6
Monte-Carlo Study

The goal of this section is threefold. The first one is to evaluate the small sample properties of the QMLE. Second, the modeling cycle strategy proposed in the paper is evaluated in small and moderate samples. Finally, the approximation capabilities of the Tree-MM are analyzed when the model is misspecified (the parametric form of the true data generating process is not a Tree-MM).

In order to attain the first two goals we simulate four different models[1]. The simulated models are:

– Model 1
A linear AR(2) model.

$$y_t = 1.0 + 0.5y_{t-1} - 0.2y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim \mathsf{NID}(0, 1.0).$$

– Model 2
A Tree-MM model with two AR(4) local models, $\mathbb{J} = \{0\}$ and $\mathbb{T} = \{1, 2\}$. The gating function is $g_0(\cdot; \gamma_0, c_0) = g(y_{t-4}; 2, 3)$ and the experts are

$$\begin{aligned} y_{1t} &= 2.0 - 0.1y_{t-1} + 0.7y_{t-2} + 0.2y_{t-4} + \varepsilon_{1t}, \quad \varepsilon_{1t} \sim \mathsf{NID}(0, 1.0) \\ y_{2t} &= 2.0 + 0.2y_{t-1} - 0.6y_{t-2} + 0.3y_{t-3} - 0.3y_{t-4} + \varepsilon_{2t}, \quad \varepsilon_{2t} \sim \mathsf{NID}(0, 0.6). \end{aligned}$$

– Model 3
A Tree-MM model with three local AR(2) models, $\mathbb{J} = \{0, 2\}$ and $\mathbb{T} = \{1, 5, 6\}$. The gating functions are $g_0(\cdot; \gamma_0, c_0) = g(y_{t-2}; 2, 1)$ and $g_2(\cdot; \gamma_2, c_2) = g(y_{t-1}; 2, 4)$. The experts are given by:

$$\begin{aligned} y_{1t} &= 0.5 - 0.4y_{t-1} + 0.7y_{t-2} + \varepsilon_{1t}, \quad \varepsilon_{1t} \sim \mathsf{NID}(0, 1.0) \\ y_{5t} &= 4.0 + 0.8y_{t-1} - 0.5y_{t-2} + \varepsilon_{5t}, \quad \varepsilon_{5t} \sim \mathsf{NID}(0, 0.6) \\ y_{6t} &= 8.0 - 0.9y_{t-1} + 0.2y_{t-2} + \varepsilon_{6t}, \quad \varepsilon_{6t} \sim \mathsf{NID}(0, 1.1). \end{aligned}$$

– Model 4
A Tree-MM model with four local AR(2) models, $\mathbb{J} = \{0, 1, 2\}$ and

[1] We discard the 50 first observations to avoid any initialization effects.

$\mathbb{T} = \{3, 4, 5, 6\}$. The gating functions are: $g_0(\cdot; \gamma_0, c_0) = g(y_{t-2}; 1, 1)$, $g_1(\cdot; \gamma_1, c_1) = g(y_{t-1}; 3, 0)$ and $g_2(\cdot; \gamma_2, c_2) = g(y_{t-1}; 2, 4)$. The experts are defined as

$$\begin{aligned} y_{3t} &= 0.7y_{t-1} - 0.3y_{t-2} + \varepsilon_{3t}, \quad \varepsilon_{3t} \sim \mathsf{NID}(0, 0.7) \\ y_{4t} &= -0.5 - 0.4y_{t-1} + 0.7y_{t-2} + \varepsilon_{4t}, \quad \varepsilon_{4t} \sim \mathsf{NID}(0, 1.0) \\ y_{5t} &= 4.0 + 0.8y_{t-1} - 0.5y_{t-2} + \varepsilon_{5t}, \quad \varepsilon_{5t} \sim \mathsf{NID}(0, 0.6) \\ y_{6t} &= 8.0 - 0.9y_{t-1} + 0.2y_{t-2} + \varepsilon_{6t}, \quad \varepsilon_{6t} \sim \mathsf{NID}(0, 1.1). \end{aligned}$$

6.1 Parameter estimation

In this section, we present the empirical results of the estimation of the parameters across the 2000 simulations. Two measures are used to evaluate the variability of the estimates: the sample standard deviation and, as a more robust alternative, the median absolute deviation around the median (MAD) defined as

$$MAD(\widehat{\theta}) = \text{median}\left(\left|\theta - \text{median}(\widehat{\theta})\right|\right). \quad (6\text{-}1)$$

We have simulated Models 2–4 with two different sample sizes: 150 and 500 observations. Tables 6.1–6.3 show estimation results for each model. From the tables, it is easily seen that the estimation turns to be rather precise, with the only exception of the slope parameter γ, which is usually overestimated. This overestimation were noticed in (47), and it is caused due the lack of observations around the transition location.

6.2 Specification Algorithm.

In this section we evaluate the performance of the specification algorithm. We evaluate four models (Models 1–4), and each model is simulated 200 times with two sample sizes: 150 and 500 observations. Table 6.4 presents the results. We define the model to be correctly specified if the sets $\mathbb{J}$, $\mathbb{T}$ and $\mathbb{S} = \{s_0, \ldots, s_{\#\mathbb{J}}\}$ are equal to the true sets $\mathbb{J}_0$, $\mathbb{T}_0$ and $\mathbb{S}_0$. The tree is incorrectly specified if any of these sets are different.

The BIC has a better performance then AIC in small and large samples. Furthermore, the building cycle has a much better performance in large samples for both information criteria, but the performance in small samples is still fairly good.

Tabela 6.1: SIMULATED MODEL 2: DESCRIPTIVE STATISTICS OF THE ESTIMATES.

The table shows the mean, standard deviation, median, and median absolute deviation the estimates of the parameters of Model 2 over 2000 simulations. 150 and 500 observations are considered.

		150				500			
Parameter	Actual	Mean	Std. Dev.	Median	MAD	Mean	Std. Dev.	Median	MAD
γ_0	2	5.06	1.75	4.68	0.89	5.01	1.65	4.66	0.83
c_0	3.00	3.01	0.23	3.02	0.15	3.00	0.22	3.00	0.15
σ_1^2	1.00	0.95	0.14	0.94	0.09	0.95	0.13	0.95	0.09
β_{01}	2.00	2.00	0.11	2.00	0.07	2.00	0.10	2.00	0.07
β_{11}	-0.10	-0.10	0.04	-0.10	0.03	-0.10	0.04	-0.10	0.03
β_{21}	0.70	0.70	0.04	0.70	0.02	0.70	0.04	0.70	0.02
β_{31}	0	0.00	0.05	0.00	0.03	0.00	0.04	0.00	0.03
β_{41}	0.20	0.20	0.06	0.20	0.04	0.20	0.06	0.20	0.04
σ_2^2	0.60	0.31	0.08	0.31	0.05	0.30	0.08	0.31	0.05
β_{02}	2.00	1.99	0.36	1.99	0.23	2.01	0.34	2.01	0.20
β_{12}	0.20	0.20	0.05	0.20	0.03	0.02	0.72	0.22	0.08
β_{22}	-0.60	-0.60	0.06	-0.60	0.04	-0.60	0.06	-0.60	0.03
β_{32}	0.30	0.03	0.05	0.30	0.03	0.03	0.04	0.30	0.03
β_{42}	-0.30	-0.30	0.06	-0.30	0.04	-0.30	0.04	-0.30	0.04

Tabela 6.2: SIMULATED MODEL 3: DESCRIPTIVE STATISTICS OF THE ESTIMATES.

The table shows the mean, standard deviation, median, and median absolute deviation the estimates of the parameters of Model 3 over 2000 simulations. 150 and 500 observations are considered.

		150				500			
Parameter	Actual	Mean	Std. Dev.	Median	MAD	Mean	Std. Dev.	Median	MAD
γ_0	2.00	5.41	2.25	4.98	1.25	4.99	1.17	4.86	0.66
c_0	1.00	0.97	0.35	0.97	0.18	0.94	0.20	0.94	0.12
σ_1^2	1.00	0.87	0.44	0.87	0.18	0.97	0.15	0.96	0.10
β_{01}	0.50	0.63	0.98	0.51	0.19	0.51	0.15	0.40	0.09
β_{11}	-0.40	-0.39	0.27	-0.41	0.08	-0.40	0.06	-0.40	0.04
β_{21}	0.70	0.62	0.48	0.66	0.10	0.68	0.08	0.69	0.05
γ_2	2.00	5.47	2.48	4.89	1.36	5.21	1.64	4.85	0.88
c_2	4.00	3.92	0.42	3.95	0.14	3.92	0.2	3.93	0.13
σ_5^2	0.60	0.34	0.09	0.33	0.05	0.35	0.04	0.35	0.03
β_{05}	4.00	3.99	0.24	4.00	0.14	4.00	0.11	4.00	0.07
β_{15}	0.80	0.80	0.06	0.80	0.04	0.80	0.03	0.80	0.02
β_{25}	-0.50	-0.50	0.05	-0.50	0.04	-0.50	0.03	-0.50	0.02
σ_6^2	1.10	1.13	0.34	1.11	0.18	1.21	0.15	1.20	0.10
β_{06}	8.00	8.09	1.48	1.09	0.79	8.07	0.62	8.06	0.39
β_{16}	-0.90	-0.90	0.21	-0.90	0.11	-0.90	0.09	-0.90	0.05
β_{26}	0.20	0.17	0.22	0.18	0.11	0.18	0.09	0.18	0.06

Tabela 6.3: SIMULATED MODEL 4: DESCRIPTIVE STATISTICS OF THE ESTIMATES.

The table shows the mean, standard deviation, median, and median absolute deviation the estimates of the parameters of Model 4 over 2000 simulations. 150 and 500 observations are considered.

		150				500			
Parameter	Actual	Mean	Std. Dev.	Median	MAD	Mean	Std. Dev.	Median	MAD
γ_0	1.00	3.41	1.62	3.07	0.67	3.38	1.14	3.16	0.52
c_0	1.00	0.98	0.42	0.94	0.19	0.96	0.23	0.94	0.12
γ_1	3.00	6.07	3.29	5.51	2.31	6.39	3.02	6.12	2.27
c_1	0.00	0.04	0.49	0.04	0.20	0.01	0.34	0.02	0.15
γ_2	2.00	5.44	3.04	4.69	2.02	5.13	2.86	4.43	1.91
c_2	4.00	3.48	0.87	3.68	0.40	3.59	0.67	3.74	0.29
σ_3^2	0.70	0.45	0.29	0.42	0.08	0.47	0.17	0.46	0.05
β_{03}	0.00	-0.02	0.35	-0.02	0.17	-0.02	0.17	-0.01	0.09
β_{13}	0.70	0.68	0.16	0.69	0.08	0.69	0.08	0.69	0.04
β_{23}	-0.30	-0.31	0.10	-0.31	0.05	-0.31	0.05	-0.31	0.03
σ_4^2	1.00	0.87	0.37	0.85	0.18	0.95	0.16	0.94	0.10
β_{04}	-0.50	-0.56	0.46	-0.57	0.28	-0.53	0.22	-0.53	0.15
β_{14}	-0.40	-0.40	0.17	-0.40	0.10	-0.40	0.08	-0.40	0.05
β_{24}	0.70	0.67	0.17	0.67	0.10	0.68	0.09	0.68	0.05
σ_5^2	0.60	0.33	0.14	0.31	0.06	0.34	0.05	0.34	0.03
β_{05}	4.00	4.00	0.19	4.00	0.11	4.00	0.08	4.00	0.05
β_{15}	0.80	0.80	0.05	0.80	0.03	0.80	0.02	0.80	0.02
β_{25}	-0.50	-0.50	0.05	-0.50	0.02	-0.50	0.02	-0.50	0.01
σ_6^2	1.10	1.14	0.67	1.04	0.31	1.28	0.37	1.23	0.18
β_{06}	8.00	7.94	2.17	8.00	1.08	7.81	1.08	7.91	0.57
β_{16}	-0.90	-0.86	0.37	-0.88	0.17	-0.84	0.18	-0.86	0.09
β_{26}	0.20	0.12	0.29	0.13	0.15	0.16	0.12	0.16	0.07

Tabela 6.4: SPECIFICATION ALGORITHM.

This table shows the number of cases where the each model is correctly/uncorrectly specified. We consider two different samples: 150 and 500 observations. Both the AIC and BIC are used to select the models.

	150				500			
	AIC		BIC		AIC		BIC	
Model	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
Model 1	163	37	172	28	197	3	200	0
Model 2	107	93	134	66	193	7	196	4
Model 3	83	117	96	104	150	50	166	34
Model 4	57	143	81	119	123	77	135	65

6.3 Approximation Capabilities

In this section we will illustrate the ability of the Tree-MM model to approximate the unknown conditional probability density functions. We simulate two AR(1)-GARCH(1,1) models and two neural network (NN) models. We generate 2000 observations, where the first 1000 are used for estimation and the remaining 1000 for out-of sample evaluation.

We use the coverage test (16) over a set of percentiles to evaluate the coverage. The test is applied to the out-of-sample period. The correlation and the mean squared error (MSE) of the one-step-ahead predictions are also used to compare the Tree-MM models with the true data generation process. The Christoffersen test consists in two likelihood ratio (LR) tests. The first one is the LR test of unconditional coverage and the second one the LR test of independence. All the AR(1)-GARCH(1,1) models have the same linear part and distinct GARCH(1,1) conditional variances. The linear model is:

$$y_t = 0.7y_{t-1} + u_t,$$

where $u_t = h_t^{1/2}\epsilon_t$, $\epsilon_t \sim \mathsf{NID}(0,1)$, and

– Model 1: GARCH(1,1)

$$h_t = 10^{-5} + 0.85h_{t-1} + 0.05u_{t-1}^2$$

– Model 2: GARCH(1,1)

$$h_t = 10^{-5} + 0.90h_{t-1} + 0.085u_{t-1}^2$$

We evaluate the conditional coverage over the following percentiles: 90%, 95%, 97.5% and 99%. Table 6.5 shows the results. LR_{uc} is the p-value of the unconditional coverage test and LR_{cc} is the p-value of the conditional coverage test.

The estimated trees usually have three leaves. The first regime is identified by very negative values of y_t, the second by near zero values of y_t and the third one by very positive values of y_t.

Now we turn to the neural network examples. We simulated a neural network with two distinct conditional variances: one constant and another one following a

Tabela 6.5: COVERAGE FOR EVALUATED AR(1)-GARCH(1,1) MODELS. The table shows the empirical coverage and the p-value of the Christoffersen test for the estimated AR(1)-GARCH(1,1) models. LR_{uc} is the p-value of the unconditional coverage test and LR_{cc} is the p-value of the conditional coverage test.

Model	Percentile	Empirical Coverage			
		90.0	95.0	97.5	99.0
Model 1	Est. Percentile	90.39	95.60	97.60	99.20
	LR_{uc}	.6954	.3890	1.000	.5769
	LR_{cc}	.2234	.1841	.8587	.0000
Model 2	Est. Percentile	90.76	95.51	97.71	99.33
	LR_{uc}	.4574	.9177	.2145	.2728
	LR_{cc}	.5418	.2276	.1183	.2729

GARCH(1,1) process. The simulated models are the following:

$$\begin{aligned} y_t &= 0.1 + 0.75y_{t-1} - 0.05y_{t-4} \\ &\quad +0.8g(0.45y_{t-1} - 0.89y_{t-4}; 2.24, -0.09) \\ &\quad -0.7g(0.44y_{t-1} + 0.89y_{t-4}; 1.12, -0.35) + u_t \end{aligned}$$

where $u_t = h_t^{1/2}\epsilon_t$, $\epsilon_t \sim \mathsf{NID}(0,1)$, and

- Model 1: $h_t = 1$

- Model 2: $h_t = 10^{-5} + 0.85h_{t-1} + 0.05\epsilon_{t-1}^2$.

Table 6.6 compares the out-of-sample performance of the estimated Tree-MM model with the true Neural Network specification. The correlation row shows the average correlation between the estimates, MSE_{NN} and $\text{MSE}_{\text{Tree-MM}}$ are the average out-of-sample MSE for the NN and Tree-MM models, respectively. Table 6.7 shows the coverage probability as well as the p-values of the conditional and unconditional tests of (16). From the results in the tables we can see that the correlation between the estimates are high and the MSEs are very close for both models, showing the approximation capabilities of the Tree-MM models. The estimated coverage probabilities are close to the nominal ones.

Tabela 6.6: FORECASTING PERFORMANCE RESULTS. The table shows the forecasting results and the correlation between the true data generating process and the estimated Tree-MM model.

Model	Correlation	MSE_{NN}	$\text{MSE}_{\text{Tree-MM}}$
Model 1	0.88	0.0223	0.0242
Model 2	0.74	2.25×10^{-3}	2.81×10^{-3}

Tabela 6.7: COVERAGE FOR EVALUATED NEURAL NETWORKS MODELS.

This table shows the empirical coverage and the p-values of the unconditional and conditional tests of (16).

Model		Empirical Coverage							
	Percentile	50.0	60.0	70.0	80.0	90.0	95.0	97.5	99.0
Model 1	Est. Percentile	52.10	62.20	70.30	79.70	89.30	95.00	97.40	99.20
	LR_{uc}	.8065	.3235	.4726	.5336	.3344	.6878	1.000	.5297
	LR_{cc}	.9224	.3761	.5964	.2020	.2939	.6664	.3583	.2349
Model 2	Est. Percentile	51.45	61.35	71.00	81.70	90.60	95.40	97.10	99.00
	LR_{uc}	.4707	.1629	.3348	.0673	.5352	.5735	.4385	1.000
	LR_{cc}	.5704	.4340	.2338	.1336	.6199	.6018	.5317	0.000