SUPPLEMENTARY MATERIAL FOR

Generalized Autoregressive Score Trees and Forests

Andrew J. Patton and Yasin Simsek

This Version: August 7, 2025

This appendix contains three sections. Section S.1 presents additional details for the t-GAS copula model considered in Section 3.4. Section S.2 presents a simulation study of the GARCH tree and forest models in the presence of parameter instability and DGP non-linearity. Section S.3 presents the results of conditional forecast comparisons of baseline and forest-based GAS models. Section S.4 presents some additional tables and figures.

S.1 Derivation of the score function for the t-GAS copula

In this section we present the score function for the t-copula analysis discussed in Section 3. We refer to Creal et al. (2013) for the details of the univariate applications (the GARCH and t-GAS models).

S.1.1 Notation

We adopt the notation of Creal et al. (2011) for ease of comparability with that article. The Kronecker product is denoted by $A \otimes B$ for any matrices A and B. $A_{\otimes}$ stands for $A \otimes A$. The function $vec(A)$ vectorizes matrix A into a column vector, and $vech(A)$ vectorizes just the lower triangle of A, which eliminates duplicates in the case that A is symmetric. The duplication matrix is implicitly defined as the solution to $\mathcal{D}\,\mathrm{vech}(A) = \mathrm{vec}(A)$. Finally, $\mathbb{E}_{t-1}$ denotes the expectation conditional on the information available up to period $t-1$.

S.1.2 The probability density function of t copula

We adopt Student's t copula specification in our empirical analysis and its probability density function is given by

$$c(\mathbf{u}_t;\Sigma_t,\nu)=\frac{\Gamma\left(\frac{\nu+2}{2}\right)\Gamma\left(\frac{\nu}{2}\right)}{\sqrt{|\Sigma_t|}\left[\Gamma\left(\frac{\nu+1}{2}\right)\right]^2}\left(1+\frac{\mathbf{x}_t'\Sigma_t^{-1}\mathbf{x}_t}{\nu}\right)^{-\frac{\nu+2}{2}}\prod_{i=1}^{2}\left(1+\frac{x_{i,t}^2}{\nu}\right)^{\frac{\nu+1}{2}} \tag{12}$$

where $\mathbf{x}_t=[x_{1,t},x_{2,t}]=[T_\nu^{-1}(u_{1,t}),T_\nu^{-1}(u_{2,t})]'$ obtained by applying the inverse of the univariate t distribution with ν degrees of freedom, $\Gamma(\cdot)$ is gamma function and Σ_t is 2-by-2 correlation matrix. We denote the off-diagonal element of Σ_t with ρ_t which is the variable of interest:

$$\Sigma_t=\begin{bmatrix}1 & \rho_t\\ \rho_t & 1\end{bmatrix} \tag{13}$$

S.1.3 The score and information matrix

We use inverse information matrix of the score function as a scaling factor in all applications. Given the complex structure of the Student's t copula, derivation of the information matrix requires tedious calculations, but Creal et al. (2011) provide a closed-form formula of both score and information matrix. Based on their results, we can write

$$\begin{aligned}\nabla_t &= \tfrac{\partial \log c_t(y_t|\Sigma_t;\nu)}{\partial f_t}\\ &= \tfrac{1}{2}(\mathcal{D}\Psi_t)'\Sigma_{t\otimes}^{-1}\left[w_t\mathbf{x}_{t\otimes}-\mathrm{vec}\left(\Sigma_t\right)\right]\\ \mathcal{I}_{t|t-1} &= \mathbb{E}_{t-1}\left[\nabla_t\nabla_t'\right]\\ &= \tfrac{1}{4}(\mathcal{D}\Psi_t)'J_{t\otimes}'\left[gG-\mathrm{vec(I)}\,\mathrm{vec(I)}'\right]J_{t\otimes}\mathcal{D}\Psi_t\end{aligned} \tag{14}$$

where $\Psi_t\equiv\frac{\partial\,\mathrm{vech}(\Sigma_t)}{\partial\rho_t}$, J_t is such that $\Sigma_t^{-1}=J_t'J_t$, $w_t\equiv\frac{\nu+2}{\nu-2+\mathbf{x}_t'\Sigma_t^{-1}\mathbf{x}_t}$, $g\equiv\frac{v+2}{v+4}$, and the explicit form of matrix G is

$$G=\begin{bmatrix}3 & 0 & 0 & 1\\ 0 & 1 & 1 & 0\\ 0 & 1 & 1 & 0\\ 1 & 0 & 0 & 3\end{bmatrix}. \tag{15}$$

We define the scaled score functions as $s_t = \mathcal{I}_{t|t-1}^{-1} \nabla_t$. As in Janus et al. (2014) we use a transformation to ensure $\rho_t \in (-1, 1)$, by setting $\rho_t = \frac{1-\exp(-\tilde{\rho}_t)}{1+\exp(-\tilde{\rho}_t)}$, where $\tilde{\rho}_t \in \mathbb{R}$. In order to obtain the scaled score function for $\tilde{\rho}_t$, we multiply the original scaled score with the derivative of the transformation function: $\tilde{s}_t = \frac{\partial \tilde{\rho}_t}{\partial \rho_t} s_t$. When we use the explicit form of each component in equation (14), we obtain the following expression for the scaled score of the t copula:

$$s_t = \left(\frac{2}{1-\rho_t^2}\right)\left(\frac{1+\rho_t^2}{g+(2g-1)\rho_t^2}\right)\left(w_t(x_{1,t}x_{2,t}-\rho_t) - \frac{\rho_t}{1+\rho_t^2}(w_t x_{1,t}^2 + w_t x_{2,t}^2 - 2)\right) \tag{16}$$

Noting that $(g, w_t) \to (1, 1)$ as $\nu \to \infty$, we thus also obtain the scaled score function for Gaussian copula:

$$s_t = \left(\frac{2}{1-\rho_t^2}\right)\left(x_{1,t}x_{2,t} - \rho_t - \frac{\rho_t}{1+\rho_t^2}(x_{1,t}^2 + x_{2,t}^2 - 2)\right). \tag{17}$$

S.2 Simulation study

In this section we consider three data generating processes (DGPs) to evaluate the performance of GAS Tree and Forest models under varying degrees of model complexity and dynamics in the conditional variance.

All DGPs assume that returns follow a zero-mean normal distribution, i.e.,

$$r_t \sim \mathcal{N}(0, \sigma_t^2), \quad t = 1, \ldots, T$$

where the variance σ_t^2 evolves over time according to different specifications described below.

- **DGP 1 (Baseline GARCH):** This is a standard GARCH(1,1) model with constant parameters:

  $$\sigma_t^2 = 0.05 + 0.80\,\sigma_{t-1}^2 + 0.18\,r_{t-1}^2.$$

  This specification captures persistence in volatility through past variance and past squared returns.

- **DGP 2 (Structural break in GARCH parameters):** In this setting, we introduce a structural break in the GARCH parameters at time $t = T/4$. While the long-run variance and total persistence remain unchanged, the contribution of past variance becomes more pronounced in the post-break period:

$$\sigma_t^2 = \begin{cases} 0.05 + 0.80\,\sigma_{t-1}^2 + 0.18\,r_{t-1}^2, & \text{if } t \le T/4, \\ 0.05 + 0.90\,\sigma_{t-1}^2 + 0.08\,r_{t-1}^2, & \text{if } t > T/4. \end{cases}$$

- **DGP 3 (Nonlinear GARCH):** In the final DGP, we allow the coefficient on the score (i.e., the past squared return) to be a nonlinear function of past returns. Specifically, we adopt the following form:

$$\sigma_t^2 = 0.1 + 0.90\,\sigma_{t-1}^2 + 0.10 \cdot \begin{cases} \dfrac{3r_{t-1}^2}{1 + \frac{3r_{t-1}^2}{4}}, & \text{if } r_{t-1} < 0, \\ \dfrac{3r_{t-1}^2}{1 + \frac{3r_{t-1}^2}{2}}, & \text{if } r_{t-1} \ge 0. \end{cases}$$

  This model generates asymmetric responses to past returns and introduces smooth nonlinear dynamics into the conditional variance process.

These DGPs are designed to reflect a range of volatility behaviors commonly observed in financial return series, from stationary GARCH dynamics to structural changes and nonlinear relations in volatility process.

We simulate 1,000 return series for each of the three DGPs described above, each consisting of $T = 2000$ observations. The first half of each sample is used for in-sample estimation, while the second half is reserved for out-of-sample evaluation. We estimate the GARCH Tree and GARCH Forest models as introduced in the main text, and benchmark their performance against the standard GARCH model as well as rolling window GARCH specifications. To construct the GAS trees and forests, we define a set of state variables comprising time, lagged return, lagged variance, and lagged squared return. Table S.1 presents the out-of-sample performance comparison across models.

Table S.1 shows that when the DGP is a standard GARCH model, with no parameter

Table S.1: **QLIKE loss by DGP and model, relative to GARCH loss.** This table presents the out-of-sample average QLIKE loss of four models: a GARCH model estimated using a rolling window of 250 or 500 observations, a GARCH Tree, and a GARCH Forest, each relative to the average QLIKE loss of a GARCH model estimated using the full training sample. Three data generating processes are considered, as described in the text above. The sample size for each simulation is 2,000, with the first half used for estimation and the second half for out-of-sample forecast comparison. The number of simulation replications is 1,000.

DGP	**GARCH**	**RW250**	**RW500**	**Tree**	**Forest**
Baseline	1.000	7.224	3.049	1.179	1.020
Break	1.000	2.133	0.938	0.973	0.701
Nonlinear	1.000	5.219	2.490	1.141	0.907

breaks, the simple GARCH model yields the lowest out-of-sample (OOS) loss. This is not surprising, as that model is correctly specified in this case, and it does not have additional, unnecessary, flexibility as in the GARCH Tree and Forest models. These two models have loss close to the baseline GARCH model, but slightly higher. The worst performing model in for this DGP is the GARCH model using a window of just 250 observations. While that model is also correctly specified, the short sample size means that parameter estimation error contaminates the forecast and leads to higher OOS loss.

In the second DGP the parameters of the GARCH model change halfway through the estimation sample. The rolling window estimator using 500 observations beats the baseline GARCH model, with OOS loss 6.2% lower, while the model using 250 observations continues to perform poorly out of sample. The GARCH Tree and Forest models are able to detect the parameter instability and both models beat the baseline GARCH model. Interestingly, the GARCH Forest model is the best-performing model for this design, with OOS almost 30% lower than the baseline GARCH model, and 23.7% lower than the GARCH model estimated using a rolling window of 500 observations.

In the third DGP, where the GARCH parameters are a nonlinear function of past returns, we see the rolling window GARCH models perform poorly; these models are capable of capturing breaks in parameters but are not well-suited to capturing omitted nonlinearities.

The GARCH Tree model is theoretically able to capture such nonlinearities, but we see in this simulation that it performs slightly worse than the baseline GARCH model, revealing that its improved fit is more than offset, out of sample, by increased estimation error. The GARCH Forest, on the other hand, is able to capture the nonlinearity and mitigate estimation error, by averaging across the trees in the forest, and attains the lowest OOS loss for this design, with loss almost 10% lower than the next best model.

S.3 Conditional forecast comparison

In the main text we compared competing methods based on their average out-of-sample loss, providing an unconditional assessment. Next, we examine whether the state variables used in the construction of the tree structures can also help predict which model outperforms the other. We focus on the relative performance of the baseline GAS model and the GAS Forest model for each of our four applications.

In this section we conduct two empirical exercises to investigate this question. First, we run linear regressions of the out-of-sample loss differences between the competing models on a constant and a state variable, following the approach of Giacomini and White (2006). This allows us to assess whether the state variable has predictive power for model performance differences. Second, we estimate the loss differences using a nonparametric kernel smoothing technique, which enables us to explore how model performance varies across different regions of the state variable's support.

In each application, we select the state variable with the highest SHAP value. Specifically, we use RVOL for the GARCH, t-GAS, and ACD models, and T10Y for the t-Copula models.

Table S.3 presents the results of the Giacomini and White (2006) regressions. The columns correspond to the four applications considered in the main paper, while the rows report the regression coefficients along with their standard errors and corresponding t-statistics. We demean the state variable so that the t-statistic for the intercept aligns with the Diebold-Mariano statistic reported in the out-of-sample comparison section.[1]

[1]The t-statistics in Table S.3 differ very slightly from those in Tables 1–4 (e.g., -11.490 vs -11.558 for the GARCH application) due to the numerical implementation of Python's built-in functions.

Table S.2: **Conditional comparisons of GAS Models.** This table reports the estimated coefficients and standard errors from linear regressions of out-of-sample loss differences on a constant and the lagged state variable, across the four applications considered in this paper. In each case, the forecasts from the GAS Forest model are compared to those of the basline GAS model. RVOL is used as the conditioning variable for the GARCH, t-GAS, and ACD models, while T10Y is used for the t-Copula model.

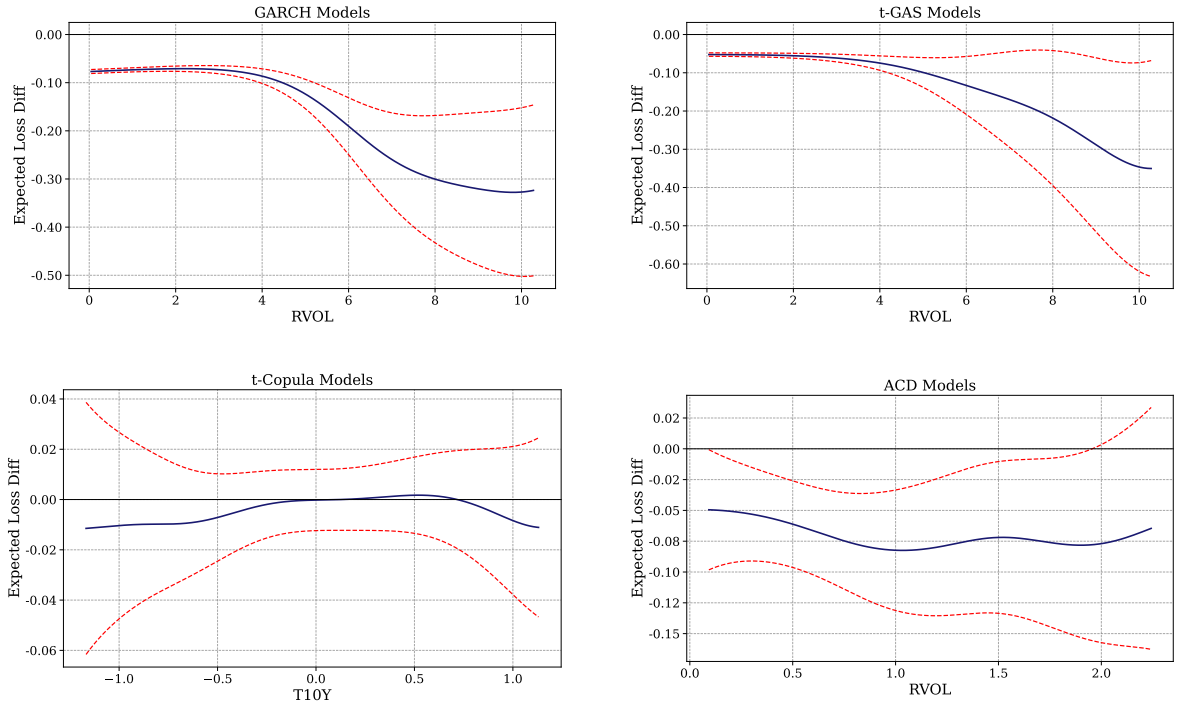
	GARCH	t-GAS	t-Copula	ACD
Intercept	-0.083	-0.059	-0.002	-0.066
(std. err.)	(0.007)	(0.006)	(0.003)	(0.012)
[t-stat]	[-11.490]	[-10.198]	[-0.813]	[-5.357]
Slope	-0.009	-0.007	0.006	-0.016
(std. err.)	(0.008)	(0.004)	(0.006)	(0.021)
[t-stat]	[-1.038]	[-2.072]	[1.057]	[-0.781]

The slope coefficients in Table S.3 indicate no evidence that the selected state variable (linearly) predicts the relative performance of forest models in the GARCH, ACD, or t-Copula applications. In contrast, for the t-GAS model we observe that RVOL has predictive power, with a slope t-statistic below -2. The negative sign of the slope indicates that the t-GAS forest model outperforms its benchmark more strongly when market volatility is high. This finding underscores the strength of tree-based methods, particularly in periods of elevated uncertainty, where forecasting return distributions is typically more challenging.

Figure S.1 displays the estimated expected out-of-sample loss differences between models as a function of the selected state variables, along with 95% pointwise confidence intervals. These estimates are obtained by applying a Gaussian kernel smoother to the loss differences. The top-left and top-right panels indicate that the GAS forest models outperform the baseline GAS model by a greater margin when RVOL is high. As expected, the confidence intervals become wider in periods of elevated market volatility, though they still exclude zero, indicating suggesting statistically significant gains.

On the other hand, the bottom panels show that the t-Copula and ACD forest models exhibit relatively stable performance across the support of T10Y and RVOL, respectively.

Figure S.1: **Conditional expected loss differences between GAS models.** This figure shows the estimated expected loss differences between the GAS forest and baseline GAS models, conditional on RVOL (top left, top right, bottom right) and T10Y (bottom left). Negative values indicate that the GAS forest model outperforms the plain GAS model. The differing support of RVOL in the bottom-right plot reflects differences in sample periods across applications.



Specifically, while the ACD forest model continues to outperform its benchmark overall, this advantage diminishes in the right tail of RVOL, where statistical significance is lost.

S.4 Additional tables and figures

Table S.3: **Out-of-sample performance of GARCH models, alternative sample splits.** This table presents results analogous to Table 1 of the main paper. That table uses training, validation and test sample splits of 30/30/40. Panel A instead uses splits of 30/20/50, and Panel B uses 30/30/10/30, where the third subsample is discarded to separate the estimation sample from the test sample. See the notes to Table 1 for more details.

	GARCH	RW250	RW500	DRF	Small G Tree	GARCH Tree	GARCH Forest
Panel A: Sample splits 30/20/50							
RW250	-4.564						
RW500	-5.092	0.344					
DRF	2.204	4.682	4.833				
Small G Tree	-4.801	-0.974	-1.259	-5.043			
GARCH Tree	-7.168	-3.492	-4.064	-7.316	-3.172		
GARCH Forest	-10.667	-4.984	-6.855	-8.484	-3.211	-0.591	
Avg loss	0.382	0.342	0.344	0.412	0.331	0.309	0.306
MCS						✓	✓
Panel B: Sample splits 30/30/10/30							
RW250	-3.964						
RW500	-4.963	-0.046					
DRF	0.770	2.837	3.019				
Small G Tree	-3.599	-0.363	-0.371	-2.899			
GARCH Tree	-6.346	-2.620	-2.885	-4.787	-3.083		
GARCH Forest	-9.567	-3.171	-4.314	-5.183	-2.212	0.111	
Avg Loss	0.410	0.361	0.360	0.427	0.355	0.327	0.327
MCS						✓	✓

Table S.4: **Out-of-sample performance of t-GAS models, alternative sample splits.** This table presents results analogous to Table 2 of the main paper. That table uses training, validation and test sample splits of 30/30/40. Panel A instead uses splits of 30/20/50, and Panel B uses 30/30/10/30, where the third subsample is discarded to separate the estimation sample from the test sample. See the notes to Table 2 for more details.

	GAS	RW250	RW500	DRF	Small GAS Tree	GAS Tree	GAS Forest
Panel A: Sample splits 30/20/50							
RW250	-0.565						
RW500	-1.085	-0.536					
DRF	-0.344	0.039	0.238				
Small GAS Tree	-4.808	-3.854	-4.158	-3.617			
GAS Tree	-6.240	-4.896	-5.531	-4.813	-0.896		
GAS Forest	-8.128	-6.022	-7.065	-6.257	-2.180	-1.979	
Avg loss	1.257	1.253	1.251	1.253	1.214	1.210	1.204
MCS							✓
Panel B: Sample splits 30/30/10/30							
RW250	-0.302						
RW500	-0.720	-0.419					
DRF	-2.294	-1.416	-1.425				
Small GAS Tree	-4.905	-3.709	-4.248	-1.840			
GAS Tree	-6.334	-4.498	-5.202	-3.055	-1.595		
GAS Forest	-8.500	-5.503	-6.885	-3.916	-2.489	-1.760	
Avg loss	1.222	1.219	1.216	1.197	1.172	1.165	1.159
MCS							✓

Table S.5: **Out-of-sample performance of t Copula GAS models, alternative sample splits.** This table presents results analogous to Table 3 of the main paper. That table uses training, validation and test sample splits of 30/30/40. Panel A instead uses splits of 30/20/50, and Panel B uses 30/30/10/30, where the third subsample is discarded to separate the estimation sample from the test sample. See the notes to Table 3 for more details.

	GAS	RW250	RW500	DRF	Small GAS Tree	GAS Tree	GAS Forest
Panel A: Sample splits 30/20/50							
RW250	2.513						
RW500	1.723	-1.825					
DRF	4.739	1.873	3.208				
Small GAS Tree	-1.686	-2.832	-2.169	-5.664			
GAS Tree	-0.918	-2.486	-1.724	-4.870	0.987		
GAS Forest	0.584	-2.034	-1.136	-5.180	2.302	1.293	
Avg Loss	-0.115	-0.105	-0.110	-0.093	-0.119	-0.118	-0.114
MCS	✓		✓		✓	✓	✓
Panel B: Sample splits 30/30/10/30							
RW250	1.836						
RW500	1.357	-0.885					
DRF	1.143	-0.155	0.322				
Small GAS Tree	-1.232	-2.010	-1.619	-1.324			
GAS Tree	-1.559	-2.181	-1.870	-1.805	-1.381		
GAS Forest	-1.219	-2.239	-2.052	-1.748	-1.077	0.665	
Avg Loss	-0.086	-0.078	-0.081	-0.079	-0.086	-0.091	-0.089
MCS	✓	✓	✓	✓	✓	✓	✓

Table S.6: **Out-of-sample performance of ACD models, alternative sample splits.** This table presents results analogous to Table 4 of the main paper. That table uses training, validation and test sample splits of 30/30/40. Panel A instead uses splits of 30/20/50, and Panel B uses 30/30/10/30, where the third subsample is discarded to separate the estimation sample from the test sample. See the notes to Table 4 for more details.

	ACD	RW250	RW500	DRF	Small ACD Tree	ACD Tree	ACD Forest
Panel A: Sample splits 30/20/50							
RW250	-8.389						
RW500	-9.328	-0.606					
DRF	-8.445	2.380	2.958				
Small ACD Tree	-12.525	-1.942	-1.738	-6.076			
ACD Tree	-10.529	-2.334	-1.946	-5.336	-1.001		
ACD Forest	-13.504	-3.947	-3.935	-9.731	-4.406	-1.666	
Avg loss	-0.377	-0.651	-0.660	-0.584	-0.686	-0.705	-0.729
MCS						✓	✓
Panel B: Sample splits 30/30/10/30							
RW250	0.919						
RW500	1.074	0.015					
DRF	6.808	2.095	2.186				
Small ACD Tree	-1.848	-1.571	-1.942	-5.475			
ACD Tree	-2.212	-2.428	-2.458	-5.365	-1.719		
ACD Forest	-4.745	-3.310	-3.574	-7.910	-3.899	-0.560	
Avg loss	-0.797	-0.768	-0.768	-0.680	-0.810	-0.858	-0.870
MCS						✓	✓

Table S.7: **Out-of-sample performance of GARCH models, MSE loss.** This table presents results analogous to Table 1 of the main paper. That table uses QLIKE as the loss function, while this table instead uses mean-squared error. See the notes to Table 1 for more details.

	GARCH	RW250	RW500	DRF	Small G Tree	GARCH Tree	GARCH Forest
RW250	-0.076						
RW500	-0.585	-0.989					
DRF	0.564	0.592	0.778				
Small G Tree	-2.047	-0.898	-0.703	-1.228			
GARCH Tree	-1.443	-0.623	-0.354	-1.304	0.566		
GARCH Forest	-2.197	-0.723	-0.481	-1.132	1.156	-0.101	
Avg Loss	7.490	7.346	6.652	9.062	5.720	6.036	5.982
MCS	✓	✓	✓	✓	✓	✓	✓

Table S.8: **Out-of-sample performance of t-GAS models, CRPS loss.** This table presents results analogous to Table 2 of the main paper. That table uses the negative log-likelihood as the loss function, while this table instead uses continuous ranked probability score (CRPS). See the notes to Table 2 for more details.

	t-GAS	RW250	RW500	DRF	Small GAS Tree	GAS Tree	GAS Forest
RW250	1.102						
RW500	0.927	-1.603					
DRF	-2.433	-2.107	-2.090				
Small GAS Tree	-0.180	-2.651	-3.209	1.668			
GAS Tree	-2.744	-2.269	-2.362	1.336	-1.806		
GAS Forest	-2.589	-2.242	-2.363	1.504	-1.656	1.875	
Avg loss	0.528	0.552	0.544	0.492	0.526	0.505	0.510
MCS	✓	✓	✓	✓	✓	✓	✓

Table S.9: **Out-of-sample performance of t Copula GAS models, CRPS loss.** This table presents results analogous to Table 3 of the main paper. That table uses the negative log-likelihood as the loss function, while this table instead uses continuous ranked probability score (CRPS). See the notes to Table 3 for more details.

	GAS	RW250	RW500	DRF	Small GAS Tree	GAS Tree	GAS Forest
RW250	1.258						
RW500	1.610	-0.495					
DRF	12.976	3.201	5.310				
Small GAS Tree	9.555	-0.061	0.662	-12.212			
GAS Tree	-5.214	-1.627	-2.331	-13.811	-12.239		
GAS Forest	-10.001	-2.847	-4.620	-14.346	-13.752	-9.512	
Avg Loss	0.944	0.954	0.951	0.979	0.954	0.941	0.931
MCS							✓

Table S.10: **Out-of-sample performance of ACD models, MSE loss.** This table presents results analogous to Table 4 of the main paper. That table uses the negative log-likelihood as the loss function, while this table instead uses continuous ranked probability score (CRPS). See the notes to Table 4 for more details.

	ACD	RW250	RW500	DRF	Small ACD Tree	ACD Tree	ACD Forest
RW250	3.318						
RW500	2.565	-0.099					
DRF	1.479	-3.304	-2.552				
Small ACD Tree	3.532	-3.314	-2.561	-0.941			
ACD Tree	-1.676	-3.324	-2.570	-2.087	-2.447		
ACD Forest	-3.667	-3.326	-2.571	-2.381	-4.472	-0.657	
Avg loss	0.085	0.959	0.917	0.089	0.086	0.083	0.083
MCS						✓	✓

Figure S.2: **Parameter estimates as a function of realized variance (RVOL) for the GARCH forest.** This figure plots the average, across bootstrap samples, values of $\omega/(1-\alpha-\beta)$ (upper-left), $\alpha+\beta$ (upper-right), and α (lower-left) from the GARCH model in equation (8), as well as the average predicted variance, σ_t^2 (lower-right) from that model. The state variable is discretized into bins based on 1% quantiles. The values for each of these quantities from the benchmark GARCH model are plotted in horizontal dashed lines. The solid lines are local quadratic polynomials fitted to the grey dots.

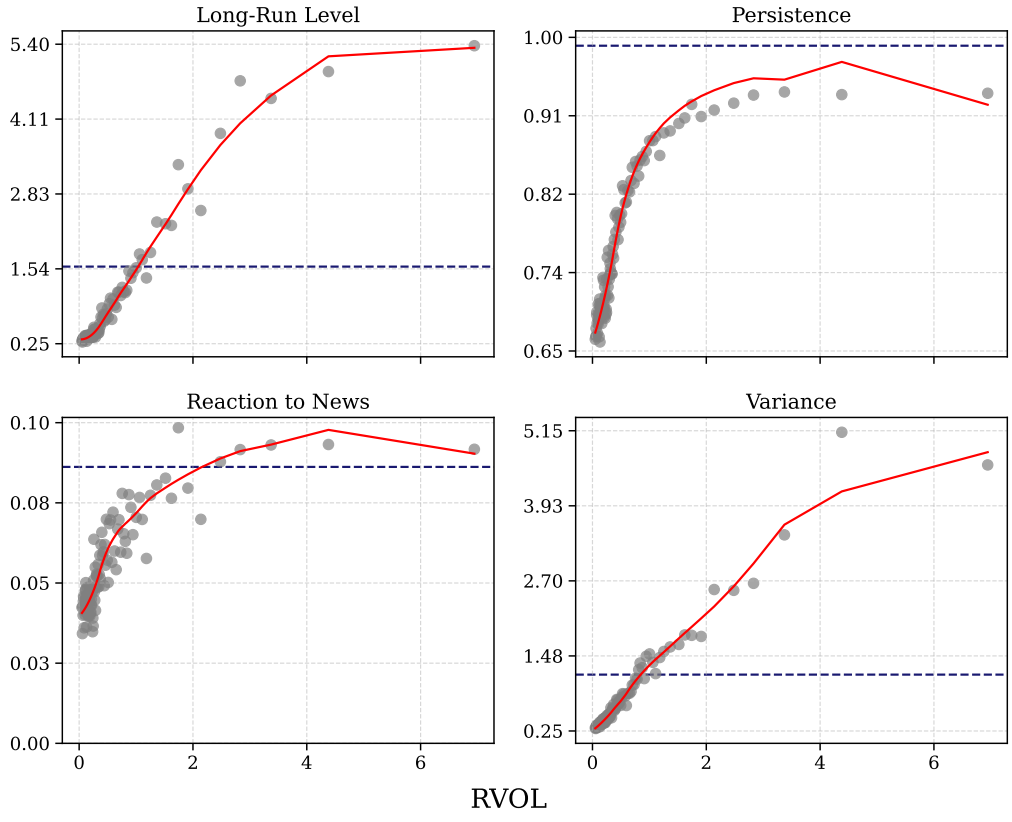


Figure S.3: **Parameter estimates as a function of stock market return (SPX) for the GARCH forest.** This figure plots the average, across bootstrap samples, values of $\omega/(1-\alpha-\beta)$ (upper-left), $\alpha+\beta$ (upper-right), and α (lower-left) from the GARCH model in equation (8), as well as the average predicted variance, σ_t^2 (lower-right) from that model. The state variable is discretized into bins based on 1% quantiles. The values for each of these quantities from the benchmark GARCH model are plotted in horizontal dashed lines. The solid lines are local quadratic polynomials fitted to the grey dots.

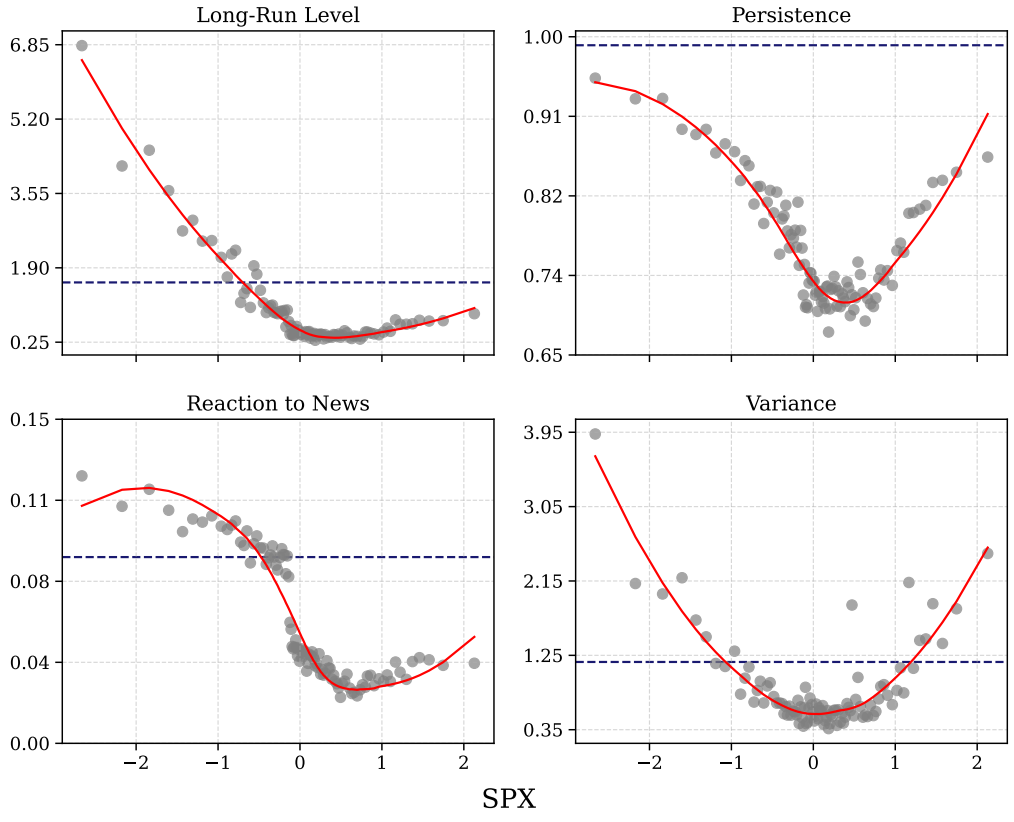


Figure S.4: **Parameter estimates as a function of realized variance (RVOL) for the Student's t GAS forest.** This figure plots the average, across bootstrap samples, values of $\omega/(1-\beta)$ (upper-left), β (upper-right), and α (lower-left) from the GAS model in equation (10), as well as the average predicted variance, σ_t^2 (lower-right) from that model. The state variable is discretized into bins based on 1% quantiles. The values for each of these quantities from the benchmark GAS model are plotted in horizontal dashed lines. The solid lines are local quadratic polynomials fitted to the grey dots.

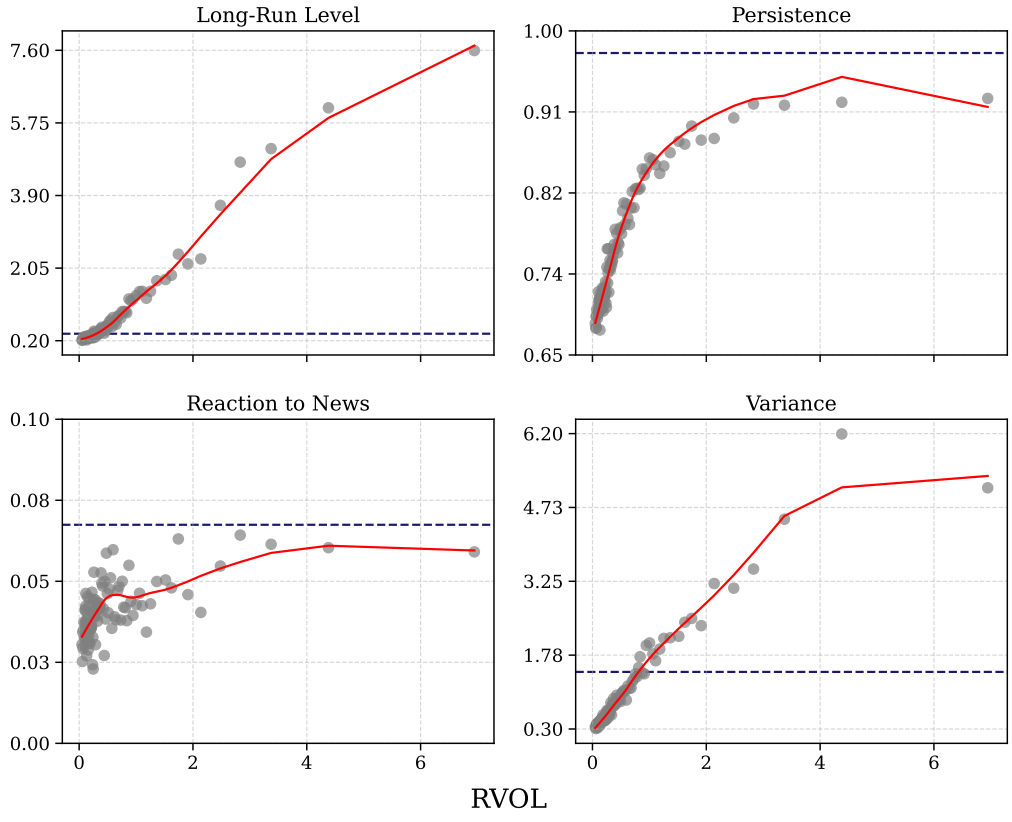


Figure S.5: **Parameter estimates as a function of the stock market return (SPX) for the Student's t GAS forest.** This figure plots the average, across bootstrap samples, values of $\omega/(1-\beta)$ (upper-left), β (upper-right), and α (lower-left) from the GAS model in equation (10), as well as the average predicted variance, σ_t^2 (lower-right) from that model. The state variable is discretized into bins based on 1% quantiles. The values for each of these quantities from the benchmark GAS model are plotted in horizontal dashed lines. The solid lines are local quadratic polynomials fitted to the grey dots.

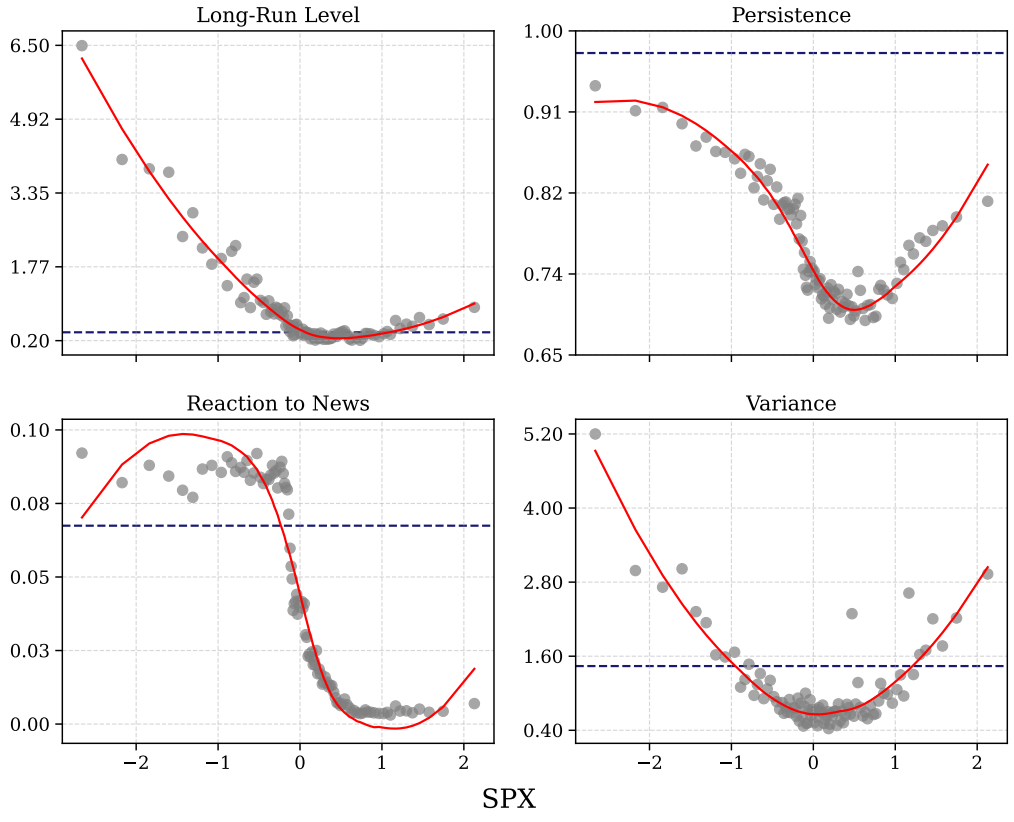


Figure S.6: **Parameter estimates as a function of the stock market return (SPX) for the Student's t copula GAS forest.** This figure plots the average, across bootstrap samples, values of $\omega/(1-\beta)$ (upper-left), β (upper-right), and α (lower-left) from the GAS model in equation (10), as well as the average predicted correlation, ρ_t (lower-right) from that model. The state variable is discretized into bins based on 1% quantiles. The values for each of these quantities from the benchmark GAS model are plotted in horizontal dashed lines. The solid lines are local quadratic polynomials fitted to the grey dots.

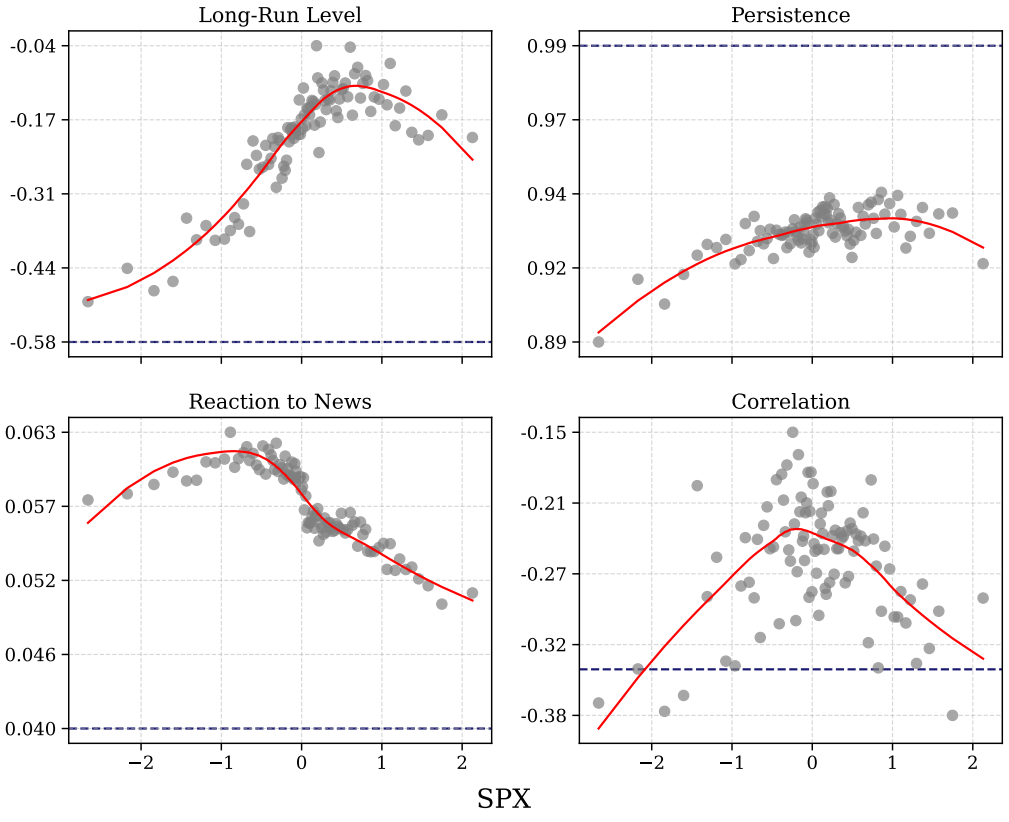


Figure S.7: **Parameter estimates as a function of the stock market return (SPX) for the ACD GAS forest.** This figure plots the average, across bootstrap samples, values of $\omega/(1-\beta)$ (upper-left), β (upper-right), and α (lower-left) from the GAS model in equation (10), as well as the average predicted duration, μ_t (lower-right) from that model. The state variable is discretized into bins based on 1% quantiles. The values for each of these quantities from the benchmark GAS model are plotted in horizontal dashed lines. The solid lines are local quadratic polynomials fitted to the grey dots.

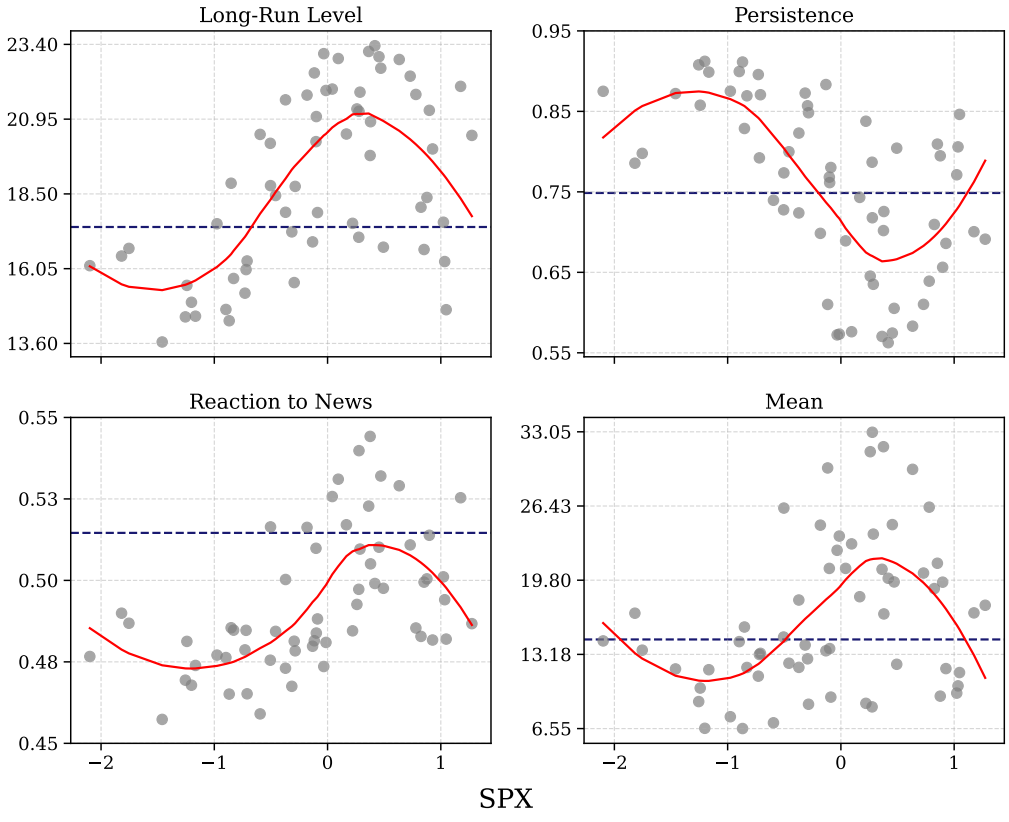


Figure S.8: **The estimated ACD tree model.** This figure depicts the tree structure for the ACD model. The tree's splits are based on SPX, T10Y and RVOLM, which refer to the S&P 500 return, the 10 year Treasury bond return and the monthly average of realized variance respectively.