplete multinormal data (Wothke 1998). They are less biased than listwise and pairwise deletion and mean-imputation methods, but this advantage depends on the missing-data rate, the covariance structure of the data and size of the sample (Wothke 1998).

Missing data problems can affect daily time series under multivariate applications like volatility spillover, extreme fluctuation or contagion modelling, where assumptions about conditional variance, covariance and correlation are critical.

3 METHODS

This paper applies and compares three different missing value handling method to capture their ability to maintain central moments, autocorrelation, volatility persistence and extreme movements. All methods were used on daily closing data of African floating currencies and EUR in USD denomination between March 8 2000 and March 6 2015.

Let us assume n foreign exchange rates (1), where the ith ($1 \leq i \leq n$) currency has the following p price for every y trading day with v sample size:

$$P_i = \begin{bmatrix} y_1 & p_{i,1} \\ \cdots & \cdots \\ y_v & p_{i,v} \end{bmatrix}. \tag{1}$$

There is also a kth ($1 \leq k \leq n, and\ k \neq i$) currency (2) with w data, and z ($z \neq y$) time indices:

$$P_k = \begin{bmatrix} z_1 & p_{k,1} \\ \cdots & \cdots \\ z_w & p_{k,w} \end{bmatrix}. \tag{2}$$

Upper $P_{1,\dots,i,k,\dots,n}$ matrices should be united for purposes of multivariate analysis, what requires the synchronization of time indices.

Listwise deletion (3) means a T cap of specific time indices to exclude all cases where at least one value is missing:

$$T = Y \cap Z. \tag{3}$$

Mean substitution (4) can be applied only on logarithmic returns due to their near zero mean and mode. Last observation carried forward (LSCF) scheme similar benefits on prices – with a zeroed logarithmic return at the end:

$$T = (Y \cup Z)\ with\ p_{i,o} = p_{i,o-1}\ for\ T \notin (Y \cap Z). \tag{4}$$

Regularized expectation maximization (EM) algorithm based on iterated linear regression analyses, but it replaces the conditional maximum likelihood estimation of regression parameters for Gaussian data (5), following Schneider (2001). For each $p_{t,i} \in P$ with missing values, the relationship between the prices with missing values (trading days) and the prices with available values is modelled by a linear regression model:

$$p_{NaN} = \mu_{NaN} + (p_a - \mu_a)B + \varepsilon \quad (5)$$

Where a represents available data, and $B \in \Re^{n_a \times n_{NaN}}$ is a matrix of regression coefficients with covariance matrix with missing and available data from n all sample markets. The $\varepsilon \in \Re^{1 \times n_{NaN}}$ residual is assumed to be a zero-mean and $C \in \Re^{n_{NaN} \times n_{NaN}}$ unknown covariance matrix vector. In each iteration of the EM algorithm, estimates of the mean $\mu \in \Re^{1 \times n}$ and of the $\Sigma \in \Re^{n \times n}$ covariance matrix are taken as given, and from these estimates, the conditional maximum likelihood estimates of the matrix of regression coefficients B and of the covariance matrix C of the residual are computed for each record with missing values – to fill each missing values are with imputed values, before recomputation of the entire μ vector and Σ matrix. Then the estimated regression coefficients will be the product of the two (missing-missing and available-missing) estimated covariance matrixes: $\hat{B} = \widehat{\Sigma_{aa}}^{-1}\widehat{\Sigma_{aNaN}}$ to estimate the residual covariance matrix later. However, the regularized EM algorithm for each record with missing values uses $\hat{B} = \left(\widehat{\Sigma_{aa}} + h^2 Diag(\widehat{\Sigma_{aa}})\right)^{-1}\widehat{\Sigma_{aNaN}}$ with a h regularization parameter to inflate diagonal elements with a $1 + h^2$ factor.

Sensitivity analysis is required to examine the bias of uncertain input on the model, where the maintenance of the central tendencies, autocorrelation is studied as well as the patterns of the percentage of missing data (Kang 2013, Graham 2012). Variance models can be affected by missing data, making model selection and parameterization biased. Different GARCH models were fitted on the data to analyse patterns of volatility persistence, following Cappeiello, Engle and Sheppard (2006). The applied GARCH(p,q), GJR GARCH(p,o,q), TARCH(p,o,q) and APARCH(p,o,q) (6-10) models can be useful to capture volatility developments and their clustering in time (heteroscedasticity).

GARCH (p,q): $$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \, \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_i \, \sigma_{t-j}^2 . \quad (6)$$

where σ_t^2 represents present variance, ω is a constant term, p denotes the lag number of squared past ε_{t-i}^2 innovations with α_i parameters, while q denotes the lag number of past σ_{t-j}^2.variances with β_i parameters to represent volatility persistence. Asymmetric GARCH models can be introduced via

$$\begin{cases} S_{t-i}^- = 1, if \, \varepsilon_{t-i} < 0 \\ S_{t-i}^- = 0, if \; \varepsilon_{t-i} \geq 0 \end{cases}$$ as a sign asymmetric reaction to decreasing returns.

(7)

GJR GARCH (p,o,q): $$\sigma_t = \omega + \sum_{i=1}^{p} \alpha_i |\varepsilon_{t-i}| + \sum_{i=1}^{o} \gamma_i S_{t-i}^- |\varepsilon_{t-i}| + \sum_{j=1}^{q} \beta_i \, \sigma_{t-j} , \quad (8)$$

TARCH (p,o,q): $$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \, \varepsilon_{t-i}^2 + \sum_{i=1}^{o} \gamma_i \, S_{t-i}^- \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_i \, \sigma_{t-j}^2 , \quad (9)$$

APARCH (p,o,q): $$\sigma_t^\delta = \omega + \sum_{i=1}^{p} \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^\delta , \quad (10)$$

where $\alpha i > 0$ (i=1,…,p), $\gamma i + \alpha i > 0$ (i=1,…,o), $\beta i \geq 0$ (i=1,…,q), $\alpha i + 0,5\ \gamma j + \beta k + < 1$ (i=1,…,p, j=1,…,o, k=1,…,q) and δ index parameter can be between 1 and 2.

Model selection was made with focus on homoscedastic residuals (using a 2 lagged ARCH-LM test), searching for the lowest Bayesian Information Criteria (BIC). This study applies DCC-GARCH6 model, following Engle (2002), to analyze the daily common movements of the selected markets.

Missing values have an impact on the density function of the data – listwise deletion assumed to make more data on the tails, while mean substitution can increase the representation of the expected value. The EM should produce data between these mean and extremes. Extreme fluctuation of the data was tested with ordinary Value-at-Risk model (11), where the weight of extreme data and the kurtosis of non-extreme data were the variables of my sensitivity analysis.

$$r = r_n + r_x^- + r_x^+,\ where\ r_x^- < \mu - 1.65 * \sigma\ and\ r_x^+ > \mu + 1.65 * \sigma\ , \qquad (11)$$

where r is a logarithmic return, μ unconditional mean, σ unconditional standard deviation, r_x^- represents extreme negative, r_x^+ extreme positive returns and r_n denotes a non-extreme subset of data (Madura 2008).

This study applies Dynamic Conditional Correlation GARCH (DCC-GARCH)7 model (12), following Engle (2002), to analyze the daily common movements of the selected markets. It is defined as:

$$r_t = \mu_t + \alpha_t, \alpha_t = H_t^{1/2} z_t, H_t = D_t R_t D_t, R_t = Q_t^{*-1} Q_t Q_t^{*-1}, Q_t = (1 - a - b)\bar{Q} + a\varepsilon_{t-1}\varepsilon_{t-1}^T + bQ_{t-1} \qquad (12)$$

Where r_t denotes log returns, α_t mean-corrected returns ($E[\alpha_t] = 0\ and\ Cov[\alpha_t] = H_t$), μ_t expected value of the conditional r_t, H_t matrix of conditional variances of α_t, $H_t^{1/2}$ after Choelsky factorization, D_t conditional standard deviations of , R_t conditional correlation matrix, z_t vector of independent and identically distributed errors, Q_t unconditional covariance matrix of the standardized errors ε_t (Cappeiello et al. 2006).

Current paper applies the following setup during the comparison of three different methods: deviation between raw and refined data was studied in central moments, autocorrelation, heteroscedasticity, normal distribution, weak stationarity, GARCH model and parameter selection and Value-at-Risk weights and kurtosis as well as dynamic conditional correlation results.

4 RESULTS AND DATA

Statistical properties of raw and refined data were compared in this section to present the underlying differences among missing value handling approaches and their impact on model parameterization.

[6]The estimation based on the Oxford MFE and UCSD toolboxs, developed by Kevin Sheppard: http://www.kevinsheppard.com/

[7]The estimation based on the Oxford MFE and UCSD toolboxs, developed by Kevin Sheppard: http://www.kevinsheppard.com/

4.1 Raw data

Floating African currencies, the Euro-fixed CFA Franc (XAF) and Euro in USD denomination was tested between March 8 2000 and March 6 2015. CFA Franc (XAF) followed strictly the euro only, due to its fixed regime, showing an appreciation against US dollar during the entire time set on Figure 1. Kenyan Shilling (KES) and South African Rand (ZAR) presented an appreciating trend before the subprime crisis only, otherwise the entire currency set suffered from depreciation.

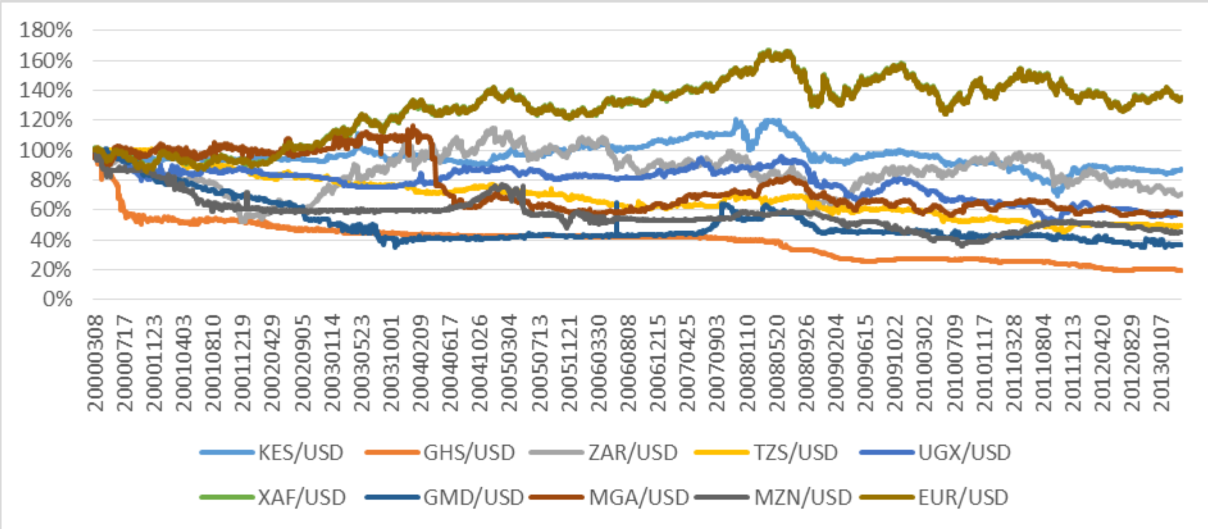

Figure 1: Developments of Selected African Currencies between 2000 and 2015 (March 8 2000=100%). Source: Bloomberg

Logarithmic returns of raw data had zero mean and low standard deviation, while symmetry appeared only for EUR and GMD (table 1). Excess kurtosis presented higher-than-expected occurrence for extreme fluctuation – pegged XAF and EUR presented a moderated level only. None of the currencies followed normal distribution and most of the data suffered from autocorrelation (except EUR) and heteroscedasticity (except KES, ZAR and EUR) at 2 lags. The entire sample was weak stationary.

Table 1: Descriptive statistics of the raw data. Source: author's calculation, using Kevin Sheppard's MFE Toolbox for Matlab

					normal distribution	autocorrelation	heteroscedasticity	stationarity
currency	mean	std	skewness	kurtosis	Jarque-Bera (p)	Ljung-Box (p)	ARCH-LM (p)	ADF (p)
KES/USD	0,00	0,01	-0,32	20,51	0,00	0,00	0,10**	0,00
GHS/USD	0,00	0,01	-1,23	33,77	0,00	0,00	0,00	0,00
ZAR/USD	0,00	0,01	-1,05	15,74	0,00	0,01	0,14**	0,00

TZS/USD	0,00	0,01	0,82	39,78	0,00	0,00	0,00	0,00
UGX/USD	0,00	0,01	-0,47	16,76	0,00	0,00	0,02	0,00
XAF/USD	0,00	0,01	0,13	5,14	0,00	0,00	0,00	0,00
GMD/USD	0,00	0,02	0,05	169,41	0,00	0,00	0,03	0,00
MGA/USD	0,00	0,01	-1,63	54,41	0,00	0,00	0,00	0,00
MZN/USD	0,00	0,01	-0,80	42,11	0,00	0,00	0,00	0,00
EUR/USD	0,00	0,01	-0,02	4,39	0,00	0,52*	0,59**	0,00

Notes: *: non autocorrelated at 2 lags, **: homoscedastic at 2 lags

Four (GARCH, TARCH, GJR-GARCH, APARCH) GARCH models with 13 different lag-number setup was fitted on the raw dataset to find the most fitting model trough searching for the lower BIC. Half of the sample preferred asymmetric models (except GHS, TZS, UGX, EUR), but previous volatilities had major role in the estimation of present volatility – innovations were important at GMD only (table 2).

Table 2: GARCH models of raw data. Source: author's calculation, using Kevin Sheppard's UCSD Toolbox for Matlab

currency	model	constant	alpha 1	alpha 2	gamma	beta 1	beta 2	delta	BIC
KES/USD	TARCH(1,1,2)	0,00	0,26		-0,03	0,46	0,29		-4,20
GHS/USD	GARCH(1,2)	0,00	0,11			0,37	0,51		-3,70
ZAR/USD	GJR GARCH(1,1,1)	0,00	0,12		-0,10	0,93			-3,21
TZS/USD	GARCH(1,1)	0,00	0,21			0,79			-3,91
UGX/USD	GARCH(1,1)	0,00	0,20			0,80			-3,82
XAF/USD*									
GMD/USD	TARCH(2,1,1)	0,00	0,30	0,30	-0,16	0,48			-2,81
MGA/USD	GJRGARCH(1,1,2)	0,00	0,02		0,03	0,47	0,49		-3,18
MZN/USD	GJRGARCH(1,1,2)	0,00	0,26		-0,11	0,29	0,51		-3,32

EUR/USD	GARCH(1,1)	0,00	0,04			0,96			-3,71

Notes: *: none of the models were able to provide homoscedastic residuals with normal distribution

Value-at-Risk (table 3) was able to create a close-to-symmetric set of non-extreme returns, while kurtosis dropped under 5. Extreme fluctuations had lower weight than 10% (except the 11% of XAF and EUR), so the method was able to capture those rare cases, which were responsible for most of the fat tailness of the data.

Table 3: Value-at-Risk properties of raw data (in USD). Source: author's calculation, following Madura (2008)

currency	KES	GHS	ZAR	TZS	UGX	XAF	GMD	MGA	MZN	EUR
mean	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
standard deviation	0,00	0,01	0,01	0,00	0,00	0,00	0,01	0,01	0,01	0,00
skewness	0,07	-0,14	-0,04	-0,02	-0,02	0,02	-0,04	0,04	0,05	0,01
kurtosis	3,48	4,34	2,51	3,16	2,84	2,52	4,07	3,68	4,24	2,48
X- treshold	-0,01	-0,02	-0,02	-0,01	-0,01	-0,01	-0,03	-0,02	-0,02	-0,01
X+ treshold	0,01	0,02	0,02	0,01	0,01	0,01	0,03	0,02	0,02	0,01
Number of X-	148	121	179	136	149	190	105	100	123	189
Number of X+	138	128	145	121	127	178	95	90	107	207
Number of non-extreme returns	3606	3573	3589	3578	3563	3455	3589	3424	3598	3517

4.2 Comparison of Methods

Temporary distribution of missing data (number of days between two missing data after each other) was not random, so the MCAR hypothesis can be rejected, we can assume MAR case only (table 4). MGA currency suffered from missingness at most, while ZAR and EUR had none of them. Listwise deletion was the most restrictive approach, while other two methods made less dramatic reduction in the length of the entire dataset.

Table 4: Number of missing days to data length under different approaches. Source: author's calculation

	Percentage of missing			normal distribution of missing days
Currency	Listwise deletion	Mean substitution	EM	Kolmogorov-Smirmov-test (p)
KES/USD	1%	1%	1%	0,00
GHS/USD	3%	2%	2%	0,00
ZAR/USD	0%	0%	0%	NaN

TZS/USD	2%	2%	2%	0,00
UGX/USD	2%	2%	2%	0,00
XAF/USD	3%	2%	2%	0,00
GMD/USD	4%	3%	3%	0,00
MGA/USD	9%	8%	8%	0,00
MZN/USD	2%	2%	2%	0,00
EUR/USD	0%	0%	0%	NaN

Appendix 1 contains the comparison of the descriptive statistics of reined data by three different approaches. Mean remained close to zero as at the raw data, but standard deviation doubled or tripled in 60% of the cases at EM method. Asymmetry of the data was completely distorted by all of the methods, but kurtosis was increased in the 40% of the cases and remained at the previous level in the other 40% at listwise deletion. Kurtosis increased in the half of the cases under mean substitution or remained stable. Kurtosis dramatically increased under EM method. Data remained non-normal distributed and weak stationer, and there were no significant changes in the autocorrelation of heteroscedasticity properties.

Results about Value-at-Risk has the same message as kurtosis (Appendix 2), where EM approach provided less VaR signals, but the "non-extreme" subset suffered from the increase of kurtosis in 80% of the sample (except ZAR and EUR). It means that VaR-based risk management can be biased by missing data if it is managed trough EM methodology. Listwise approach presented the lower impact on VaR properties.

Listwise approach in volatility modelling (Appendix 3) had a moderate impact on parameters only and suggested a different model for MGA and MZN, while it was now possible to fit a GARCH model on XAF data. Innovation parameters increased while previous volatility decreased a bit, and models presented a better fit – despite the expected higher fragmentation of the approach. Mean substitution pushed MGN and GHS currencies towards the more complicated APARCH model, but only GHS lost its former symmetric design. This approach increased the parameters of volatility persistence with similar BIC. The EM approach suggested asymmetric models instead of former symmetric models (for KES, GHS, TZS, UGX), while three former asymmetric preference decreased to symmetry (GMD, MGA, MZN) – but BIC increased almost everywhere, suggesting that it was harder to find well-fitting models with homoscedastic residuals. ZAR and EUR was completely unaffected by the different approaches (despite that they had to lost the most value to meet listwise deletion standards), while MGA and MZN was completely the subject of missing data management.

Despite significant share of the European Union and the US in sample countries balance of payments, unconditional correlation was mediocre between EUR/USD and ZAR/USD as well as high between XAF/USD and EUR/USD in the selected time period under listwise deletion and mean substitution (Appendix 4). Other markets remained isolated from international currency fluctuations, while ZAR correlated with XAF as well. Results were augmented by the EM method, signing medium common movement ($\rho = 0.64$) between GHS and UGX and mediocre levels ($0.3< \rho <0.5$) among GHS and KES, TZS, GMD, MGA, MZN or between UGX and KES, TZS and UGX, TZS and MZN, GMD and MZN. Due to the former results about the biased results of the EM procedure this sudden appearance of mediocre correlations is hardly questionable. Sample currencies were tested by Dynamic Conditional Correlation (DCC) against EUR/USD were only ZAR and XAF presented some sort of common movement, like it happened before at the unconditional case. Listwise deletion provided higher mean in both cases, but DCC of ZAR/USD-EUR/USD was significant different from Mean substitution or EM counter pairs only – according to the two sampled t-test (where p=0.55 and p=0.00 for all other cases).

Table 5: Dynamic Conditional Correlation moments – with EUR/USD. Source: author's calculation

	Listwise deletion				Mean substitution				EM			
	mean	std	skewness	kurtosis	mean	std	skewness	kurtosis	mean	std	skewness	kurtosis
KES/USD	0,0451	0,0489	0,1880	2,3352	0,0357	0,0459	0,5519	2,6715	0,0147	0,0354	-1,2180	3,4391
GHS/USD	-0,0010	0,0166	0,0417	6,9048	-0,0055	0,0094	0,1457	8,0598	0,0294	0,0157	0,7023	2,2311
ZAR/USD	0,4290	0,1775	-0,6407	3,4664	0,4119	0,1711	-0,4950	2,9529	0,4119	0,1711	-0,4950	2,9529
TZS/USD	-0,0037	0,0216	0,4166	10,2784	-0,0351	0,0088	-0,3345	1,9282	-0,0059	0,0023	-13,0668	201,9501
UGX/USD	0,0384	0,0214	0,2245	4,8841	0,0569	0,0167	0,2014	1,7215	0,0343	0,0228	0,4147	3,8064
XAF/USD	0,8373	0,1755	-1,3723	4,6628	0,7686	0,2313	-1,4152	5,3184	0,7642	0,2403	-1,4034	5,0187
GMD/USD	-0,0321	0,0151	-0,5513	2,1798	-0,0296	0,0156	-0,5877	2,2286	0,0060	0,0073	0,7218	20,5067
MGA/USD	0,0634	0,0322	-0,1573	2,7352	0,0531	0,0214	-0,0327	2,5063	0,0558	0,0117	0,5373	2,1534
MZN/USD	-0,0054	0,0191	-0,1388	2,1902	-0,0055	0,0390	-0,8071	2,6819	-0,0187	0,0350	-0,9365	3,1454

5 CONCLUSION

The maximum-likelihood-based Expectation Maximization (EM) models have high popularity nowadays to manage missing data in query data due to its ability to maintain the covariance matrix of the data. However, compared to listwise deletion or mean substitution methods, the EM method presented many biases on daily closing data of financial time series. This application increased the second and fourth moment dramatically, providing a reduced VaR-signal performance and made conditional volatility more asymmetric. Contagion studies can be also biased by the choice of method, but results were limited here.

Results of application comparison in current paper suggest the usage of mean substitution or listwise deletion for daily financial time series due to their tendency to have a close-to-zero first momentum.

References

Baraldi, P., Di Maio, F., Genini, D., Zio, E. (2015). Reconstruction of missing data in multidimensional time series by fuzzy similarity. *Applied Soft Computing Journal*, 26: 1-9

Cappeiello, L., Engle, R. F., Sheppard, K., (2006). Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns. *Journal of Financial Econometrics*, 4: 537-572

Ceylan, Y., Sipan, A., Cem, I., Inci, B. (2013). Comparison of missing value imputation methods in time series: the case of Turkish meteorological data. *Theoretical & Applied Climatology*, 112(1/2): 143-167

Engle, R. F., (2002). Dynamic Conditional Correlation - A Simple Class of Multivariate GARCH Models. *Journal of Business and Economic Statistics*, 20(3): 377-389

Graham, J. W. (2012). Missing Data Analysis and Design. Springer New York Heidelberg Dordrecht London

Houari, R., Bounceur, A., Kechadi, T., Euler, R. (2013). *A New Method for Estimation of Missing Data Based on Sampling Methods for Data Mining*. CCSEIT

Juan Carlos, F. G., Dusko, K., Bello, L., Amilcar, C. (2010). An Evolutionary Approach for Imputing Missing Data in Time Series. *Journal of Circuits, Systems & Computers*, 19(1): 107-121

Junger, W.L., Ponce de Leon, A. (2015). Imputation of missing data in time series for air pollutants. *Atmospheric Environment*, 102: 96-104

Kang, H. (2013). The prevention and handling of the missing data. *Korean J Anesthesiol*, 64(5): 402-406

Keim, D. B., Stambaugh, R. F. (1984). A further investigation of the weekend effect in stock returns. *The Journal of Finance*, 39(3): 819-835

Madura, J. (2008). *International Financial Management*. Mason: Thomson

Robins, R. P., Smith, G. P. (2015). *No More Weekend Effect*. Critical Finance Review, Forthcoming.

Ruud, P. A. (1991). Extensions of estimation methods using the EM algorithm. *Journal of Econometrics*, 49(3): 305-341

Shahid, M. N., Mehmood, Z. (2015). Calendar Anomalies in Stock Market: A Case of KSE 100 Index. *International Journal of African and Asian Studies*, 7: 16-23

Schneider, T. (2001). Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *Journal of Climate*, 14: 853-871.

Wothke, W. (1998). *Longitudinal and multi-group modelling with missing data*. Mahwah. NJ: Lawrence Erlbaum Associates.

Appendix 1 – Descriptive statistics of refined data

	currency	mean	std	skewness	kurtosis	normal distribution Jarque-Bera (p)	autocorrelation Ljung-Box (p)	heteroscedasticity ARCH-LM (p)	stationarity ADF (p)
Listwise deletion	KES/USD	0,00	0,01	0,28	18,60	0,00	0,00	0,08	0,00
	GHS/USD	0,00	0,01	1,78	39,88	0,00	0,00	0,00	0,00
	ZAR/USD	0,00	0,01	1,07	17,89	0,00	0,22	0,50	0,00
	TZS/USD	0,00	0,01	-0,87	30,73	0,00	0,00	0,00	0,00
	UGX/USD	0,00	0,01	0,46	16,63	0,00	0,00	0,07	0,00
	XAF/USD	0,00	0,01	-0,06	5,08	0,00	0,00	0,00	0,00
	GMD/USD	0,00	0,02	-0,03	169,73	0,00	0,00	0,03	0,00
	MGA/USD	0,00	0,01	1,77	58,07	0,00	0,00	0,00	0,00
	MZN/USD	0,00	0,01	0,92	49,84	0,00	0,00	0,00	0,00
	EUR/USD	0,00	0,01	-0,05	4,59	0,00	0,83	0,86	0,00
Mean substitution	KES/USD	0,00	0,01	0,32	20,60	0,00	0,00	0,22	0,00
	GHS/USD	0,00	0,01	1,25	34,57	0,00	0,00	0,00	0,00
	ZAR/USD	0,00	0,01	1,05	15,74	0,00	0,01	0,14	0,00
	TZS/USD	0,00	0,01	-0,84	40,52	0,00	0,00	0,00	0,00

	UGX/USD	0,00	0,01	0,46	16,99	0,00	0,00	0,02	0,00
	XAF/USD	0,00	0,01	-0,14	5,26	0,00	0,00	0,00	0,00
	GMD/USD	0,00	0,02	-0,05	174,91	0,00	0,00	0,03	0,00
	MGA/USD	0,00	0,01	1,61	57,71	0,00	0,00	0,00	0,00
	MZN/USD	0,00	0,01	0,81	43,03	0,00	0,00	0,00	0,00
	EUR/USD	0,00	0,01	-0,02	4,39	0,00	0,51	0,59	0,00
EM	KES/USD	0,00	0,01	0,98	181,04	0,00	0,00	0,02	0,00
	GHS/USD	0,00	0,03	-0,02	358,51	0,00	0,00	0,06	0,00
	ZAR/USD	0,00	0,01	1,05	15,74	0,00	0,01	0,14	0,00
	TZS/USD	0,00	0,02	0,04	198,98	0,00	0,00	0,00	0,00
	UGX/USD	0,00	0,02	0,03	121,75	0,00	0,00	0,00	0,00
	XAF/USD	0,00	0,01	-0,21	13,08	0,00	0,00	0,00	0,00
	GMD/USD	0,00	0,03	-0,07	74,16	0,00	0,00	0,00	0,00
	MGA/USD	0,00	0,04	0,11	24,16	0,00	0,00	0,00	0,00
	MZN/USD	0,00	0,02	0,14	55,40	0,00	0,00	0,00	0,00
	EUR/USD	0,00	0,01	-0,02	4,39	0,00	0,51	0,59	0,00

Appendix 2 – Value-at-Risk of refined data (in USD)

	currency	KES	GHS	ZAR	TZS	UGX	XAF	GMD	MGA	MZN	EUR
Listwise deletion	mean	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	standard deviation	0,00	0,01	0,01	0,00	0,00	0,01	0,01	0,01	0,01	0,00
	skewness	-0,05	0,09	0,06	0,06	0,00	-0,02	0,02	-0,02	-0,02	0,00
	kurtosis	3,49	4,31	2,54	3,03	2,87	2,52	4,05	3,69	4,02	2,50
	X- threshold	-0,01	-0,02	-0,02	-0,01	-0,01	-0,01	-0,03	-0,02	-0,02	-0,01
	X+ threshold	0,01	0,02	0,02	0,01	0,01	0,01	0,03	0,02	0,02	0,01
	No X-	130	111	119	119	115	159	84	87	82	160
	No X+	136	105	153	135	130	169	97	90	91	173
	No Normal	3146	3196	3140	3158	3167	3084	3231	3235	3239	3079
Mean substitution	mean	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	standard deviation	0,00	0,01	0,01	0,00	0,00	0,00	0,01	0,01	0,01	0,00
	skewness	-0,07	0,13	0,04	0,01	0,00	-0,03	0,05	-0,10	-0,02	0,00
	kurtosis	3,51	4,47	2,51	3,25	2,88	2,59	4,37	3,95	4,38	2,48
	X- threshold	-0,01	-0,02	-0,02	-0,01	-0,01	-0,01	-0,03	-0,02	-0,02	-0,01
	X+ threshold	0,01	0,02	0,02	0,01	0,01	0,01	0,03	0,02	0,02	0,01
	No X-	138	128	145	121	129	188	97	102	111	189
	No X+	148	124	179	135	149	192	107	105	123	207
	No Normal	3626	3660	3588	3656	3634	3532	3708	3705	3678	3516
EM	mean	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	standard deviation	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,00
	skewness	-0,05	0,04	0,04	-0,16	-0,03	-0,02	0,08	-0,20	-0,05	0,00
	kurtosis	5,58	9,62	2,51	7,01	5,63	2,66	5,30	10,66	5,11	2,48
	X- threshold	-0,02	-0,05	-0,02	-0,03	-0,03	-0,01	-0,06	-0,07	-0,04	-0,01

	X+ threshold	0,02	0,06	0,02	0,03	0,03	0,01	0,06	0,07	0,04	0,01
	No X-	45	26	145	38	56	144	76	127	94	189
	No X+	50	30	179	41	67	159	84	129	113	207
	No Normal	3817	3856	3588	3833	3789	3609	3752	3656	3705	3516

Appendix 3 – GARCH models of refined data

	currency	model	constant	alpha 1	alpha 2	gamma	beta 1	beta 2	delta	BIC
Listwise deletion	KES/USD	TARCH(1,1,2)	0,00	0,23		0,04	0,40	0,36		-4,12
	GHS/USD	GARCH(1,2)	0,00	0,13			0,35	0,52		-3,68
	ZAR/USD	TARCH(1,1,1)	0,00	0,05		0,07	0,91			-3,13
	TZS/USD	GARCH(1,1)	0,00	0,25			0,75			-3,96
	UGX/USD	GARCH(1,1)	0,00	0,21			0,79			-3,80
	XAF/USD	GARCH(1,1)	0,00	0,04			0,95			-3,57
	GMD/USD	TARCH(2,1,1)	0,01	0,15	0,38	0,10	0,42			-2,76
	MGA/USD	APARCH(1,1,1)	0,00	0,02		0,05	0,94		3,98	-3,22
	MZN/USD	GARCH(1,2)	0,00	0,23			0,30	0,47		-3,36
	EUR/USD	GARCH(1,1)	0,00	0,05			0,95			-3,63
Mean substitution	KES/USD	TARCH(1,1,2)	0,00	0,23		0,03	0,44	0,32		-4,20
	GHS/USD	APARCH(1,1,1)	0,00	0,05		-0,01	0,89		3,66	-3,69
	ZAR/USD	GJRGARCH(1,1,1)	0,00	0,02		0,10	0,93			-3,21
	TZS/USD	GARCH(1,1)	0,00	0,20			0,80			-3,92
	UGX/USD	GARCH(1,1)	0,00	0,19			0,81			-3,83
	XAF/USD	GARCH(1,2)	0,00	0,06			0,02	0,91		-3,62
	GMD/USD	TARCH(2,1,1)	0,00	0,14	0,28	0,17	0,49			-2,83
	MGA/USD	APARCH(1,1,1)	0,00	0,02		0,07	0,95		3,44	-3,22

	MZN/USD	TARCH(1,1,2)	0,00	0,14		0,12	0,29	0,52		-3,32
	EUR/USD	GARCH(1,1)	0,00	0,04			0,96			-3,71
	KES/USD	GJRGARCH(1,1,2)	0,00	0,09		-0,05	0,00	0,94		-3,72
	GHS/USD	TARCH(1,1,2)	0,00	0,31		-0,10	0,00	0,74		-2,44
	ZAR/USD	GJRGARCH(1,1,1)	0,00	0,02		0,10	0,93			-3,21
	TZS/USD	TARCH(1,1,2)	0,00	0,72		-0,43	0,30	0,20		-3,18
	UGX/USD	APARCH(1,1,1)	0,00	0,16		-0,19	0,34		4,00	-3,04
	XAF/USD	none of the models was able to provide homoscedastic residuals								
	GMD/USD	GARCH(1,1)	0,00	0,34			0,50			-2,28
	MGA/USD	GARCH(1,2)	0,00	0,30			0,38	0,32		-2,35
	MZN/USD	GARCH(1,1)	0,00	0,40			0,55			-2,75
EM	EUR/USD	GARCH(1,1)	0,00	0,04			0,96			-3,71

Appendix 4 – Unconditional correlations – calculated on homoscedastic GARCH residuals

	KES/USD	GHS/USD	ZAR/USD	TZS/USD	UGX/USD	XAF/USD	GMD/USD	MGA/USD	MZN/USD	EUR/USD	
KES/USD	0,00	0,00	0,05	0,04	0,09	0,06	0,00	0,05	0,00	0,05	
GHS/USD	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
ZAR/USD	0,05	0,00	0,00	0,00	0,00	0,38	0,00	0,00	0,00	0,43	
TZS/USD	0,04	0,00	0,00	0,00	0,08	0,00	0,00	0,00	0,04	0,00	
UGX/USD	0,09	0,00	0,00	0,08	0,00	0,04	0,05	0,00	0,00	0,04	
XAF/USD	0,06	0,00	0,38	0,00	0,04	0,00	0,00	0,07	0,00	0,83	Listwise deletion

GMD/USD	0,00	0,00	0,00	0,00	0,05	0,00	0,00	0,00	0,00	0,00	
MGA/USD	0,05	0,00	0,00	0,00	0,00	0,07	0,00	0,00	0,00	0,06	
MZN/USD	0,00	0,00	0,00	0,04	0,00	0,00	0,00	0,00	0,00	0,00	
EUR/USD	0,05	0,00	0,43	0,00	0,04	0,83	0,00	0,06	0,00	0,00	
KES/USD	0,00	0,00	0,03	0,05	0,08	0,06	0,00	0,00	0,00	0,04	
GHS/USD	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
ZAR/USD	0,03	0,00	0,00	0,00	0,00	0,35	0,00	0,00	0,00	0,41	
TZS/USD	0,05	0,00	0,00	0,00	0,08	0,00	0,00	0,00	0,00	0,00	
UGX/USD	0,08	0,00	0,00	0,08	0,00	0,04	0,04	0,00	-0,05	0,00	
XAF/USD	0,06	0,00	0,35	0,00	0,04	0,00	0,00	0,06	0,00	0,76	
GMD/USD	0,00	0,00	0,00	0,00	0,04	0,00	0,00	0,00	0,00	0,00	
MGA/USD	0,00	0,00	0,00	0,00	0,00	0,06	0,00	0,00	0,00	0,05	
MZN/USD	0,00	0,00	0,00	0,00	-0,05	0,00	0,00	0,00	0,00	0,00	
EUR/USD	0,04	0,00	0,41	0,00	0,00	0,76	0,00	0,05	0,00	0,00	Mean substitution
KES/USD	0,00	0,32	0,00	0,17	0,33	0,05	0,08	0,13	0,14	0,03	
GHS/USD	0,32	0,00	0,00	0,46	0,64	-0,03	0,33	0,38	0,46	0,00	
ZAR/USD	0,00	0,00	0,00	0,00	0,00	0,30	0,00	0,00	0,03	0,41	EM

TZS/USD	0,17	0,46	0,00	0,00	0,41	0,00	0,29	0,29	0,38	0,00	
UGX/USD	0,33	0,64	0,00	0,41	0,00	0,00	0,25	0,27	0,29	0,03	
XAF/USD	0,05	-0,03	0,30	0,00	0,00	0,00	0,00	0,04	-0,03	0,66	
GMD/USD	0,08	0,33	0,00	0,29	0,25	0,00	0,00	0,07	0,36	0,00	
MGA/USD	0,13	0,38	0,00	0,29	0,27	0,04	0,07	0,00	0,22	0,00	
MZN/USD	0,14	0,46	0,03	0,38	0,29	-0,03	0,36	0,22	0,00	0,00	
EUR/USD	0,03	0,00	0,41	0,00	0,03	0,66	0,00	0,00	0,00	0,00	

Notes: only significant ($p<0.05$) unconditional correlations

Contact information

Gábor Dávid Kiss, Ph.D.
University of Szeged, Faculty of Economics and Business Administration
6722 Szeged, Kálvária sgt. 1, Hungary
0036305010578
kiss.gabor.david@eco.u-szeged.hu
Marianna Sávai
University of Szeged, Faculty of Economics and Business Administration
6722 Szeged, Kálvária sgt. 1, Hungary
savai.marianna@eco.u-szeged.hu