STAVROS DEGIANNAKIS AND CHRISTOS FLOROS

MODELLING AND FORECASTING HIGH FREQUENCY FINANCIAL DATA

Modelling and Forecasting High Frequency Financial Data

Stavros Degiannakis and Christos Floros

palgrave
macmillan

First published 2015 by
PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries

ISBN: 978–1–137–39648–8

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

A catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data

Degiannakis, Stavros, author.
Modelling and forecasting high frequency financial data / Stavros Degiannakis, Christos Floros.
pages cm
ISBN 978–1–137–39648–8 (hardback)
1. Finance–Mathematical models. 2. Speculation–Mathematical models.
3. Technical analysis (Investment analysis)–Mathematical models.
I. Floros, C. (Christos), author. II. Title.

HG106.D44 2015
332.01′5195–dc23 2015013168

Modelling and Forecasting High Frequency Financial Data

To Aggelos, Andriana and Rebecca

Stavros Degiannakis

To Ioanna, Vasilis-Spyridon, Konstantina-Artemis and Christina-Ioanna

Christos Floros

List of Figures

List of Tables

Acknowledgments

Dr. Stavros Degiannakis and Dr. Christos Floros acknowledge the support from the European Community's Seventh Framework Programme (Marie Curie FP7-PEOPLE-IEF & FP7-PEOPLE-RG) funded under grant agreements no. PIEF-GA-2009-237022 & PERG08-GA-2010-276904. We would like to thank Dr. Enrique Salvador and Dr. Thomas Poufinas for their constructive contribution to the 7th chapter. We are also grateful to the many people (colleagues at the university, researchers, traders and financial market practitioners), who by occasional informal exchange of views have had an influence on these aspects as well. Most importantly, we wish to express our gratitude to our families for their support.

List of Symbols and Operators

$d(.)$	indicator function; i.e. $d(y_t > 0) = 1$ if $y_t > 0$, and $d(y_t > 0) = 0$ otherwise
r_f	risk-free rate
AIC	Akaike's information criterion
r_{annual}	annual risk-free interest rate
P_{ask,t_i}	ask price
γ_i	asymmetric parameter in ARCH model
δ	asymmetric power parameter in ARCH model
c_i	autoregressive coefficients
$\bar{\Psi}_{(.)}$	average of predictive loss/evaluation function, i.e. $\bar{\Psi}_{(MSE)} = \tilde{T}^{-1} \sum_{t=1}^{\tilde{T}} (RV_{t+1\|t}^{(\tau)} - RV_{t+1}^{(\tau)})^2$
μ_2	average of the daily realized variances
μ_1	average of the squared closed-to-open log-returns
P_{bid,t_i}	bid price
$BV_t(m)$	Bi-Power variation at m sampling frequency
CTS	calendar time sampling
w_i	coefficients of HAR model
$\mu_t^{(m_i)}$	conditional mean estimation of model m_i
μ_t	conditional mean
$\mu(.)$	conditional mean's functional form
σ_t	conditional standard deviation
$\hat{\sigma}_t^2$	conditional variance estimate
$g(.)$	conditional variance's functional form
CGR	Correlated Gamma Ratio distribution
ρ	correlation coefficient
dq_t	counting process
$\sigma_{m,i}$	covariance of market returns and asset
η_{12}	covariance of realized variance and squared closed-to-open log-returns
$F_{X_{(1)}}(x; a, C_{12...n})$	cumulative distribution function of $X_{(1)}$
$F_J^{*\,n}(t)$	cumulative distribution function of durations J

$F_{T_n}(t)$	cumulative distribution function of epoch T_n
$F(.)$	cumulative distribution function
$DM_{(.,.)}$	Diebold Mariano Statistic
$\Psi_t^{(.,.)}$	difference of loss/evaluation functions (evaluation differential)
$k^{(m_i)}$	dimension of vector of unknown parameters $\beta^{(m_i)}$
$J_i = T_i - T_{i-1}$	durations
$\{T_i\}_{i=1}^{M}$	epochs
τ	equidistance points (sub-intervals) in time
ω_1, ω_2	estimated parameters of $RV_{t(HL^*)}^{(\tau)}$
$\hat{\theta}^{(T)}$	estimator of θ based on a sample of size T
d'	exponent of the fractional differencing operator$(1-L)^{d'}$ in ARFIMA models
d	exponent of the fractional differencing operator$(1-L)^{d}$ in FIGARCH & FIAPARCH models
$\hat{y}_t^{(m)}$	forecasts of y_t from model m
$L_T(.)$	full sample log-likelihood function based on a sample of size T
$\Gamma(.)$	Gamma function
X_{m_i}	half sum of squared standardized one-step-ahead prediction errors of model m_i
HQ	Hannan and Quinn information criterion
$^c\inf$	infimum
I_t	information set
$\varepsilon_{t-i\|t}$	in-sample fitted error at time $t-i$ based on information available at time t
$y_{t-i\|t}$	in-sample fitted value of conditional mean at time $t-i$ based on information available at time t
$p(t)$	instantaneous (unobserved) asset price
$\log p(t)$	instantaneous logarithmic asset price
$\sigma(t)$	instantaneous variance of the rate of return
$\sigma_{[a,b]}^{2(IQ)}$	integrated quarticity
$\sigma_{[a,b]}^{2(IV)}$	integrated variance over the interval $[a,b]$
P_{lip,t_j}	interpolated price
Ku	kurtosis
$J_{t,\alpha}^{(LM)}$	L&M statistic
L	lag operator
y_{t_j}	log-return over the sub-interval $\left[t_j - t_{j-1}\right]$
y_t	log-returnsover the sub-interval $[t, t-1]$

$\Psi_t^{(.)}$	loss/evaluation function that measures the distance between volatility and its forecast
$\varepsilon_{t_j} = \log P_{t_j} - \log p_{t_j}$	market microstructure noise
$\mathbf{X}_t^{(m_i)}$	matrix of $x_t^{(m_i)}$ explanatory variables
$\sigma_{i,t}^{(MAD)}$	Median Absolute Deviation
MMG	Minimum Multivariate Gamma distribution
$X_{(1)}$	minimum value of X_{m_i}
$m_{(1)}$	model with the lowest value of X_{m_i}
d_i	moving average coefficients
$MSE^{(\tau)}$	MSE loss (or evaluation) function
$\widehat{T}$	number of forecasts for out-of-sample evaluation
n	number of models or variables
$\tilde{T}$	number of observations for out-of-sample forecasting
$\breve{T}$	number of observations for rolling sample
$\breve{\theta}$	number of parameters of vector θ
T	number of total observations
$\{P_{t_j}\}_{j=1}^{\tau}$	observed asset price
$RV_{t+1\|t}^{(\tau)}$	one-day-ahead realized variance at time $t+1$ based on information available at time t
$RV_{(un),t+1\|t}^{(\tau)}$	one-day-ahead realized variance at time $t+1$ based on information available at time t (unbiased estimator)
$y_{t+1\|t}$	one-step-ahead conditional mean at time $t+1$ based on information available at time t
$y_{t+1\|t}^{(m_i)}$	one-step-ahead conditional mean at time $t+1$ based on information available at time t of model m_i
$\sigma_{t+1\|t}^2$	one-step-ahead conditional variance at time $t+1$ based on information available at time t
$h_{t+1\|t}^2$	one-step-ahead estimate of integrated quarticity given the information available at time t
$\varepsilon_{t+1\|t}$	one-step-ahead prediction error at time $t+1$ based on information available at time t
$z_{t+1\|t}$	one-step-ahead standardized prediction error at time $t+1$based on information available at time t
q	order of ARCH form
p	order of GARCH form
k	order of the autoregressive model
l	order of the moving average model
$d_{i,t}$	outlyingness measure
$C(L)$	polynomial of autoregressive model - AR
$B(L)$	polynomial of FIGARCH & FIAPARCH models
$\Phi(L)$	polynomial of FIGARCH & FIAPARCH models
$D(L)$	polynomial of moving average model - MA

P_{pre,t_j}	previous tick price
$lRange_{[a,b]}$	price log-range
$c^E(t)$	price of the European call option at time t
$Range_{[4],[a,b]}$	price range, four-data-points
$Range_{[a,b]}$	price range
$f(.)$	probability density function
$QQ_t(m)$	Quad-power quarticity at m sampling frequency
$\mathfrak{R}$	
$RJ_t(m)$	realized jumps at m sampling frequency
$\mathbf{ROWQCov}_t$	Realized Outlyingness Weighted Covariation
$ROWQuarticity_t(m)$	Realized Outlyingness Weighted Quarticity at m sampling frequency
$ROWVar_t$	Realized Outlyingness Weighted Variance
$RV^{[2q]}_{[a,b]}$	realized power variation of order 2q
$\mathbf{RCov}_t$	realized quadratic covariation
$RV^{(\tau)}_t$	realized volatility at time t, divided in τ points in time
$RV_{[a,b]}$	realized volatility for the time interval $[a,b]$
$RV^{(\tau)}_{t(n)}$	realized volatility of n-trading-days
$RV^{(\tau)}_{t(FKO)}$	realized volatility with Fleming's et al dynamic scaling
$RV^{(\tau)}_{t(HL^*)}$	realized volatility with Hansen and Lunde's interday adjustment
$RV^{(\tau)}_{t(Martens)}$	realized volatility with Marten's interday adjustment
$\gamma_d(i)$	sample autocovariance of i^{th} order
$m = \frac{b-a}{\tau-1}$	sampling frequency
SBC	Schwarz information criterion
SH	Shibata information criterion
$f_d(0)$	spectral density at frequency zero
$N(.)$	standard normal density function
$z_t \sim N(0,1)$	standard normal distribution
$W(t)$	standard Wiener process
SPA	Superior Predictive Ability statistic
t_j	time index, $t_j \ni [a,b]$
$[a,b]$	time interval
$TQ_t(m)$	Tri-power quarticity at m sampling frequency
$z^{(m_i)}_t$	unpredictable component of model m_i
ε_t	unpredictable component
η_2	variance of the daily realized variances
σ^2_m	variance of the market
η_1	variance of the squared closed-to-open log-returns
$\mathbf{Y}_t$	vector of dependent variable y_t
θ	vector of estimated parameters for the conditional mean and variance

$\theta^{(t)}$	vector of estimated parameters for the conditional mean and variance at time t
$\mathbf{w}$	vector of estimated parameters for the density function f
$\mathbf{x}_{t-1}^{(m_i)}$	vector of explanatory variables of m_i regression model
β	vector of parameters for estimation in regression model
$\beta^{(m_i)}$	vector of parameters for estimation of m_i regression model
v_t	vector of predetermined variables included in I_t
$\Psi_t^{(ES)}$	loss/evaluation function for Expected Shortfall
$\Psi_t^{(VaR)}$	loss/evaluation function for VaR
$ES_t^{(a)}$	Expected Shortfall of a portfolio at confidence level a
$VaR_t^{(a)}$	Value-at-Risk of a portfolio at confidence level a
$r_{h,t}$	log-return of the hedged portfolio at time t
$\mathbf{R}_t$	conditional correlation matrix
$\mathbf{r}_t$	vector of returns
$\mathbf{z}_t$	vector of standardized error term (residuals)
ρ_{ij}	constant correlation of spot and future price returns
$\sigma_{f,t}^2$	variance of future price returns at time t
$\sigma_{s,t}\sigma_{f,t}$	covariance of spot and future price returns at time t
$\sigma_{s,t}^2$	variance of spot price returns at time t
$\Sigma_t(.)$	conditional variance-covariance matrix
ε_t	vector of error term (residuals)
$\mu_{\mathbf{t}}(.)$	vector of conditional mean
Δ	first order log-difference
Δ	log difference
R^2	coefficient of determination
$Cov(.,.)$	covariance
$MVHR$	minimum-variance hedge ratio
$diag(.)$	diagonal matrix