Forecasting Daily FTSE 100 Returns: A Box-Jenkins Analysis, 2017-2019

Abhav Singh 23b3035 Dhairya Kantawala 23b3321 Saubhagya Singh 23b3025
Aryan Mahajan 23b1076 Yash Palwe 23b1823 Shivam Singh 24b3311

Project report, Group A, EC 602 Time Series Econometrics, Instructor: Prof. Aditi Chaubal

1 Objectives

We apply the univariate Box-Jenkins methodology[1], in the notation of our course notes[2], to daily returns on the FTSE 100 index over 2017-2019. Our objectives are to:

(i) find the order of integration of the index and identify an ARIMA(p, d, q) model for the mean of daily returns;

(ii) test that model's errors for conditional heteroskedasticity and model the variance with the ARCH family, including the asymmetric GJR and EGARCH models and GARCH-in-mean;

(iii) forecast the mean and the variance one day ahead over 2019, a year held out from all estimation, and test the forecasts against simple benchmarks.

We follow the seven Box-Jenkins steps in order and check the key tests by Monte Carlo simulation at our sample size.

2 Data

The series is the daily close of the FTSE 100 price index, P_t, from Yahoo Finance[3] for 30 December 2016 to 31 December 2019. We checked the closes against official figures on two dates (differences of at most 0.05 index points) and against the iShares Core FTSE 100 ETF, whose daily returns have a correlation of 0.97 with ours. We model percentage log returns, $Z_t = 100(1 - B) \ln P_t$, 758 trading days in all. Identification, estimation and selection use 2017-2018 only ($T = 505$). The 253 days of 2019 are used only to evaluate forecasts.

Table 1. Descriptive statistics of daily returns Z_t (%)

Sample	N	Mean	Std. dev.	Ann. vol.	Skewness	Kurtosis	Jarque-Bera
2017	252	0.029	0.539	8.6%	−0.30	5.11	50.3***
2018	253	−0.053	0.801	12.7%	−0.30	4.04	15.3***
2017-2018 (estimation)	505	−0.012	0.684	10.9%	−0.38	4.85	84.1***
2019 (forecast period)	253	0.045	0.740	11.7%	−0.44	5.17	57.9***

Annualised volatility is $\sqrt{252}$ times the daily standard deviation. ***, **, *: significant at 1%, 5% and 10% in all tables.

Figure 1 shows the level and the returns. The index rose to a record 7877.5 on 22 May 2018 and then fell 16.4% by 27 December 2018 in the global sell-off of the fourth quarter. Large moves cluster around dated news, such as the snap-election call of April 2017, the volatility spike of February 2018 and the US-China tariff escalation of August 2019. Volatility rose from 8.6% a year in 2017 to 12.7% in 2018, and a Levene test rejects equal variances in the two years ($p < 0.001$). The mean return is -0.012% a day with a standard error of 0.030, so it cannot be told apart from zero.

Returns are fat-tailed and skewed to the left, and the Jarque-Bera test rejects normality in every year (Table 1). Seven training days lie more than three standard deviations from the mean, against 1.4 expected under normality. A Student-t distribution with 5.3 degrees of freedom fits the tails that the normal misses, and it improves the log-likelihood by 14.7 for one extra parameter.

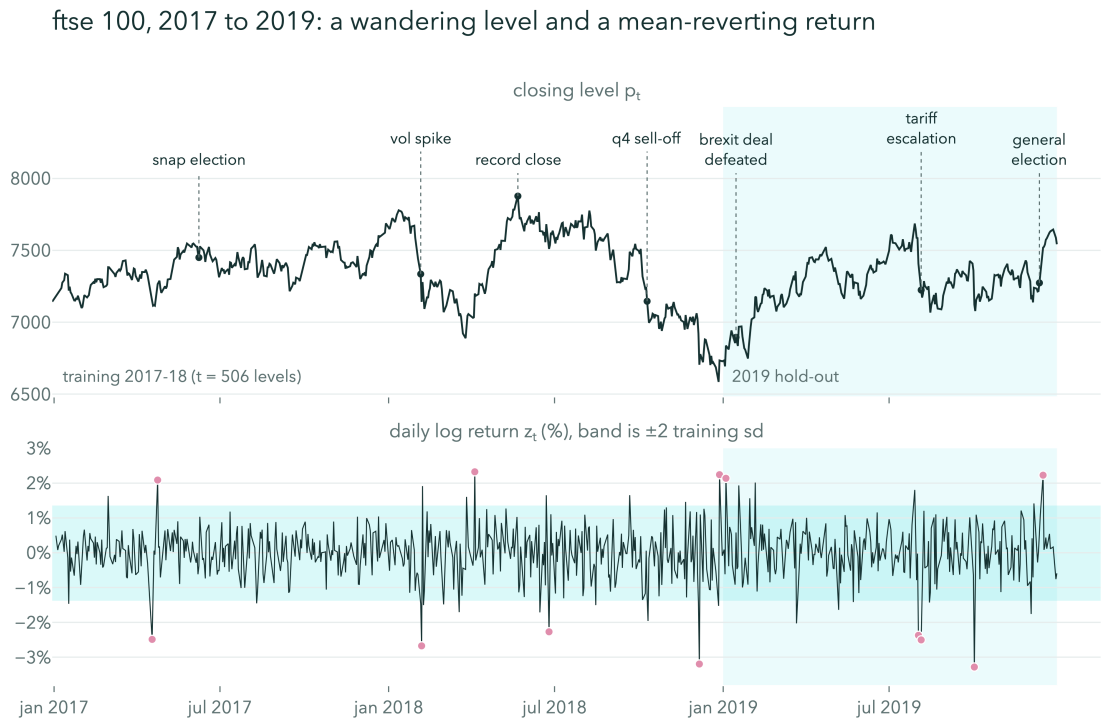


Figure 1. FTSE 100 closing level (top) and daily log returns with a band of ±2 estimation-sample standard deviations (bottom). Shaded: 2019 forecast period.

3 Results and analysis

3.1 Stationarity (step 3)

Neither the augmented Dickey-Fuller (ADF) nor the Phillips-Perron (PP) test rejects a unit root in $\ln P_t$ in any of the three Dickey-Fuller cases, and both reject very strongly for Z_t (Table 2). We therefore set $d = 1$. The ADF lag length was chosen by SBC up to Schwert's maximum of 17 and came out at zero for every series. The mean daily change in $\ln P_t$ is not significant ($t = -0.39$), so the drift case without a trend is the relevant one.

Table 2. Unit root and stationarity tests, 2017-2018

	ADF		Phillips-Perron		KPSS	
Series	Drift	Trend	Drift	Trend	Level	Trend
P_t	−1.73	−1.87	−1.89	−1.96	0.265	0.255***
$\ln P_t$	−1.68	−1.84	−1.82	−1.91	0.272	0.257***
Z_t	−23.39***	−23.45***	−23.39***	−23.49***	0.225	0.049
5% critical value	−2.87	−3.42	−2.87	−3.42	0.461	0.148

ADF and PP test the null of a unit root, KPSS the null of stationarity. ADF lags chosen by SBC. The no-constant case agrees ($\ln P_t$: −0.40; Z_t: −23.40***).

The KPSS test rejects stationarity of $\ln P_t$ around a trend but not around a constant, so the evidence is not one-sided. We therefore checked the ADF test by simulation (Figure 2). With $T = 506$ and normal, Student-t or GARCH-t shocks matched to the data, the simulated 5% critical value is -2.84 to -2.92 against the tabulated -2.87, and the size stays between 4.6% and 5.7%. Fat tails and volatility clustering therefore do not distort the test. Its power is the problem. The drift-case regression implies a root of 0.984 for $\ln P_t$, and if the level were a stationary AR(1) with that root, the ADF test would reject in only 26% of samples. A unit root in the log index rests on the tests together with the economic prior of weak-form efficiency.[4] Over-differencing a root of 0.98 costs little in one-step forecasts, while treating

a random walk as stationary would build false mean reversion into them.

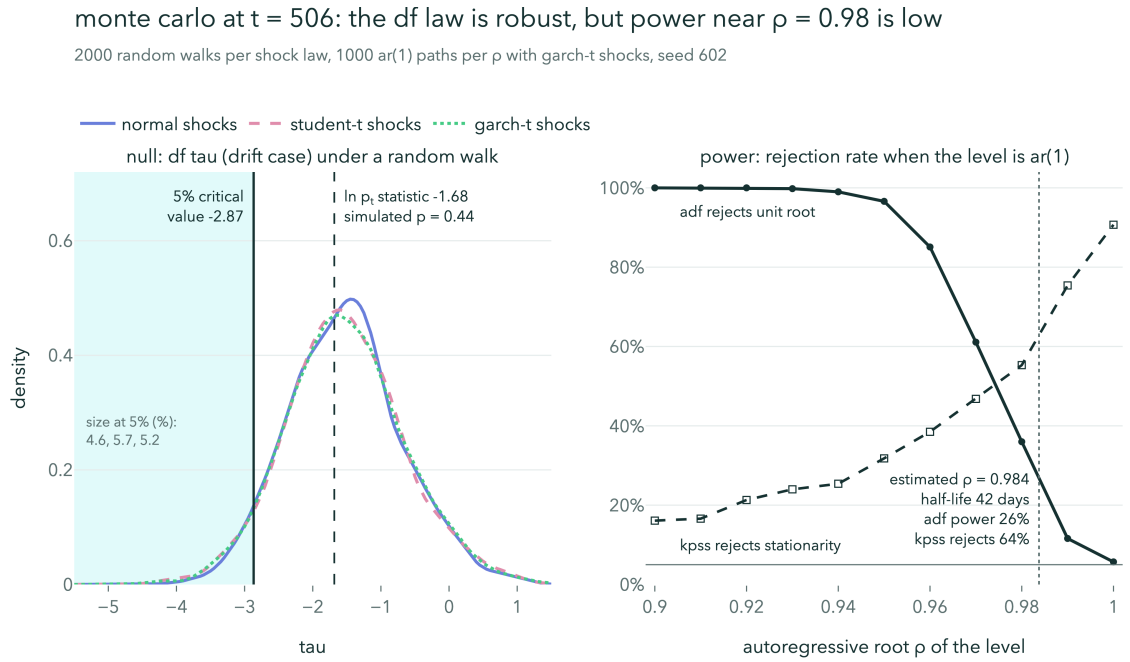


Figure 2. Monte Carlo study of the ADF test at $T = 506$. Left: simulated null distribution of the drift-case statistic for three shock laws, with our statistic for $\ln P_t$. Right: rejection rate when the level is a stationary AR(1) with root ρ (GARCH-t shocks), and the KPSS rejection rate.

3.2 Identification (step 2)

No sample autocorrelation or partial autocorrelation of Z_t up to lag 20 leaves the Bartlett 95% band of ± 0.087 (Figure 3). Since squared returns are autocorrelated, we also drew a heteroskedasticity-robust band, which is wider at the lags where the ARCH effects sit (1, 4 and 14) and no wider elsewhere; no lag leaves it either. Only lags 4 and 5 touch the narrower $\pm 1/\sqrt{N}$ band, and under pure noise about six of twenty lags would. The Ljung-Box statistic is $Q(10) = 4.88$ ($p = 0.90$) and its smallest p-value over $m = 1, \ldots, 20$ is 0.34. The variance ratios at 2, 5, 10 and 20 days (0.96, 0.93, 0.97 and 0.97) are all within their confidence intervals of one. Neither the ACF nor the PACF cuts off, both are flat, and the identification table points to white noise, with AR(1) and MA(1) as the smallest alternatives to test.

3.3 Estimation and selection (steps 4 and 5)

To check the identification, we estimated every ARMA(p, q) with $p, q \leq 3$ by conditional least squares on a common sample ($T = 502$) and compared $\mathrm{AIC} = T \ln \mathrm{SSR} + 2n$ with $\mathrm{SBC} = T \ln \mathrm{SSR} + n \ln T$. Table 3 lists the main candidates in the course sign convention, $\phi(B) Z_t = c + \theta(B) \varepsilon_t$ with $\phi(B) = 1 - \phi_1 B - \ldots$ and $\theta(B) = 1 - \theta_1 B - \ldots$. No AR or MA coefficient of the small models is significant, and the constant is not either.

SBC selects white noise, and every other model is at least 6 points worse (Figure 4). AIC prefers ARMA(3,3) by 2.0 points over white noise. The ARMA(3,3) reduces the SSR by 3.1%, which lowers $T \ln \mathrm{SSR}$ by 16.0. AIC charges 14 for its seven parameters and SBC charges 43.5. The four models that beat white noise on AIC all have a root of modulus below 1.05.

The roots of the ARMA(3,3) show why it is not a real improvement. Its inverse AR roots (0.979, -0.956, -0.403) almost coincide with its inverse MA roots (1.000, -1.000, -0.352), so $\phi(B)$ and $\theta(B)$ share common factors and nearly cancel, leaving $Z_t \approx \varepsilon_t$. Two of the MA roots also lie on the unit circle, so the fitted model is not invertible. Exact maximum likelihood finds a different pair of cancelling factors, and under it AIC selects white noise as well.

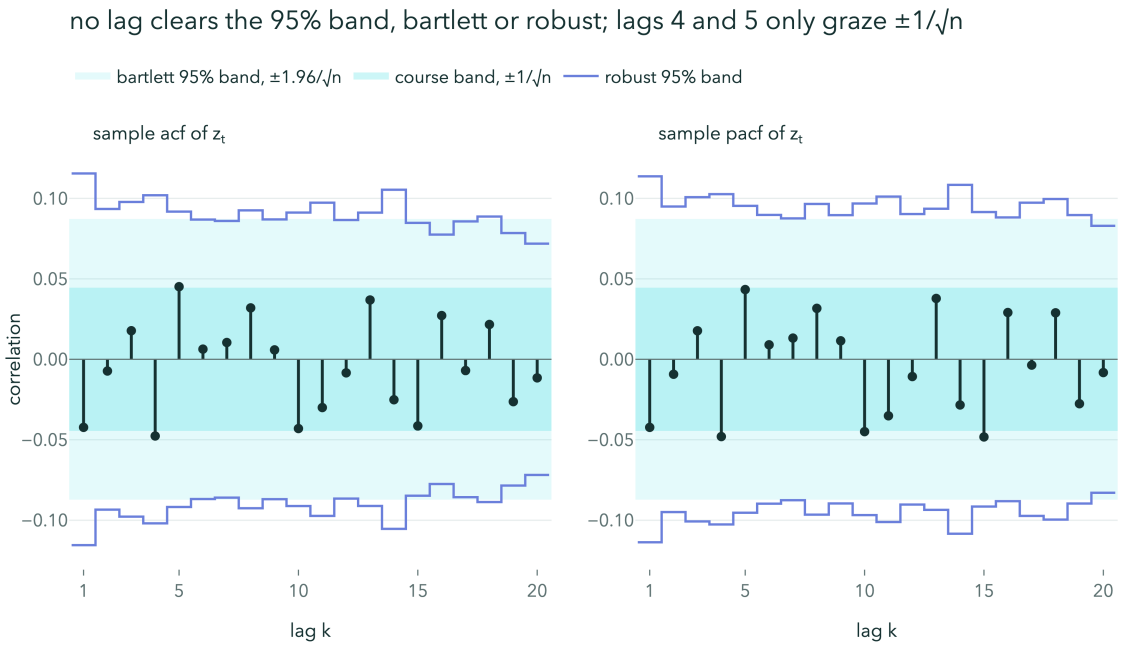


Figure 3. Sample ACF and PACF of Z_t, 2017-2018, with the $\pm 1/\sqrt{N}$ band, the Bartlett 95% band and a heteroskedasticity-robust 95% band.

Table 3. Conditional least squares estimates of candidate mean models, 2017-2018 ($T = 502$)

Model	Coefficients (standard errors)	SSR	AIC	SBC
White noise	none	235.32	2741.4	**2741.4**
Constant	$c = -0.013$ (0.031)	235.23	2743.2	2747.4
AR(1)	$c = -0.014$ (0.031), $\phi_1 = -0.043$ (0.045)	234.80	2744.3	2752.7
MA(1)	$c = -0.013$ (0.029), $\theta_1 = 0.044$ (0.045)	234.79	2744.3	2752.7
ARMA(1,1)	$c = -0.011$ (0.026), $\phi_1 = 0.224$ (0.965), $\theta_1 = 0.266$ (0.955)	234.77	2746.2	2758.9
ARMA(2,2)	$\phi_1 = 0.012$, $\phi_2 = 0.951^{***}$, $\theta_1 = 0.004$, $\theta_2 = 0.996^{***}$	229.92	2739.8	2760.8
ARMA(3,3)	$\phi = (-0.381^{**}, 0.945^{***}, 0.378^{**})$, $\theta = (-0.352^{**}, 1.000^{***}, 0.352^{*})$	227.94	**2739.4**	2768.9

The constant of ARMA(2,2) and ARMA(3,3) is below 0.002 in absolute value and is not shown. Bold: minimum of each criterion over all 17 models. The full grid is in our command files.

A Monte Carlo study confirms that the AIC gain is what overfitting noise looks like. We drew 500 white-noise samples of 505 days, either i.i.d. normal or with the volatility model estimated in step 6, and refitted the whole grid on each. AIC picks a model other than white noise in 61% and 69% of the samples, and finds a gain of at least 2.0 in 40% and 54%. SBC picks a model other than white noise in only 2% and 7%. The selected model is therefore $Z_t = \varepsilon_t$, so $\ln P_t$ follows an ARIMA(0,1,0), a random walk without drift.

3.4 Error diagnostics and the variance (step 6)

The residuals of the selected model are the returns themselves. They are uncorrelated, but their squares are not: the first autocorrelation of $\hat{\varepsilon}_t^2$ is 0.20, and lags 4 and 14 also leave the band. The ARCH LM test (TR^2 from regressing $\hat{\varepsilon}_t^2$ on m of its lags) and the Ljung-Box test on squares reject constant variance at the 1% level (Table 4, Panel A). The white-noise model with constant variance therefore fails step 6.

We model the conditional variance h_t of $\varepsilon_t = \sqrt{h_t}\, u_t$. Besides ARCH($q$) and GARCH(1,1), we



Figure 4. AIC and SBC of every ARMA(p, q), $p, q \leq 3$, relative to white noise. Hollow markers: models with a root of modulus below 1.05.

estimate

$$\text{GJR:} \quad h_t = \omega + (\alpha + \gamma I_{t-1})\varepsilon_{t-1}^2 + \beta h_{t-1}, \quad I_{t-1} = 1 \text{ if } \varepsilon_{t-1} < 0, \tag{1}$$

$$\text{EGARCH:} \quad \ln h_t = \omega + \alpha(|u_{t-1}| - \mathrm{E}|u_{t-1}|) + \gamma u_{t-1} + \beta \ln h_{t-1}, \tag{2}$$

$$\text{GARCH-M:} \quad Z_t = \mu + \delta h_t + \varepsilon_t, \quad h_t \text{ GARCH(1,1)}, \tag{3}$$

following Glosten, Jagannathan and Runkle,[5] Nelson[6] and Engle, Lilien and Robins,[7] with normal or Student-t u_t, by maximum likelihood from several starting values. A positive γ in GJR, or a negative γ in EGARCH, means that falls raise next-day variance more than rises of the same size.

Table 4. ARCH diagnostics and variance models, 2017-2018 ($T = 505$)

Panel A. Tests on the white-noise errors

	LM, $m = 1$	LM, $m = 5$	LM, $m = 10$	$Q^2(5)$	$Q^2(10)$
Statistic	19.79***	24.82***	25.71***	29.7***	31.2***
p-value	<0.001	<0.001	0.004	<0.001	<0.001

Panel B. Variance models

Model	$\ln L$	SBC	Persist.	$Q^2(10)$ p	LM(5) p	Status
Constant variance	−524.0	1054.3	0	0.001	<0.001	fails
ARCH(1)	−509.7	1031.8	0.24	0.25	0.045	fails
ARCH(5)	−504.9	1047.2	0.41	0.99	0.95	passes
GARCH(1,1), normal	−509.5	1037.7	0.94	0.63	0.33	passes
GARCH(1,1), t	−495.2	1015.3	1.00	0.035	0.003	nonstationary
GJR(1,1), t	**−491.7**	**1014.6**	0.53	0.42	0.08	**accepted**
EGARCH(1,1), t	−480.8	992.8	1.00	0.36	0.19	nonstationary
GARCH-M, t	−494.2	1025.6	1.00	0.023	0.001	nonstationary

SBC $= -2 \ln L + k \ln T$. Panel B diagnostics use standardised residuals $\hat{u}_t$. Persistence is $\alpha + \beta + \gamma/2$, or β for EGARCH.

Before estimation we fixed an acceptance rule: among covariance-stationary fits (persistence below 0.999), take the lowest SBC whose standardised residuals pass $Q(10)$, $Q^2(10)$ and LM(5) at 5%. ARCH(1) fails, and ARCH(5), GARCH(1,1) and GJR-t pass. The rule selects **GJR(1,1) with Student-t errors** (Table 5), $\hat{h}_t = 0.229 + 0.514\, I_{t-1}\hat{\varepsilon}^2_{t-1} + 0.268\, \hat{h}_{t-1}$ with $\hat{\nu} = 7.2$, where $\hat{\alpha}$ sits at its lower bound of zero. Its persistence is 0.53, a half-life of about one day, and its implied unconditional variance of 0.48 matches the sample variance of 0.47.

Table 5. Variance model estimates, 2017-2018 (robust standard errors in brackets)

Model	ω	α	γ	β	ν	δ
ARCH(1)	0.356***	0.238***				
	(0.037)	(0.083)				
GARCH(1,1), normal	0.031	0.077		0.861**		
	(0.115)	(0.156)		(0.395)		
GARCH(1,1), t	0.001	0.000		1.000***	6.3***	
	(0.001)	(0.007)		(0.006)	(1.8)	
GJR(1,1), t	0.229	0.000	0.514	0.268	7.2***	
	(0.684)	(0.970)	(0.360)	(2.166)	(2.7)	
GJR(1,1), normal	0.198	0.000	0.402**	0.369		
	(0.177)	(0.190)	(0.190)	(0.530)		
EGARCH(1,1), t, interior mode	−0.322	0.067	−0.334***	0.629**	7.5***	
	(0.225)	(0.101)	(0.122)	(0.251)	(2.2)	
GARCH-M, t	0.001	0.000		1.000***	6.1***	−0.244
	(0.001)	(0.007)		(0.007)	(1.7)	(0.193)

Zero mean except GARCH-M ($\hat{\mu} = 0.122$, s.e. 0.085). The EGARCH row is the highest local maximum with $\beta < 0.999$; the global maximum has β at one (Table 4).

The individual GJR-t coefficients are poorly identified, as the robust standard errors show, but the shape of the response to news is clear (Figure 5). Only falls raise next-day variance: a 3% fall lifts h_t to 5.0, while a 3% rise leaves it at 0.35, below the sample variance. The asymmetry is significant at the 1% level in the interior EGARCH mode ($\hat{\gamma} = -0.33$, s.e. 0.12) and at 5% in a normal GJR ($\hat{\gamma} = 0.40$, s.e. 0.19). This is the leverage effect of equity returns.

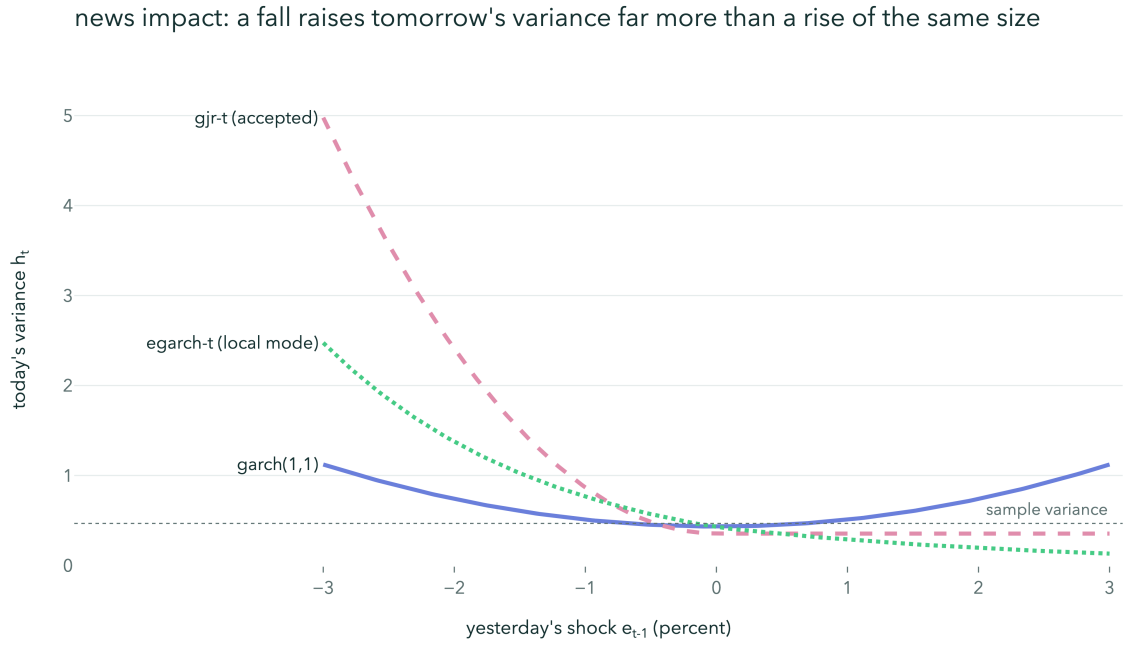


Figure 5. News impact curves: next-day variance as a function of today's shock, with h_{t-1} at the sample variance, for the accepted GJR-t, the EGARCH-t interior mode and GARCH(1,1).

The standardised residuals of the accepted model pass every test: $Q(10)$ has $p = 0.95$, $Q^2(10)$ has $p = 0.42$, and the LM test has $p = 0.29$, 0.08 and 0.41 at $m = 1$, 5 and 10 (Figure 6). Their kurtosis falls from 4.85 to 4.27, close to that of a $t(7.2)$, and the quantile plot follows the fitted t through the body and the lower tail. As a further check we simulated 1,000 samples of 505 days from the accepted model. The observed kurtosis sits at the 33rd percentile of the simulated values and the first autocorrelation of squares at the 60th, whereas a normal GARCH(1,1) reaches the observed kurtosis in only 0.5% of samples.

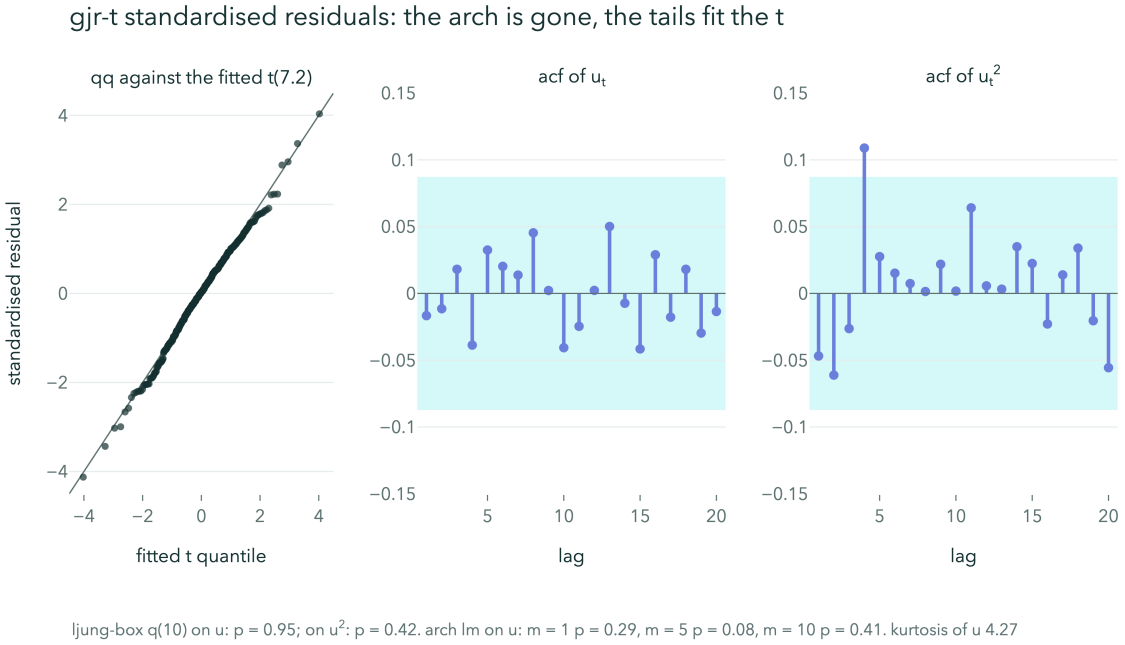


Figure 6. Standardised residuals $\hat{u}_t$ of the accepted GJR-t: quantile plot against the fitted $t(7.2)$, and the ACF of $\hat{u}_t$ and $\hat{u}_t^2$ with 95% bands.

Three fits beat GJR-t on likelihood but fail the rule: GARCH-t, the global EGARCH-t maximum and GARCH-M all drive persistence to one. With $\hat{\alpha} = 0$ and $\hat{\beta} = 1$ the GARCH-t variance becomes a slow linear rise, and a profile likelihood in α confirms that this is the true maximum rather than a failure of the optimiser. These fits track the step up in variance from 2017 to 2018 rather than shocks that decay, and our stationarity condition excludes them. GARCH-M finds no risk premium ($\hat{\delta} = -0.24$, s.e. 0.19, $p = 0.20$; with the standard deviation in the mean, -0.31, $p = 0.21$). A premium of a few per cent a year is about 0.02% a day, too small for two years of data to detect, so this result is uninformative rather than evidence against a premium.

3.5 Forecasting the mean in 2019 (step 7)

Each 2019 forecast uses data up to the previous day only. The mean models are re-estimated every day on an expanding window. The variance models are re-estimated on the first trading day of each month and filtered daily in between. A positive Diebold-Mariano (DM) statistic means that the model loses to the benchmark (Table 6).

No mean model beats the random walk (Figure 7). Every DM statistic is positive, and ARMA(1,1) is significantly worse at the 1% level (DM = 2.63, $p = 0.009$). Its losses build up from August, when its nearly cancelling AR and MA terms amplify the swings around the tariff escalation. No model predicts the sign of the next return better than chance: the hit rates of 43% to 49% are below the 54% share of up days, and the Pesaran-Timmermann test never rejects. In 2019, none of these one-step forecasts was more accurate than the zero forecast, as the ARIMA(0,1,0) predicts.

Table 6. One-step-ahead forecast evaluation, 2019 (253 days)

Panel A. Mean; benchmark: random walk (zero return)					
Model	RMSE	MAE	DM vs RW	p	Correct sign
Random walk	0.7396	0.5570			
Historical mean	0.7402	0.5574	1.93*	0.055	49.0%
AR(1)	0.7419	0.5591	1.64	0.10	42.7%
MA(1)	0.7419	0.5592	1.64	0.10	43.1%
ARMA(1,1)	0.7854	0.5945	2.63***	0.009	46.6%

Panel B. Variance; benchmark: EWMA ($\lambda = 0.94$)					
Model	QLIKE	DM vs EWMA	p	VaR 1% hits (p)	VaR 5% hits (p)
Constant variance	0.371	−0.46	0.65	6 (0.06)	9 (0.27)
EWMA	0.384			7 (0.02)	8 (0.15)
GARCH(1,1), normal	**0.354**	−2.40**	0.017	6 (0.06)	10 (0.43)
GARCH(1,1), t	0.399	0.94	0.35	5 (0.17)	10 (0.43)
GJR(1,1), t	0.413	0.97	0.34	5 (0.17)	12 (0.85)
EGARCH(1,1), t	0.522	2.33**	0.020	5 (0.17)	14 (0.70)

QLIKE = $\hat{\sigma}_t^2/h_t - \ln(\hat{\sigma}_t^2/h_t) - 1$, with $\hat{\sigma}_t^2$ the high-low range variance scaled to close-to-close returns; lower is better. VaR hits: days with a loss beyond the one-day Value-at-Risk (2.5 and 12.7 expected), Kupiec p-value in brackets.

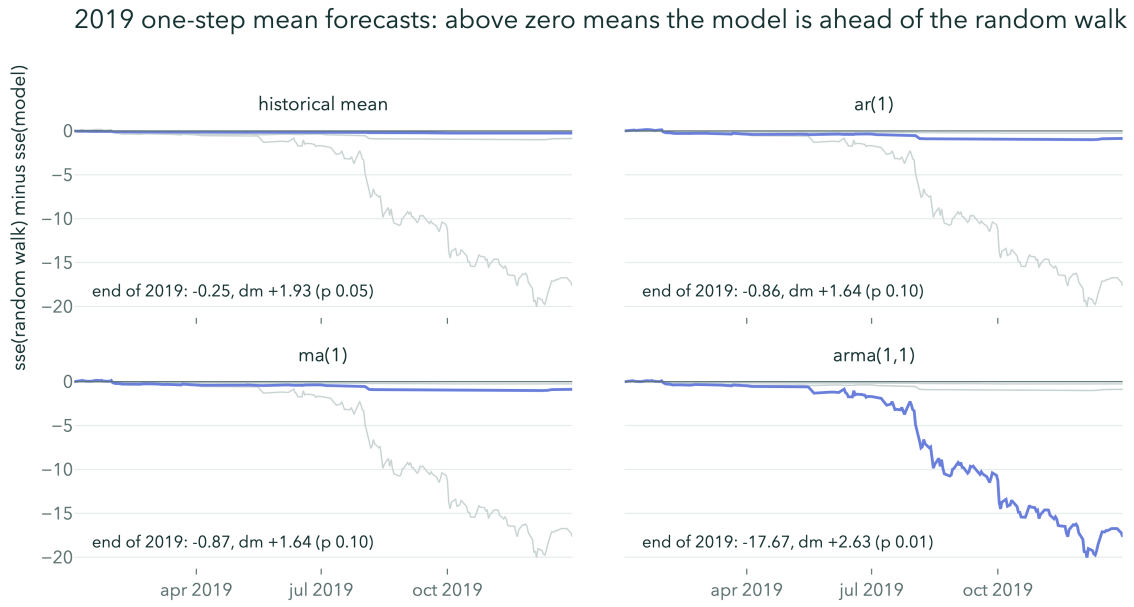


Figure 7. Cumulative squared-error difference between the random walk and each mean model over 2019; above zero, the model is ahead.

3.6 Forecasting the variance in 2019 (step 7)

We compare variance forecasts with the QLIKE loss against a range-based proxy, the Parkinson high-low variance scaled so that its 2017-2018 mean equals that of Z_t^2. QLIKE ranks forecasts correctly even when the proxy is noisy.[8] GARCH(1,1) with normal errors has the lowest loss (Table 6, Panel B). It beats EWMA significantly (DM $= -2.40$, $p = 0.017$) and also beats GARCH-t (-3.09, $p = 0.002$), while EGARCH-t is significantly worse than EWMA ($p = 0.020$). Figure 8 shows that EGARCH-t loses most of its ground in early August 2019, when its forecast standard deviation stayed near 0.5% while realised volatility jumped.

The accepted GJR-t loses to a constant variance, though not significantly ($p = 0.09$). With persistence of 0.53 it forgets a shock within a day, so it mostly reacts to yesterday's fall. Its likelihood also

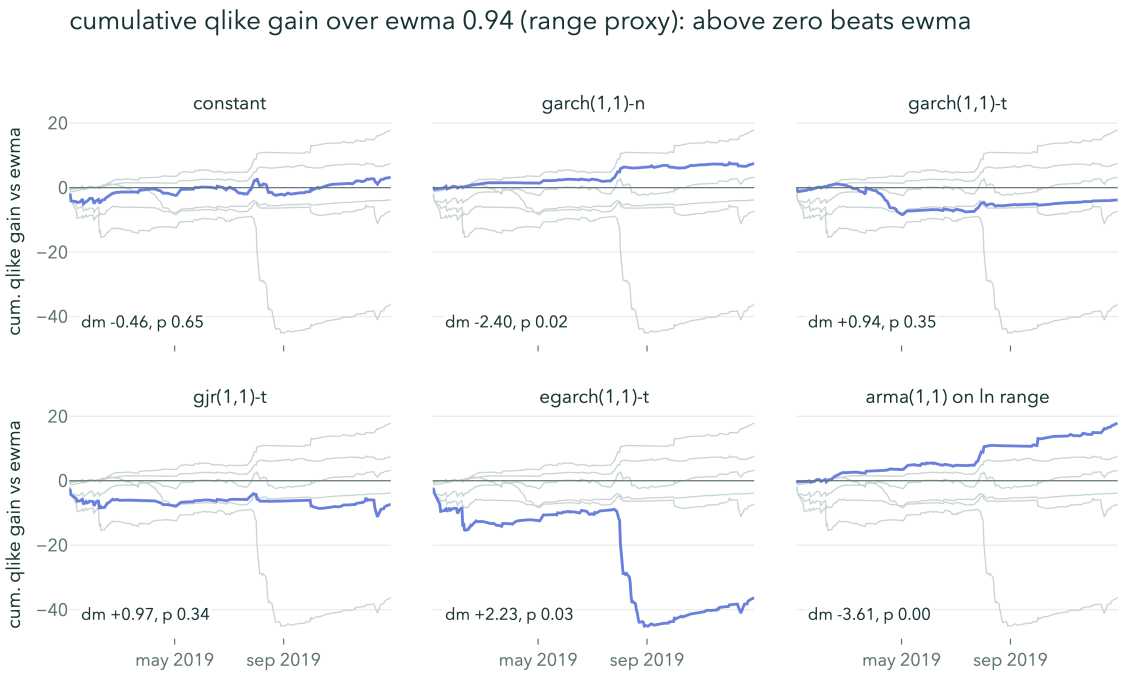


Figure 8. Cumulative QLIKE gain of each variance forecast over EWMA in 2019, one model highlighted per panel; above zero, the model beats EWMA. The range-based ARMA(1,1) in the last panel uses high-low data and is shown for reference only.

has two modes: the monthly re-estimates stay in the low-persistence mode in January and February 2019 and jump to $\hat{\beta} \approx 0.97$ from March, so the forecasting model changes character during the year. Richer volatility models often fit better in sample and forecast worse,[9] and step 7 is where this shows. A Box-Jenkins model of the log range itself, an ARMA(1,1), forecasts better still (QLIKE 0.313), but it uses high-low data and so lies outside our return-only comparison.

The one-day 95% intervals from GJR-*t* cover 94.9% of 2019 returns. All models pass the 5% Kupiec VaR backtest, and only EWMA, which uses normal quantiles, fails at 1%.

With 253 forecasts these tests have little power. We simulated returns from an AR(1) with GARCH-*t* errors, re-estimated the AR(1) daily as in our forecasting exercise, and applied the DM test against the zero forecast. With a true coefficient of 0.05, close to our in-sample estimate of -0.043, the test favours the AR(1) in only 1.1% of samples; at 0.20 it does so in 21%, and it would take about 2,500 test days to reach 95%. The 1% Kupiec test detects a true exceedance rate of 2% only a quarter of the time and needs about 1,000 days to reach 80% power. Failing to beat the random walk therefore shows that predictability is small, not that it is zero.

3.7 Forecast for January 2020

As the final output of step 7, we re-estimated the accepted model on all of 2017-2019 ($\hat{\omega} = 0.077$, $\hat{\alpha} = 0$, $\hat{\gamma} = 0.188$, $\hat{\beta} = 0.751$, $\hat{\nu} = 6.5$) and simulated 20,000 paths of the 22 trading days of January 2020 (Figure 9). The mean forecast is zero, as ARIMA(0,1,0) implies, and the one-day forecast standard deviation is 0.68%. On the realised data, 21 of the 22 daily returns fall inside the 95% band. The one miss is the -2.32% fall of 27 January, the first coronavirus sell-off. The cumulative return of -3.46% at the end of the month stays inside its 95% band of -7.2% to 5.9%.

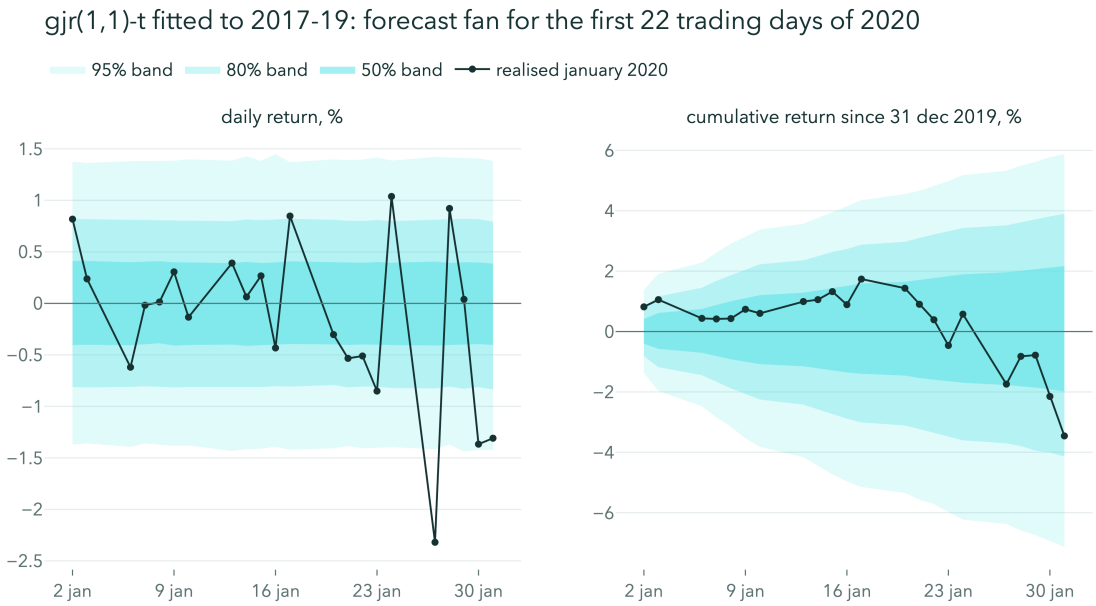


Figure 9. January 2020 forecast fan (50%, 80%, 95% bands) from GJR-*t* fitted to 2017-2019, daily (left) and cumulative (right), with realised values.

4 Conclusion

The log FTSE 100 is integrated of order one, and daily returns over 2017-2018 are white noise, so the Box-Jenkins model for the index is ARIMA(0,1,0). SBC selects it, AIC's ARMA(3,3) is a near-cancelling overfit, and in 2019 no ARMA forecast beats the random walk, while ARMA(1,1) is significantly worse. The errors are not homoskedastic. The ARCH LM test rejects at the 1% level, and by our pre-stated rule the variance is best described by a GJR(1,1) with Student-*t* errors, which captures the leverage effect and the fat tails. Out of sample, a plain GARCH(1,1) with normal errors forecasts the 2019 variance best and significantly beats the EWMA benchmark, so the model that wins in sample is not the one that forecasts best.

Our simulations qualify these results: at this sample size the ADF test cannot separate a unit root from a root of 0.98, and the forecast tests could not detect small but real predictability. Part of the measured ARCH effect reflects the step up in volatility from 2017 to 2018, which is also why several fits run to unit persistence. A variance break in 2018 and a longer forecast sample are the natural extensions.

5 References

[1] Box, G. E. P., Jenkins, G. M., Reinsel, G. C. and Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*, 5th ed. Wiley.

[2] Kantawala, D. (Fall 2026). *EC 602: Time Series Analysis*, LaTeX lecture notes. dhairyakantawala.github.io/notes.

[3] Yahoo Finance. FTSE 100 (^FTSE) daily prices. finance.yahoo.com.

[4] Fama, E. F. (1970). Efficient capital markets: a review of theory and empirical work. *J. Finance*, 25(2), 383-417.

[5] Glosten, L. R., Jagannathan, R. and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *J. Finance*, 48(5), 1779-1801.

[6] Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 59(2), 347-370.

[7] Engle, R. F., Lilien, D. M. and Robins, R. P. (1987). Estimating time varying risk premia in the term structure: the ARCH-M model. *Econometrica*, 55(2), 391-407.

[8] Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *J. Econometrics*, 160(1), 246-256.

[9] Hansen, P. R. and Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *J. Applied Econometrics*, 20(7), 873-889.

[10] Anthropic (2026). *Claude Code* (Claude Opus 5.5), AI assistant.