

Threshold heterogeneous autoregressive modeling for realized volatility

Sein Moon^a, Minsu Park^a, Changryong Baek^{1,a}

^aDepartment of Statistics, Sungkyunkwan University

Abstract

The heterogeneous autoregressive (HAR) model is a simple linear model that is commonly used to explain long memory in the realized volatility. However, as realized volatility has more complicated features such as conditional heteroscedasticity, leverage effect, and volatility clustering, it is necessary to extend the simple HAR model. Therefore, to better incorporate the stylized facts, we propose a threshold HAR model with GARCH errors, namely the THAR-GARCH model. That is, the THAR-GARCH model is a nonlinear model whose coefficients vary according to a threshold value, and the conditional heteroscedasticity is explained through the GARCH errors. Model parameters are estimated using an iterative weighted least squares estimation method. Our simulation study supports the consistency of the iterative estimation method. In addition, we show that the proposed THAR-GARCH model has better forecasting power by applying to the realized volatility of major 21 stock indices around the world.

Keywords: heterogeneous autoregressive (HAR), threshold HAR, realized volatility, GARCH

1. 서론

변동성(volatility)은 자산의 가격이 일정 기간 동안 얼마나 변화하였는가를 나타내는 것으로, 변동성이 크다는 것은 일반적으로 시장으로 유입되는 정보의 양이 많아짐을 의미한다. 금융 시계열에서 변동성에 대한 정확한 추정과 예측은 옵션 거래와 리스크 관리 등 여러 가지 측면에서 중요한 역할을 한다. 예를 들어 Kim과 Yoon (2015)은 KOSPI 200 옵션시장에서 변동성의 상승과 하락에 따라 외국인 기관투자자의 옵션 매도와 매수 형태가 바뀐다는 것을 발견하였다. 또한 Brooks와 Persaud (2003)는 금융 시장에서 변동성의 중요성을 다시 한번 강조하고, 리스크 관리 측면에서 여러 가지 변동성 모델의 예측 성능을 비교하였다. 이처럼 변동성은 금융 시장을 이해하는데 가장 핵심적인 역할을 수행한다. 변동성을 측정하는 방법 중 로그-수익률(log-return)은 계산과 비교가 편리하다는 이유에서 널리 쓰이고 있다. 관측 시점 t 에서 자산가격 P_t 에 대한 로그-수익률은

$$r_t = \log P_t - \log P_{t-1}$$

으로 정의되며, 보통 일별 자료에 대한 로그-수익률을 통해 변동성에 대한 측도로 널리 쓰이고 있다.

This work was supported by the Basic Science Research Program from the National Research Foundation of Korea (NRF-2022R1F1A1066209).

¹ Corresponding author : Department of Statistics, Sungkyunkwan University, 25-2 Sungkyunkwan-ro, Jongno-gu, Seoul 03063, Korea. E-mail : crbaek@skku.edu

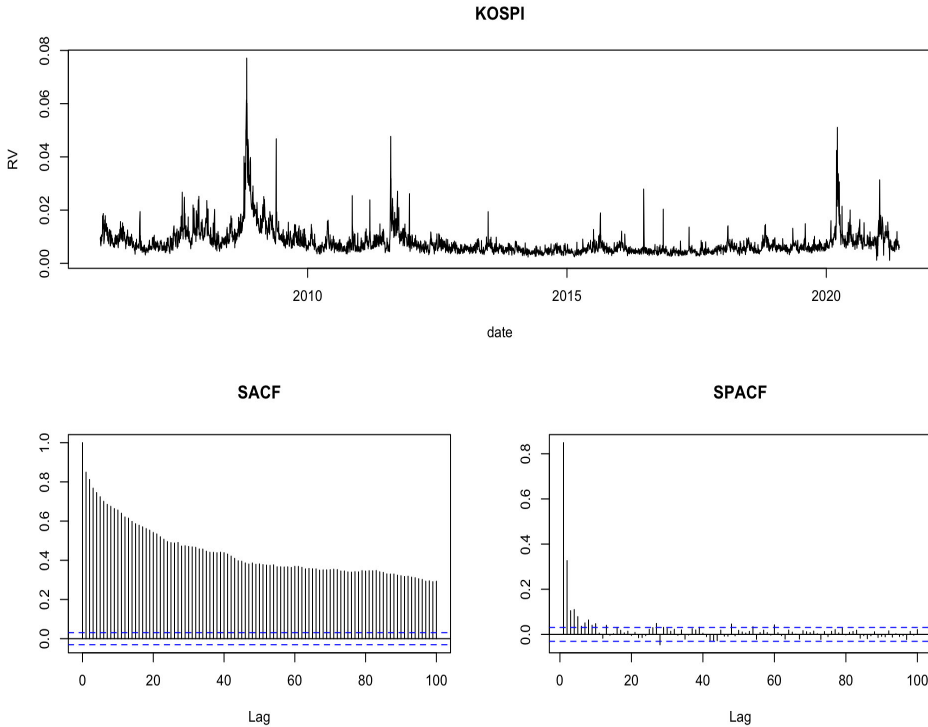


Figure 1: *RV, SACF, and SPACF plot of KOSPI data.*

하지만, 최근 데이터를 실시간으로 수집하는 기술이 발달하면서 고빈도 금융 자료를 활용하여 관측되지 않는 변동성을 추정하는 연구가 활발히 진행되고 있다. 구체적으로 Andersen 등 (2003)은 로그-수익률을 기반으로 일중(intraday) 자료를 활용한 실현 변동성(realized volatility; RV)을 변동성 추정치로 제안했다. 실현 변동성은 하루를 M 개의 등간격으로 나누어 얻은 로그-수익률 제곱합의 제곱근이며, 다음과 같이 나타낸다.

$$RV_t^d = \sqrt{\sum_{j=0}^{M-1} r_{t-j\Delta}^2}, \quad (1.1)$$

여기에서 $\Delta = 1d/M$ 이다. 실현 변동성은 다른 금융 시계열 데이터와 마찬가지로 두꺼운 꼬리 분포, 조건부 이분산성(conditional heteroscedasticity), 장기기억성(long memory), 레버리지 효과 및 변동성 집중(volatility clustering)과 같은 정형화된 사실(stylized facts)을 가짐이 잘 알려져 있다. 두꺼운 꼬리 분포는 통상 로그변환을 통해서 실증자료 분석을 하고 있어 실현 변동성에 대한 통계적인 모형은 장기기억성과 이분산성, 레버리지 효과를 반영하는 모형을 찾는 데 집중되고 있다.

예를 들어 Figure 1은 2006년 1월 4일부터 2021년 5월 31일까지 5분 간격으로 집계된 KOSPI 지수의 실현 변동성에 대한 그래프이다. 2008년에는 서브프라임 모기지(subprime mortgage) 사태로 인한 세계 금융 위기, 2011년에는 유럽 재정 위기, 2020년에는 코로나 팬데믹 등 여러 경제 위기로 인해 큰 변동성을 보여 이분산성이 한눈에 보인다. 또한 표본자기상관함수(SACF) 그래프를 보면 시차가 커지더라도 그 값이 매우 천천히 감소하고 있는 것으로 미루어 보아 실현 변동성이 장기기억성을 가진다고 볼 수 있다.

Corsi (2009)는 이러한 장기기억성을 잘 반영하고 있는 HAR (heterogeneous autoregressive) 모형을 제안했다. HAR 모형은 일일, 주간, 월간 변동성들의 합으로 이루어진 간단한 선형 모형으로 추정이 간편하고 예측력이 뛰어나다는 장점이 있다. 또한 Corsi 등 (2008)에서는 일반적인 HAR 모형의 잔차에 조건부 이분산성과 변동성 집중 현상(volatility clustering)이 있음을 발견하고, 이를 설명하기 위해 HAR 모형의 오차항에 GARCH 모형을 적합시켜 HAR-GARCH 모형을 제안했다. 그러나 금융 시계열에서 변동성은 음의 수익률과 양의 수익률에 대해서 비대칭성을 보이고 있고 음의 수익률이 더 강하게 시장에 영향을 주는 레버리지 효과를 보인다. 다시 말해 호재와 악재에 있어 다른 영향력을 받는 특징이 있고 비선형 모형을 적합시킨다면 이러한 특징을 보다 잘 반영할 것이다.

대표적인 비선형 모형으로는 Tong (1978) 및 Lim과 Tong (1980)에서 제안한 임계자기회귀모형(threshold autoregressive; TAR) 모형이 있다. 이들은 감소와 증가 패턴에 비대칭성을 고려하여 임계값을 기준으로 서로 다른 선형 모형 적합시켰다. Pypko (2015)는 실현 변동성의 변화에도 비대칭성이 있음을 발견하고, TAR 모형을 기반으로 국면에 따라 모형을 다르게 하여 비선형 구조를 가지는 임계 HAR (threshold HAR; THAR) 모형을 제안했다. 또한 S&P 500 지수의 실현 변동성 데이터를 이용하여 실제로 금융 위기가 있을 경우에 비선형 모형이 더 정확한 예측력을 가짐을 보였다. 마찬가지로 Jawadi 등 (2020)는 에너지 시장에서 석유와 가스의 가격 변동성이 내·외부적인 정보나 거래량에 따라 비대칭적이고 비선형적임을 발견하고, 이를 반영한 threshold augmented HAR 모형을 제안하여 예측력을 높였다. 또한 Motegi 등 (2020)는 이동 평균 임계 HAR (moving average threshold HAR; MAT-HAR)제안하고 일본의 산업생산지수 자료를 이용하여 HAR모형과 TAR모형을 결합한 모형이 예측력이 더 뛰어나다는 것을 밝혔다.

본 논문에서는 장기기억성 및 조건부 이분산성과 비선형성을 가지는 실현 변동성 모형을 보다 효과적으로 다루기 위해서 GARCH 오차를 가지는 임계 HAR (THAR-GARCH) 모형을 제안하고 모형의 유용성에 대해서 연구하자고 한다. 본 논문의 구성은 다음과 같다. 제 2장은 THAR-GARCH모형과 추정 방법에 대하여 설명한다. 제 3장에서는 모의실험을 통해 2장에서 제시한 추정 방법이 적절하지 확인한다. 제 4장에서는 세계 각국의 주요 21개 실현 변동성 데이터 분석을 통해 제안한 THAR-GARCH 모형의 예측 성능을 비교하였다. 마지막으로 제 5장에서는 본 논문에서 제안한 모형에 대한 결과를 정리하고 추후에 있을 연구에 대해 논의하였다.

2. 방법론

금융 시장의 변동성을 분석하기 위한 기본적인 HAR 모형은 다음과 같다.

$$Y_t^d = c + \phi^d Y_{t-1}^d + \phi^w Y_{t-1}^w + \phi^m Y_{t-1}^m + a_t, \quad (2.1)$$

여기에서 a_t 는 오차항이며, Y_t^d , Y_t^w , Y_t^m 는 각각 일일 변동성, 주간 변동성, 월간 변동성으로 주간 변동성과 월간 변동성은 주식 시장이 열리는 평균 일수가 주 5일과 월 22일임을 고려하여 다음과 같이 나타낸다.

$$Y_{t-1}^w = \frac{1}{5} \sum_{j=1}^5 Y_{t-j}^d, \quad Y_{t-1}^m = \frac{1}{22} \sum_{i=1}^{22} Y_{t-i}^d. \quad (2.2)$$

HAR 모형의 계수인 ϕ^d , ϕ^w , ϕ^m 은 약정상성을 만족하며 일반적으로 오차항 a_t 는 평균이 0이고 분산이 1인 백색 잡음 과정을 따른다고 가정한다. 그러나 Corsi 등 (2008)는 실현 변동성에서 확인되는 이분산성 및 변동성 집중 현상을 고려하여 HAR-GARCH(p, q) 모형을 제안했다. 구체적으로 HAR-GARCH(p, q) 모형은 오차항에

Table 1: MSE for DGP1

n	ϕ_1^d	ϕ_1^w	ϕ_1^m	ϕ_2^d	ϕ_2^w	ϕ_2^m	ω	α	β
	0.5	0.3	0.1	0.4	0.15	0.2	0.1	0.15	0.6
200	0.0151	0.0355	0.0399	0.0338	0.0753	0.0685	0.0094	0.0195	0.0879
500	0.0048	0.0097	0.0090	0.0096	0.0237	0.0204	0.0089	0.0194	0.0833
1000	0.0021	0.0042	0.0040	0.0039	0.0106	0.0093	0.0087	0.0185	0.0804

Table 2: MSE for DPG2

n	ϕ_1^d	ϕ_1^w	ϕ_1^m	ϕ_2^d	ϕ_2^w	ϕ_2^m	ω	α	β
	0.5	0.3	0.1	0.4	0.15	0.2	0.1	0.1	0.8
200	0.0146	0.0293	0.0394	0.0557	0.0694	0.0699	0.0720	0.0089	0.0685
500	0.0047	0.0090	0.0091	0.0081	0.0230	0.0187	0.0652	0.0084	0.0678
1000	0.0022	0.0046	0.0046	0.0042	0.0102	0.0091	0.0459	0.0081	0.0536

Table 3: MSE for DPG3

n	ϕ_1^d	ϕ_1^w	ϕ_1^m	ϕ_2^d	ϕ_2^w	ϕ_2^m	ω	α	β
	0.65	0.15	0.1	0.25	0.1	0.05	0.1	0.15	0.6
200	0.0103	0.0186	0.0288	0.0547	0.0856	0.0689	0.0092	0.0198	0.0989
500	0.0038	0.0084	0.0087	0.0135	0.0248	0.0247	0.0087	0.0191	0.0983
1000	0.0020	0.0038	0.0039	0.0057	0.0115	0.0101	0.0078	0.0190	0.0964

Table 4: MSE for DPG4

n	ϕ_1^d	ϕ_1^w	ϕ_1^m	ϕ_2^d	ϕ_2^w	ϕ_2^m	ω	α	β
	0.65	0.15	0.1	0.25	0.1	0.05	0.1	0.1	0.8
200	0.0104	0.0227	0.0318	0.0467	0.0952	0.0654	0.0653	0.0087	0.0638
500	0.0041	0.0081	0.0079	0.0135	0.0237	0.0231	0.0604	0.0085	0.626
1000	0.0023	0.0031	0.0036	0.0085	0.0132	0.0113	0.0549	0.0082	0.0584

다음과 같은 가정이 추가된다.

$$a_t = \sigma_t \epsilon_t,$$

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i a_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2, \quad (2.3)$$

여기에서 ϵ_t 는 평균이 0이고 분산이 1인 백색 잡음 과정을 가지며, a_t 의 정상성을 위해서 $\omega > 0$, $\alpha_i \geq 0$, $\beta_i \geq 0$ 와 $\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) < 1$ 의 조건을 가정한다. 예를 들어 일반적으로 널리 사용되는 GARCH(1,1) 모형의 식과 조건은 다음과 같다.

$$\sigma_t^2 = \omega + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad 0 \leq \alpha_1, \beta_1 \leq 1, (\alpha_1 + \beta_1) < 1. \quad (2.4)$$

HAR 모형은 제한된(constrained) AR(22) 모형이라고 생각할 수 있으며 선형모형이다. HAR 모형은 간단하면서도 예측력이 뛰어나다는 장점이 있지만 갑작스럽게 바뀌는 국면(regime)에 대해 설명하기엔 제한적이다. 따라서 국면에 따라 변화를 반영할 수 있는 비선형모형이 필요로 한다. 대표적인 비선형 모형으로는

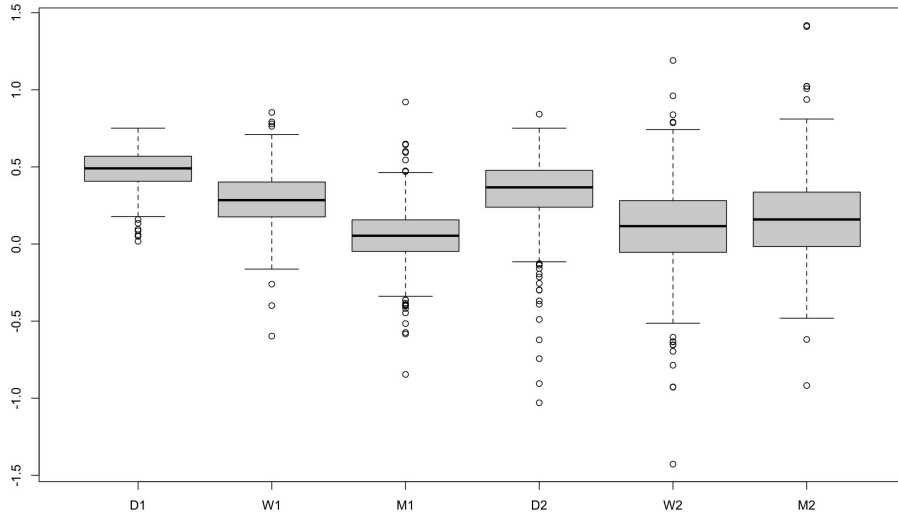


Figure 2: Boxplot of estimates in DGPI, $n = 200$.

Tong (1978), Lim과 Tong (1980)가 제시한 임계자기회귀 모형(threshold autoregressive; TAR)이 있으며 형태는 다음과 같다.

$$Y_t = \begin{cases} \beta_0^1 + \beta_1^1 Y_{t-1} + \dots + \beta_p^1 Y_{t-p} + \epsilon_t & \text{if } Z_{t-d} \leq \gamma, \\ \beta_0^2 + \beta_1^2 Y_{t-1} + \dots + \beta_p^2 Y_{t-p} + \epsilon_t & \text{if } Z_{t-d} > \gamma, \end{cases} \quad (2.5)$$

여기에서 Z_{t-d} 는 임계 변수(threshold variable)로 외생 변수나 $(Y_{t-1}, \dots, Y_{t-p})$ 의 함수가 될 수 있으며, 이를 기준으로 두 개의 식으로 나뉠 수 있다. d 는 지연 모수(delay parameter)이고 γ 는 임계값(threshold value)으로 모두 실수이다. Pypko (2015)는 HAR 모형에 TAR 모형과 같이 영향력을 달리하는 THAR 모형을 제안했다. 간단하게 2-국면 THAR 모형은 다음과 같이 나타낼 수 있다.

$$Y_t^d = \begin{cases} c_1 + \phi_1^d Y_{t-1}^d + \phi_1^w Y_{t-1}^w + \phi_1^m Y_{t-1}^m + a_t & \text{if } Z_{t-d} \leq \gamma, \\ c_2 + \phi_2^d Y_{t-1}^d + \phi_2^w Y_{t-1}^w + \phi_2^m Y_{t-1}^m + a_t & \text{if } Z_{t-d} > \gamma. \end{cases} \quad (2.6)$$

본 논문에서는 기본 HAR 모형(2.1)과 THAR 모형(2.6)에 이분산성을 고려하여 오차항 a_t 가 GARCH 모형을 따르는 모형으로 확장하고자 한다. 다만, GARCH(1,1) 모형이 여러 복잡한 GARCH 확장 모형에 비교해서 결코 뒤쳐지지 않는다는 Hansen과 Lunde (2005) 연구에 기반하여 GARCH(1,1)을 모의실험 및 실증자료 분석에서 고려하였다. 또한 임계 변수로 Y_{t-1} 과 식 (2.2)에서 정의한 Y_{t-1}^w, Y_{t-1}^m 을 고려하고자 한다.

먼저 THAR-모형의 계수를 추정하기 위해서는 Hwang과 Hong (2021)에서와 같이 가중최소제곱추정법(weighted least squares; WLS)을 사용하였다. 우선 모형(2.6)를 0 또는 1의 값을 갖는 지시 함수 $I(\cdot)$ 를 이용하면 다음과 같이 표현할 수 있다.

$$Y_t^d = (c_1 + \phi_1^d Y_{t-1}^d + \phi_1^w Y_{t-1}^w + \phi_1^m Y_{t-1}^m) I(Z_{t-d} \leq \gamma) + (c_2 + \phi_2^d Y_{t-1}^d + \phi_2^w Y_{t-1}^w + \phi_2^m Y_{t-1}^m) I(Z_{t-d} > \gamma) + a_t. \quad (2.7)$$

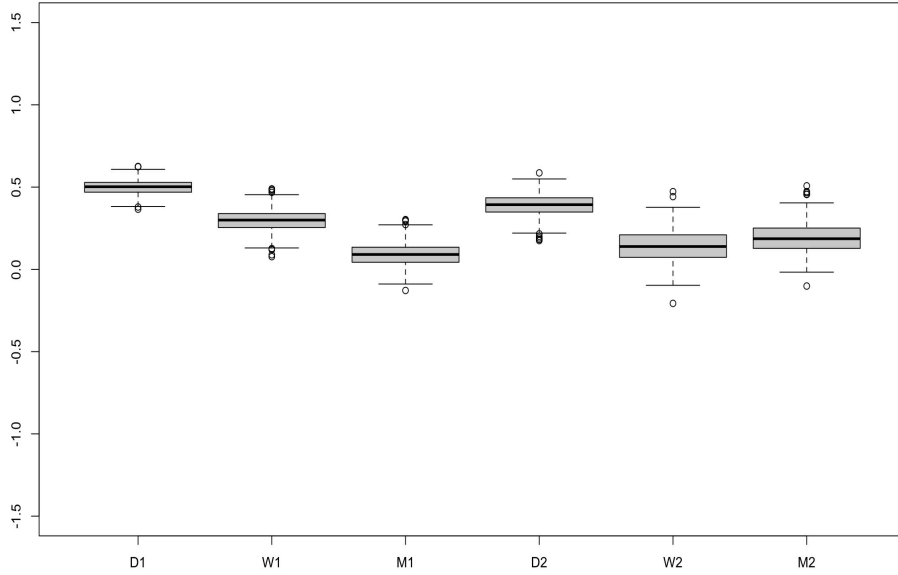


Figure 3: Boxplot of estimates in DGP1, $n = 1000$.

이를 회귀방정식 $Y = X\beta + \epsilon$ 로 나타내면, $Y = (Y_{23}^d, Y_{24}^d, \dots, Y_n^d)'$, $\beta = (c_1, \phi_1^d, \phi_1^w, \phi_1^m, c_2, \phi_2^d, \phi_2^w, \phi_2^m)$, $\epsilon = (a_{23}, a_{24}, \dots, a_n)$ 이고 $I(Z_{t-d} \leq \gamma)$ 와 $I(Z_{t-d} > \gamma)$ 를 각각 $I^{(1)}$ 과 $I^{(2)}$ 라고 하면 디자인 행렬(design matrix) X 는 다음과 같이 주어진다.

$$X = \begin{pmatrix} I_{22}^{(1)} & Y_{22}^d I_{22}^{(1)} & Y_{22}^w I_{22}^{(1)} & Y_{22}^m I_{22}^{(1)} & I_{22}^{(2)} & Y_{22}^d I_{22}^{(2)} & Y_{22}^w I_{22}^{(2)} & Y_{22}^m I_{22}^{(2)} \\ I_{23}^{(1)} & Y_{23}^d I_{23}^{(1)} & Y_{23}^w I_{23}^{(1)} & Y_{23}^m I_{23}^{(1)} & I_{23}^{(2)} & Y_{23}^d I_{23}^{(2)} & Y_{23}^w I_{23}^{(2)} & Y_{23}^m I_{23}^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ I_{n-1}^{(1)} & Y_{n-1}^d I_{n-1}^{(1)} & Y_{n-1}^w I_{n-1}^{(1)} & Y_{n-1}^m I_{n-1}^{(1)} & I_{n-1}^{(2)} & Y_{n-1}^d I_{n-1}^{(2)} & Y_{n-1}^w I_{n-1}^{(2)} & Y_{n-1}^m I_{n-1}^{(2)} \end{pmatrix}$$

주어진 임계값 γ 에 대해서, 본 논문이 제안한 THAR-GARCH 모형의 추정은 다음과 같이 반복적인 과정을 통해서 이루어진다.

Step 1 : 일반적인 최소제곱추정법(ordinary least square squares; OLS)를 이용하여 초기값 $\hat{\beta}^{(0)}$ 를 추정한다.

$$\hat{\beta}^{(0)} = (X'X)^{-1} X'Y.$$

Step 2 : 잔차 $\hat{\epsilon} = Y - X\hat{\beta}^{(0)}$ 를 구하고, 이 잔차가 조건부 평균이 0인 GARCH 모형을 따름을 이용하여 조건부 분산 $\hat{\sigma}_t^2$ 를 추정한다.

Step 3 : 가중치를 $\hat{w}_t = 1/\hat{\sigma}_t^2$ 라고 두고 가중최소제곱법을 이용하여 $\hat{\beta}^{(k)}$ 를 추정한다.

$$\hat{\beta}^{(k)} = (X'WX)^{-1} X'WY,$$

이때 W 는 대각 행렬 $\text{diag}(\hat{\sigma}_{22}^2, \dots, \hat{\sigma}_n^2)^{-1}$ 이다.

Step 4 : Steps 2와 3을 반복하여 구한 $\hat{\beta}^{(k+1)}$ 가 사전에 정의된 유의수준 ε 에 대해서

$$|\hat{\beta}^{(k+1)} - \hat{\beta}^{(k)}| \leq \varepsilon$$

를 만족하면 $\hat{\beta}^{(k)}$ 를 최종 추정된 계수로 선택한다.

Remark. 임계값 γ 에 대한 추정은 기본적으로 격자 탐색(grid search)을 이용하여 잔차제곱합을 가장 작게 만드는 값을 찾는다. 예를 들어 먼저 초기값 γ_0 을 설정한 후, 위의 설명한 방법에 따라 THAR-GARCH 모형의 모수 $\hat{\beta}^{(0)}$ 을 추정한다. 이 후 격자 탐색(grid search)을 통해 잔차제곱합을 가장 작게 만드는 $\hat{\gamma}^{(1)}$ 를 찾고 이를 이용하여 계수 $\hat{\beta}^{(1)}$ 를 업데이트한다. 이러한 과정을 $\gamma^{(s+1)} \approx \hat{\gamma}^{(s)}$ 까지 반복하여 최종 γ 를 선택한다. 금융 변동성에 대한 정형화된 사실이 0에 대한 비대칭성을 가지고, 기존의 연구 자료에서도 임계값을 추정할 경우 0에 매우 가깝게 나와, 예를 들어 Pypko (2015), 본 논문에서 사용한 모의실험과 실증 분석에서는 γ 추정을 생략하고 0으로 설정하였다.

3. 모의 실험

본 장에서는 가중최소제곱추정법을 이용한 THAR-GARCH(1, 1) 모형의 계수 추정이 적절한지 확인하기 위해 모의실험을 진행하였다. 추정해야하는 THAR 모형의 계수로는 $\phi = (\phi_1^d, \phi_1^v, \phi_1^m, \phi_2^d, \phi_2^v, \phi_2^m)$ 가 있고, GARCH (1, 1) 모형의 계수로는 $\theta = (\omega, \alpha, \beta)$ 가 있다. 임계값을 기준으로 HAR 모형의 계수는 약정상성을 만족하되 장기기억성을 보이게 선택하였고, GARCH (1, 1) 모형의 경우도 일치성을 위해 수식 (2.4)을 만족하는 변수 중에 $\alpha + \beta$ 값이 높은 영속성(high-persistence)를 갖도록 변수를 선택하였다. 또한 임계값을 기준으로 계수 값의 변화가 크지 않은 경우 $\phi = (0.5, 0.3, 0.1, 0.4, 0.15, 0.2)$ 와 비교적 큰 차이가 나는 경우 $\phi = (0.65, 0.15, 0.1, 0.25, 0.1, 0.05)$ 두 경우를 고려하여 추정력을 비교하였다. 또한 GARCH 모형의 모수인 θ 도 $\theta = (0.1, 0.15, 0.6)$ 와 $\theta = (0.1, 0.1, 0.8)$ 두 가지를 고려하여 영속성에 따라 2가지 경우를 고려하여, 모의실험에서는 총 4가지의 data generating process (DGP)를 다음과 같이 사용하였다.

DGP 1 $\phi = (0.5, 0.3, 0.1, 0.4, 0.15, 0.2)$, $\theta = (0.1, 0.15, 0.6)$,

DGP 2 $\phi = (0.5, 0.3, 0.1, 0.4, 0.15, 0.2)$, $\theta = (0.1, 0.1, 0.8)$,

DGP 3 $\phi = (0.65, 0.15, 0.1, 0.25, 0.1, 0.05)$, $\theta = (0.1, 0.15, 0.6)$,

DGP 4 $\phi = (0.65, 0.15, 0.1, 0.25, 0.1, 0.05)$, $\theta = (0.1, 0.1, 0.8)$.

임계 변수는 하루 전 시점의 변동성인 일일 변동성 Y_{t-1} 로 선택하였으며, 임계값은 0으로 두었다. 표본수는 200, 500, 1,000개를 사용하였고 추정 성능을 비교하기 위한 기준으로 평균제곱오차(mean square error; MSE)를 사용하였다. MSE 추정을 위한 모의실험 반복수는 500이며 Tables 1-4에 결과를 정리하였다. 먼저, 모든 경우에 대해서 표본수가 증가할수록 MSE값이 0으로 감소하는 것을 확인할 수 있어 제안한 추정 방법이 일치추정량임을 확인할 수 있다. 또한 DGP1과 DGP2, 혹은 DGP3과 DGP4의 비교를 통해서 GARCH 모수의 $\alpha + \beta$ 가 증가하여 이분산성에 대한 종속성이 다소 증가하더라도 THAR 모형의 계수가 잘 추정됨을 살펴볼 수 있어 GARCH 모형의 확장이 추정을 어렵게 만들지는 않음을 알 수 있다.

추정에 대한 변동폭을 보다 면밀히 살펴보기 위해 Figures 2와 3은 DGP1에서, Figures 4와 5는 DGP3에서 표본수가 각각 200과 1,000일 때 얻은 계수 추정치들을 상자 그림으로 나타냈다. 여기에서 D1, W1, M1, D2, W2, M2는 각각 $\phi_1^d, \phi_1^v, \phi_1^m, \phi_2^d, \phi_2^v, \phi_2^m$ 을 뜻한다. 표본수가 증가함에 따라 상자의 크기가 줄어들며, 이상치들도 평균에 더 가까워지는 것을 확인할 수 있다.

다음으로는 모의실험을 통해 기존의 HAR-GARCH 모형과 비교하여 제안한 THAR-GARCH 모형의 예측 우수성을 보이고자 한다. 예측력을 비교하기 위한 척도로는 1단계 표본 외 예측 방법을 사용하여 구한

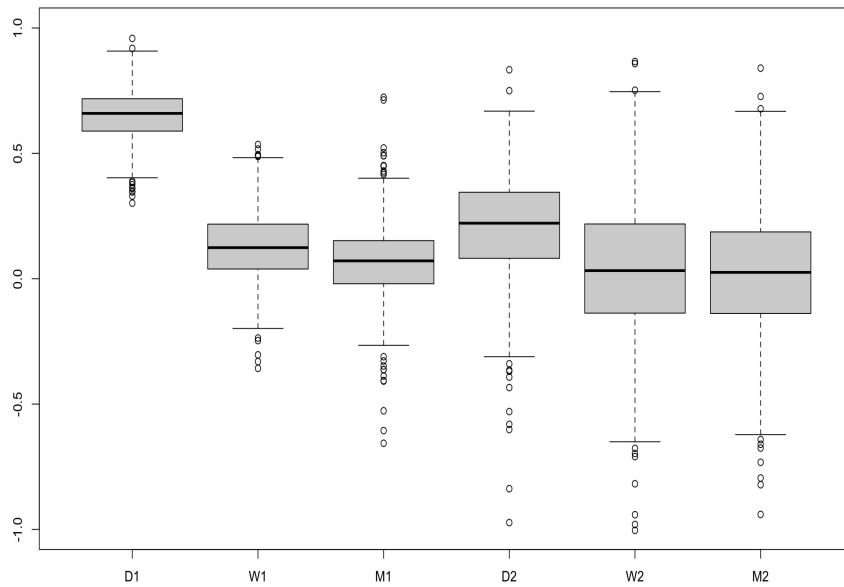


Figure 4: Boxplot of estimates in DGP3, $n = 200$.

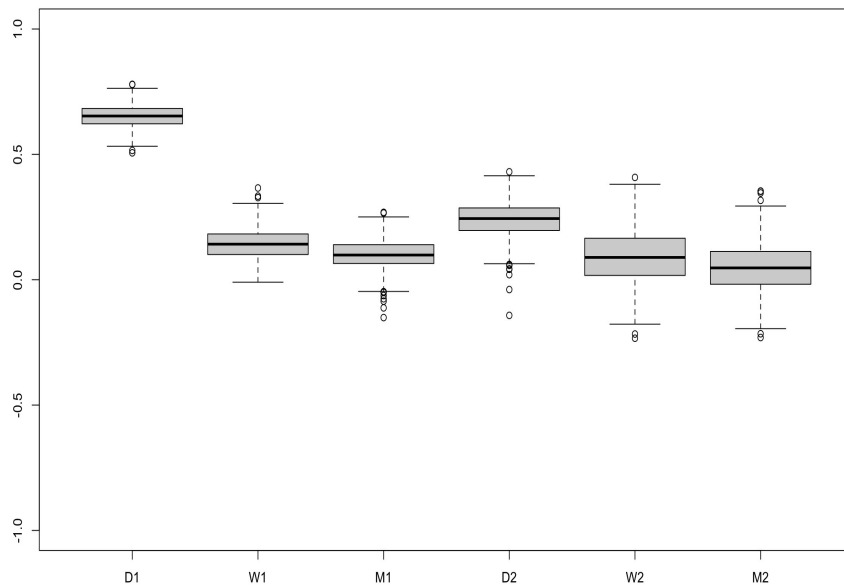


Figure 5: Boxplot of estimates in DGP3, $n = 1000$.

예측값들을 가지고 평균 제곱 예측 오차(mean squared prediction error; MSPE)를 이용하였다. 1단계 표본 외 예측 방법은 예측하고자 하는 시점의 한 시점 전까지의 데이터를 이용하여 모형을 적합한 후 예측하고자 하는

Table 5: 1-step-ahead out-of-sample forecasting comparison in terms of MSPE

	DGP1	DGP2	DGP3	DGP4
HAR-GARCH	0.4362	1.0245	0.40560	1.0185
THAR-GARCH	0.4311	1.0236	0.04029	1.0044

Table 6: Summary statistics for 21 RVs around the world

Name	Mean	Median	Maximum	Minimum	SD	Skewness	Kurtosis
AEX	0.0079	0.0067	0.0648	0.0017	0.0050	4.2625	35.0414
All Ordinaries	0.0065	0.0053	0.0691	0.0018	0.0046	5.5147	53.9577
BEL20	0.0079	0.0068	0.0607	0.0020	0.0047	3.9356	28.9121
IBOVESPA	0.0094	0.0085	0.0636	0.0023	0.0049	4.5221	37.0838
DJIA	0.0073	0.0058	0.0772	0.0014	0.0058	4.4546	36.1897
CAC40	0.0085	0.0072	0.0660	0.0021	0.0052	3.7541	28.0961
FTSE100	0.0083	0.0069	0.0817	0.0012	0.0056	4.8030	43.4703
DAX	0.0084	0.0074	0.0558	0.0020	0.0048	3.2253	21.6704
S&P/TSX	0.0056	0.0046	0.0607	0.0012	0.0040	5.4121	52.7562
Hang Seng	0.0078	0.0071	0.0487	0.0021	0.0035	3.2878	24.2646
IBEX35	0.0095	0.0082	0.0742	0.0031	0.0051	4.0021	33.9969
NASDAQ	0.0079	0.0063	0.0772	0.0016	0.0055	3.6047	27.6389
KOSPI	0.0065	0.0057	0.0511	0.0012	0.0034	4.5146	38.7311
IPC Mexico	0.0074	0.0066	0.0723	0.0024	0.0035	5.1298	69.8808
Nikkei	0.0072	0.0060	0.0620	0.0014	0.0046	3.8655	28.4311
NIFTY50	0.0071	0.0061	0.0966	0.0011	0.0047	7.5642	103.3148
Russel2000	0.0079	0.0065	0.0555	0.0017	0.0051	3.4892	23.4861
S&P500	0.0070	0.0054	0.0644	0.0011	0.0056	4.1215	31.2837
SSEC	0.0094	0.0074	0.0644	0.0021	0.0066	3.2904	19.0470
SMI	0.0069	0.0059	0.0745	0.0025	0.0048	7.0628	73.6283
EURO STOXX50	0.0091	0.0077	0.0735	0.0001	0.0060	3.9654	30.3934

시점의 예측값을 계산하며, MSPE의 구체적인 식은 다음과 같다.

$$\text{MSPE} = \frac{1}{T_0} \sum_{t=T-T_0+1}^T (Y_t - \hat{Y}_t)^2, \quad (3.1)$$

여기에서 T 는 전체 자료의 개수, T_0 는 검증자료의 개수이며 \hat{Y}_t 는 1부터 $(t-1)$ 시점까지의 자료를 이용해 1단계 표본의 예측을 한 값이다. 모의 실험에서는 각 DGP에서 대해서 $T = 500$ 시점의 데이터를 생성했으며, 이 중에서 400개의 시점을 학습 데이터로, $T_0 = 100$ 개의 시점을 시험용 데이터로 설정하였다. Table 5는 각 모형에 대한 MSPE값을 나타내며, THAR-GARCH 모형에서의 MSPE값이 작은 것을 보아 임계값을 고려한 자료에서 더 좋은 예측력을 보인다는 것을 확인하였다.

4. 실증자료분석

본 장에서는 HAR-GARCH 모형과 THAR-GARCH 모형의 예측력을 비교하고자 한다. 실증 자료로는 Oxford-Man Institute of Quantitative Finance (OMIQF; <http://realized.oxfordman.ox.ac.uk>)에서 제공하는 세계 각국의 주요 증권 지수의 실현 커널(realized kernel) 변동성을 사용하였다. 실현 커널 변동성은 시장의 미시구조에 의한 오차(market microstructure noise)를 줄이는 실현 변동성 (1.1)의 개선된 추정치로 Barndorff-Nielsen

Table 7: MSPE for 1-step-ahead out-of-sample forecasts

Name	HAR-GARCH	THAR_D	THAR_D-GARCH	THAR_W-GARCH	THAR_M-GARCH
AEX	0.105448	0.105169	0.104889	0.106401	0.105680
All Ordinaries	0.113596	0.113403	0.113382	0.114974	0.114500
BEL20	0.113779	0.113311	0.113126	0.114812	0.113907
IBOVESPA	0.062550	0.062484	0.062484	0.062166	0.062418
DJIA	0.061465	0.059839	0.059757	0.060049	0.061452
CAC40	0.116266	0.116561	0.116492	0.118177	0.117162
FTSE100	0.149640	0.149966	0.150089	0.150363	0.149630
DAX	0.077106	0.076811	0.076922	0.076567	0.076579
S&P/TSX	0.088152	0.085132	0.084807	0.085512	0.085959
Hang Seng	0.061467	0.060769	0.060135	0.059936	0.060381
IBEX35	0.088421	0.087066	0.087331	0.088636	0.087914
NASDAQ	0.068657	0.068987	0.068621	0.067996	0.067057
KOSPI	0.042591	0.041639	0.041548	0.043359	0.042893
IPC Mexico	0.063748	0.063984	0.064107	0.063785	0.064343
Nikkei	0.063918	0.064228	0.064638	0.066349	0.064369
NIFTY50	0.061627	0.062671	0.062600	0.061631	0.061720
Russel2000	0.082908	0.082596	0.082445	0.083149	0.082274
S&P500	0.066742	0.065944	0.065138	0.065549	0.066249
SSEC	0.072236	0.072111	0.071408	0.072226	0.073547
SMI	0.059801	0.060663	0.060501	0.059514	0.059692
EURO STOXX50	0.115670	0.117169	0.117100	0.116890	0.116099
Sum	1.735787	1.729143	1.728520	1.738042	1.733827

등 (2008)에 의해 제안된 뒤 실현 변동성 모형에 있어서 가장 널리쓰이는 측도이다. OMIQF는 실현 변동성을 구하기 위해 5분 단위로 집계된 로그 수익률이 사용한다. 분석한 구간은 구체적으로 2015년 01월 1일부터 2022년 5월 31일까지 총 1,934시점이며 세계 각국의 주요 증권 지수인 AEX, All Ordinaries, BEL20, IBOVESPA, DJIA, CAC40, FTSE100, DAX, S&P/TSX, Hang Seng, IBEX35, NASDAQ, KOSPI, IPC Mexico, Nikkei, Nifty50, Russell2000, S&P500, SSEC, SMI, Euro STOXX50 총 21개의 지수의 실현 변동성을 데이터로 사용하였다. 이 중에서 2015년 1월 1일부터 2021년 12월 31일까지 총 1,827시점을 학습 데이터로 사용하여 모형에 적합시켰으며, 2022년 1월 3일부터 2022년 5월 31일까지 총 107시점을 시험 데이터로 사용하여 예측력을 구하였다. 평일임에도 불구하고 각 나라의 공휴일로 인해 주식 시장이 열리지 않아 생기는 결측값들은 선형 보간법을 이용하여 채워주었다. Table 6는 사용한 데이터들의 요약통계량을 나타낸 것으로 양의 왜도 값을 갖는 것을 보아 자료가 왼쪽에 치우쳐져 비대칭성을 보임을 알 수 있다. 또한, 3 이상의 높은 첨도 값을 갖는 것을 보아 자료들이 꼬리가 두꺼운 첨예 분포(leptokurtic) 형태를 가짐을 알 수 있다. 따라서 실현 변동성 데이터에 로그 변환한 후 중심화 한 데이터를 가지고 분석을 진행하였다.

예측력을 비교하는 척도로는 모의실험에서 사용한 식 (3.1)의 MSPE를 사용하였다. Table 7은 각 증권 지수의 실현 변동성에 대하여 HAR-GARCH(1, 1) 모형과 THAR-GARCH(1, 1) 모형에 적합시켜 1단계 표본 외 예측 방법을 통해 구한 MSPE 결과이다. 추가로 임계 변수를 일일 변동성 뿐만 아니라 주간 변동성과 월간 변동성을 함께 고려하여 비교하였다. 중심화된 데이터를 사용하였으므로 임계값은 0으로 하였다. 예를 들어 THAR_D-GARCH는 일변량인 Y_{t-1} 의 값이 양수인지 음수인지에 따라서 THAR-GARCH(1, 1) 모형을 적합한 결과이다. 비슷하게 THAR_W-GARCH는 주간변량인 Y_{t-1}^w 가 임계값인 0을 기준으로 양수인지 음수인지에 따라 변화하는 THAR-GARCH 모형을 적합하였을 때의 표본 외 예측 오차 값을 나타낸다. 오차항에 GARCH

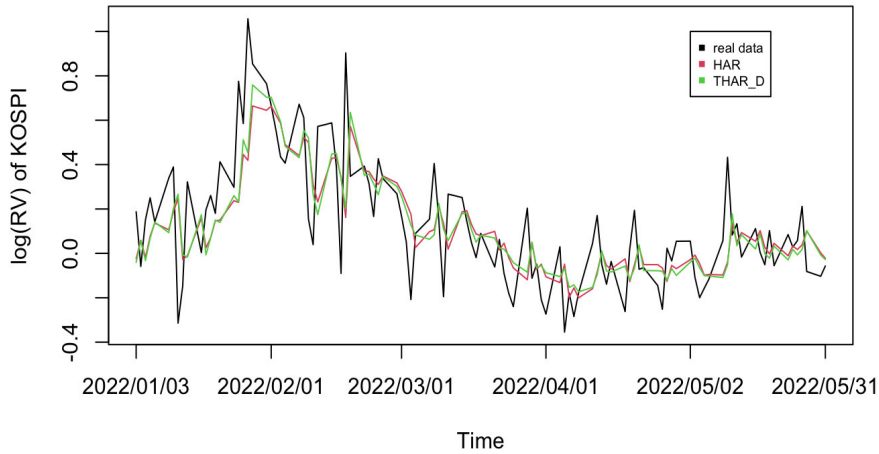


Figure 6: Forecasting plot of KOSPI RV.

가정을 추가한 모형들을 비교하기에 앞서, 오차항이 평균이 0이고 분산인 1인 백색잡음 과정을 따르는 일반적인 THAR_D 모형과의 비교를 통해 오차항의 GARCH 가정이 적절한지 알아보았다.

그 결과 CAC40, IPC Mexico, Nikkei, NIFTY50, EURO STOXX50의 경우 임계값을 고려하지 않은 HAR-GARCH(1, 1) 모형이 가장 우수한 예측력을 보였으며 나머지 16개의 주가지수에 대해서는 임계값을 고려한 THAR-GARCH 모형이 더 좋은 예측력을 보였다. 임계값을 정하는 변수로는 AEX, All Ordinaries, BEL20, DJIA, S&P/TSX, IBEX35, KOSPI, S&P500, SSEC는 일변량인 Y_{t-1} 을 썼을 때 가장 좋은 예측력을 보였다. 주간 변동량인 Y_{t-1}^w 를 쓸 경우에는 IBOVESPA, DAX, Hang Seng과 SMI이 가장 좋은 예측력을 보였으며 월변량인 Y_{t-1}^m 의 경우 FTSE100, NASDAQ과 Russel2000이 가장 좋은 예측력을 보였다. 주가 지수에 따른 차이는 있지만 일반적으로 임계점을 고려하고 GARCH 오차를 고려한 모형이 더 좋은 예측 성능을 보임을 알 수 있다. 또한 임계점을 선택하는 기준으로는 대체적으로 전 시점에서의 일일 실현 변동성을 이용하는 경우 더 높은 예측력을 보이는 경우가 많았다. GARCH 오차의 사용 여부에 대해서 THAR_D 모형과 비교해 보면, THAR_D 모형이 전반적으로 HAR-GARCH 모형에 비해서 MSPE를 작게 만들어 주기는 하였으나, THAR_D-GARCH 모형에 비해서는 성능이 낮았고 모든 모형 중에서 가장 작은 MSPE를 주는 주가 지수는 없었다. 따라서 THAR 모형에서도 금융시장의 조건부 이분산성을 고려한 GARCH 오차항을 넣는 모형이 더 좋은 성능을 보임을 확인할 수 있었다. Figure 6은 2022년 1월 3일부터 2022년 5월 31일까지 KOSPI 지수의 일일 실현 변동성의 HAR-GARCH(1, 1) 모형과 일일 변동성을 임계 변수로 설정한 THAR-GARCH(1, 1) 모형의 예측 결과 그래프이다. 그림에서도 THAR-GARCH(1, 1) 모형이 변동이 큰 peak에서 좀 더 실측값에 일치하는 예측값을 주는 것을 관찰할 수 있다.

5. 결론 및 논의점

실현 변동성은 두꺼운 꼬리 분포, 변동성 군집, 장기기억성 등 여러 가지 특성들을 가지고 있다. 본 논문에서는 이러한 실현 변동성의 특징을 고려하여 비선형 모형인 THAR-GARCH 모형을 제안하고 추정 방법 및 예측에서의 성능을 살펴보았다. 모형의 계수 추정을 위해서는 조건부 분산을 가중치로 두는 가중최소제곱추정법을 이용하였으며, 모의실험을 통해 여러 경우에서도 적절한 방법임을 확인하였다. 또한 세계 각국의 주요 증권 지수의 실현 변동성을 사용하여 실증 분석한 결과 예측력을 높이는 모형임을 알 수 있었다. 특히 임계 변수를

일일 변동성으로 설정하였을 때 주로 높은 예측력을 보였다.

본 논문에서 제안한 THAR-GARCH 모형의 임계 변수는 THAR 모형의 변수인 일일, 주간 및 월간 변동성을 고려하였으나 다른 거시 경제 변수 등도 충분히 고려해 볼 만하다. 또한 GARCH 오차에 대해서도 변화점에 따라 GARCH 모형의 계수가 변화하는 모형을 생각해 본다면 금융시장이 가지는 역동성(dynamic)을 보다 잘 반영할 수 있을 것이다. 또한 고차원 THAR-GARCH 모형 등으로의 확장도 생각해 볼 수 있어 후속 연구를 통해 THAR 모형의 확장을 기대해 본다.

References

- Andersen TG, Bollerslev T, Diebold FX, and Labys P (2003). Modeling and forecasting realized volatility, *Econometrica*, **71**, 579–625.
- Barndorff-Nielsen OE, Hansen PR, Lunde A, and Shephard N (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise, *Econometrica*, **76**, 1481–1536.
- Brooks C and Persaud G (2003). Volatility forecasting for risk management, *Journal of Forecasting*, **22**, 1–22.
- Corsi F (2009). A simple approximate long-memory model of realized volatility, *Journal of Financial Econometrics*, **7**, 174–196.
- Corsi F, Mittnik S, Pigorsch C, and Pigorsch U (2008). The volatility of realized volatility, *Econometric Reviews*, **27**, 46–78.
- Hansen PR and Lunde A (2005). A forecast comparison of volatility models: Does anything beat a GARCH (1, 1)?, *Journal of Applied Econometrics*, **20**, 873–889.
- Hwang E and Hong WT (2021). A multivariate HAR-RV model with heteroscedastic errors and its WLS estimation, *Economics Letters*, **203**, 109855.
- Jawadi F, Ftiti Z, and Louhichi W (2020). Forecasting energy futures volatility with threshold augmented heterogeneous autoregressive jump models, *Econometric Reviews*, **39**, 54–70.
- Kim SJ and Yoon SJ (2015). Institutional investors' volatility trading in the koshi 200 options market, *The Korean Journal of Financial Management*, **32**, 1–33.
- Lim K and Tong H (1980). Threshold autoregressions, limit cycles, and data, *Journal of the Royal Statistical Society, B*, **42**, 245–92.
- Motegi K, Cai X, Hamori S, and Xu H (2020). Moving average threshold heterogeneous autoregressive (mat-har) models, *Journal of Forecasting*, **39**, 1035–1042.
- Pypko S (2015). Volatility forecast in crises and expansions, *Journal of Risk and Financial Management*, **8**, 311–336.
- Tong H (1978). *On a threshold model*, In Chen C (Eds), *Pattern Recognition and Signal Processing* (pp. 575-586), Sijthoff & Noordhoff, Netherlands.

Received February 1, 2023; Revised March 14, 2023; Accepted March 21, 2023

임계 HAR 모형을 이용한 실현 변동성 분석

문세인^a, 박민수^a, 백창룡^{1,a}

^a성균관대학교 통계학과

요약

HAR 모형은 간단한 선형 모형으로 실현 변동성의 장기기억성을 비교적 잘 설명할 수 있어 널리 쓰이고 있다. 하지만, 실현 변동성은 조건부 이분산성, 레버리지 효과, 변동성 집중 등과 같은 복잡한 특징을 보이고 있기에 단순 HAR 모형을 확장할 필요가 있다. 따라서 본 연구는 조건부 이분산성을 설명하는 GARCH 모형에 임계값에 따라 계수가 달라지는 비선형 모형인 임계 HAR 모형(THAR-GARCH)을 제안하고 그 추정 방법 및 예측 성능에 대해서 살펴보고자 한다. 보다 구체적으로 오차항의 등분산 가정을 벗어났기 때문에 모형의 계수를 추정하기 위해서 반복적인 가중최소제곱추정법을 제안하고 모의실험을 통해 일치성을 보였다. 또한 전 세계 21개의 주요 주가 지수의 실현 변동성에 대한 예측 오차를 비교함으로써 제안한 GARCH 오차를 가지는 임계 HAR 모형이 일반적으로 더 우수한 예측력을 보임을 확인하였다.

주요용어: HAR모형, THAR 모형, 실현 변동성, GARCH
