

Bitcoin'in Nabzını Tutmak

Saatlik Verilerle Bir Volatilite Avı

GARCH ailesi modelleriyle iki yıllık, saat saat bir Bitcoin hikâyesi

Emine Nur Aktaş — Yazılım Bilişim Akademisi, 4. Dönem Veri Analitiği Programı

1. Bir Sorudan Doğan Proje

Her bitirme projesinin bir başlangıç anı vardır. Benimki şuydu: Bitcoin fiyat grafiğine bakarken kendime sordum — "Bu dalgalanmanın kendisi, yani şiddeti, önceden tahmin edilebilir bir şey mi?"

Çoğu insan kripto para dendiğinde önce "fiyat ne olacak" diye düşünür. Ama finans dünyasında asıl kıymetli olan çoğu zaman farklı bir soru: *piyasa ne kadar sert oynayacak?* Bu sorunun cevabı, bir yatırımcının pozisyon büyüklüğünden bir borsanın teminat gereksinimine, bir risk yöneticisinin gecelik stres testinden bir türev ürünün fiyatlanmasına kadar pek çok kararın temelini oluşturuyor.

İşte bu proje, tam olarak bu ikinci soruya, yani **volatilite tahminine**, sistematik ve dürüst bir cevap arama girişimi. Sonunda ortaya "mükemmel" bir model çıkmadı — ama ortaya, ne işe yaradığını ve nerede zorlandığını açıkça bilen bir model çıktı. Bu makalede, o yolculuğun tamamını — verinin ilk indirildiği andan nihai modelin risk yönetiminde sınıldığı ana kadar — adım adım anlatacağım.

Kullanılan araçlar: Python, `yfinance`, `arch`, `statsmodels`, `matplotlib` **Veri:** BTC-USD, saatlik, 2 yıl (2024-08-20 → 2026-08-19) **Nihai model:** EGARCH(1,1) × Skewed-t



2. Fiyat mı, Volatilite mi? Bir Kavram Ayrımı

Projeye başlamadan önce netleştirmem gereken bir kavram karışıklığı vardı. "Volatilite tahmini" dediğimde çoğu kişi "yani fiyatı mı tahmin ediyorsun?" diye soruyor. Hayır — ve aradaki fark, aslında finansal ekonometrinin en temel derslerinden biri.

Fiyat tahmini, finans literatüründe genel kabul gören görüşe göre neredeyse imkânsızdır; piyasalar büyük ölçüde **rastgele yürüyüş** (random walk) sergiler — yani dünkü bilgiyle yarının fiyatını sistematik olarak öngörmek çok zordur. Ama **volatilitenin** kendine has bir hafızası vardır. Piyasa sakinse, genelde bir süre daha sakin kalır; sert bir dönem yaşıyorsa, o sertlik genelde birkaç gün ya da hafta sürer. Buna literatürde "**volatilite kümelenmesi**" deniyor, ve bu örüntü GARCH ailesi modellerinin üzerine inşa edildiği temel taş.

Yani ben bu projede "yarın Bitcoin kaç dolar olacak" sorusuna değil, "yarın piyasa ne kadar sert oynayacak" sorusuna cevap aradım. Teknik dille ifade edersek: amacım, BTC-USD saatlik logaritmik getirilerinin **koşullu varyansını** modellemek ve örneklem dışında test etmektir.



3. Kısa Bir Literatür Molası

Boş bir sayfadan başlamadım. Bitcoin volatilitesi üzerine yapılmış birkaç önemli çalışmaya baktım ve kendi bulgularımı bunlarla karşılaştırmayı planladım:

- **Katsiampa (2017)** — Bitcoin için EGARCH modelinin en iyi performansı verdiğini bulmuş. (Spoiler: ben de aynı sonuca ulaştım.)
- **Chu vd. (2017)** — Kripto para birimlerinde GARCH etkisinin genel olarak güçlü olduğunu göstermiş.
- **Bouri vd. (2017)** — Bitcoin'de kaldıraç (leverage) etkisinin zayıf olduğunu rapor etmiş.
- **Dyhrberg (2016)** — Bitcoin'in volatilité yapısı bakımından altına benzediğini öne sürmüştü.
- **Nelson (1991)** — EGARCH modelinin orijinal, temel makalesi.

Bu liste bana hem bir yol haritası hem de sonunda kendi bulgularımı tartışabileceğim bir referans noktası verdi. Nitekim ilerleyen bölümlerde göreceğiniz gibi, bazı noktalarda literatürle tam örtüştüm, bazı noktalarda (özellikle kaldıraç etkisinin gücü konusunda) küçük bir ayrıştım.



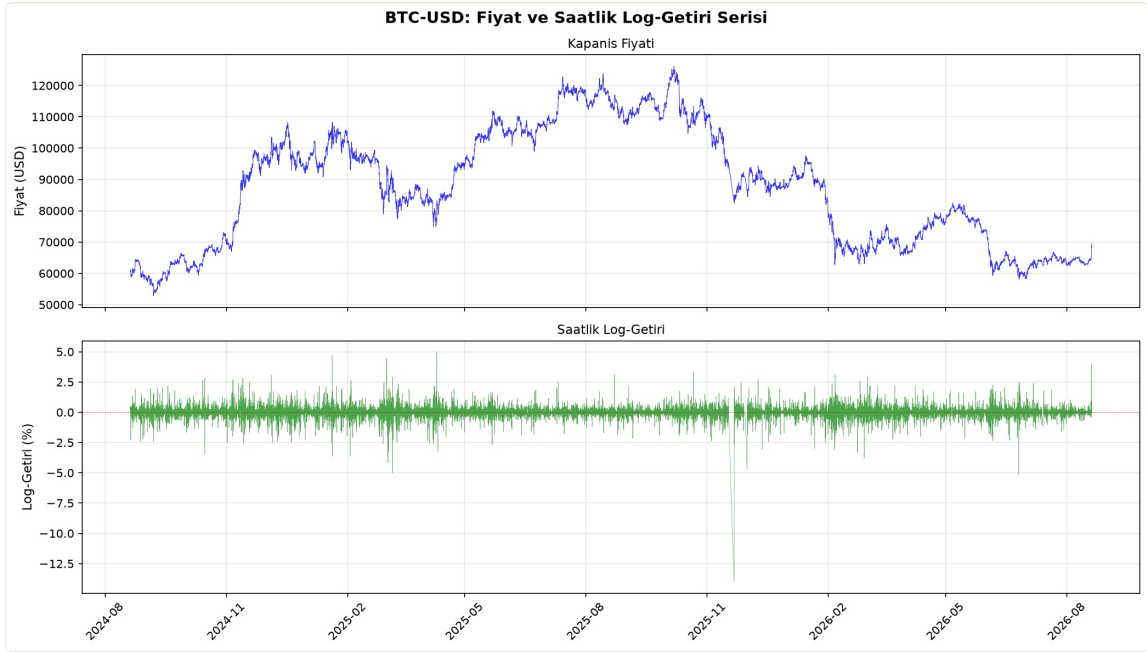
4. Veri: Ham Halinden Kullanılabilir Hale

BTC-USD paritesinin saatlik kapanış fiyatlarını `yfinance` kütüphanesi üzerinden indirdim — `interval='1h'`, `period='730d'` parametreleriyle, yani tam iki yıllık bir pencere. Ham veri 17.328 gözlemden oluşuyordu.

Gerçek dünyadaki her veri seti gibi, bu veri de kusursuz değildi. Saatlik zaman damgalarını taradığımda toplam 190 saatlik eksiklik buldum. Bunları ikiye ayırdım:

- **Küçük boşluklar (≤ 3 saat):** Toplam 2 saat — lineer interpolasyonla dolduruldu.
- **Büyük boşluklar (> 3 saat):** Toplam 188 saat, 5 ayrı olay — örneklemden tamamen çıkarıldı.

Neden büyük boşlukları doldurmadım? Çünkü uzun bir boşluğu yapay olarak doldurmaya çalışmak, gerçekte hiç yaşanmamış, düzgün bir getiri serisi "uydurmak" anlamına geliyor — bu da istatistikleri (özellikle *kurtosis*'i) yapay şekilde şişiriyor. Nitekim büyük boşluk interpolasyonunu kaldırdığımda fazladan kurtosis değeri 46,89'dan 9,82'ye düştü — yaklaşık 4,8 kat bir fark. Bu tek başına, veri temizleme kararlarının nihai istatistiksel sonuçları ne kadar değiştirebileceğinin çarpıcı bir kanıtı.



BTC-USD saatlik fiyat serisi (üst panel) ve log-getiri serisi (alt panel). Alt paneldeki keskin sivri uç, Kasım 2025 çöküşüne işaret ediyor.

Peki bu 5 büyük boşluk ne zaman, neden oldu? İşte burada işler ilginçleşiyor.



5. Verideki Sessizlikler: Rastgele mi, Değil mi?

Beş büyük boşluğu tek tek incelediğimde, aralarında çarpıcı bir örüntü buldum:

Tarih	Süre	Fiyat Değişimi	Tür
2025-11-17	100 saat	-%13,07	KRASH (gerçek piyasa olayı)
2025-11-28	20 saat	-%0,20	API kesintisi
2025-11-29	35 saat	-%4,66	İkinci dalga
2026-05-05	12 saat	+%0,10	API kesintisi
2026-07-13	21 saat	-%1,19	API kesintisi

17 Kasım 2025'te Bitcoin, sadece birkaç gün içinde 95.061 dolardan 82.640 dolara geriledi — yüzde 13'ün üzerinde bir düşüş. Ve tam bu kriz anında, **yfinance** servisi 100 saat boyunca veri sağlayamadı. 29 Kasım'daki ikinci boşluk da çöküşün ikinci dalgasıyla örtüşüyor. Geri kalan üç boşlukta ise büyük bir fiyat hareketi yok — muhtemelen sıradan teknik API kesintileri.

Bunun anlamı şu: **veri setindeki eksiklikler tamamen rastgele değil**. İstatistikte buna "missing not at random" (rastgele olmayan eksiklik) deniyor — kriz dönemlerinde veri sağlayıcı da sekteye uğruyor. Bu, modelin kriz anlarındaki dinamikleri tam olarak öğrenemediği anlamına geliyor ve bu projenin en önemli metodolojik sınırlılıklarından biri olarak raporumda ayrıca ele alındı.

Peki 3 saatlik eşik neden bu şekilde seçildi, keyfi bir karar mıydı? Hayır — raporumda ayrı bir duyarlılık analizi yaptım: eşiği 2, 4 ve 6 saat olarak değiştirdiğimde nihai model seçimi (EGARCH × Skewed-t) değişmedi. Bu, sonuçların bu belirli eşiğe aşırı duyarlı olmadığını, yani sağlam (robust) olduğunu gösteriyor.

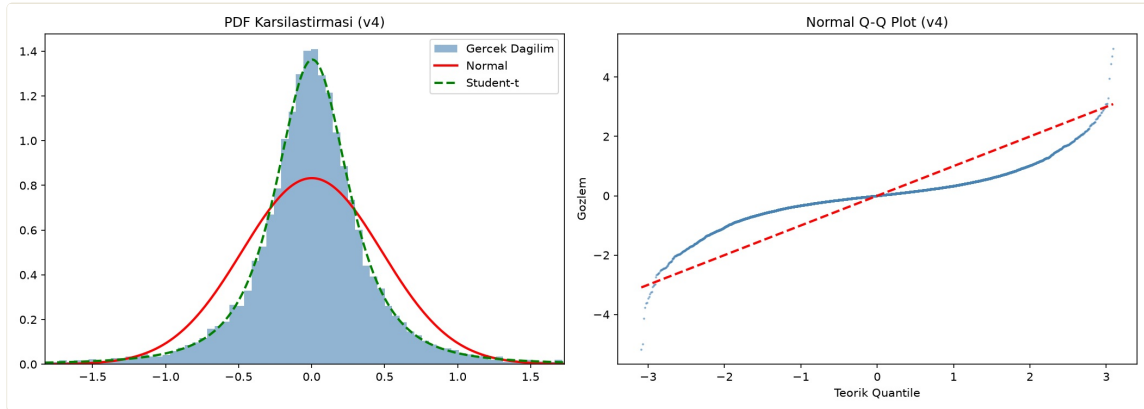


6. Bitcoin'in İstatistiksel Karakteri

Temiz veriye (17.324 gözlem) tanımlayıcı istatistiklerle baktığımda, kripto paralara özgü tanıdık bir profil ortaya çıktı:

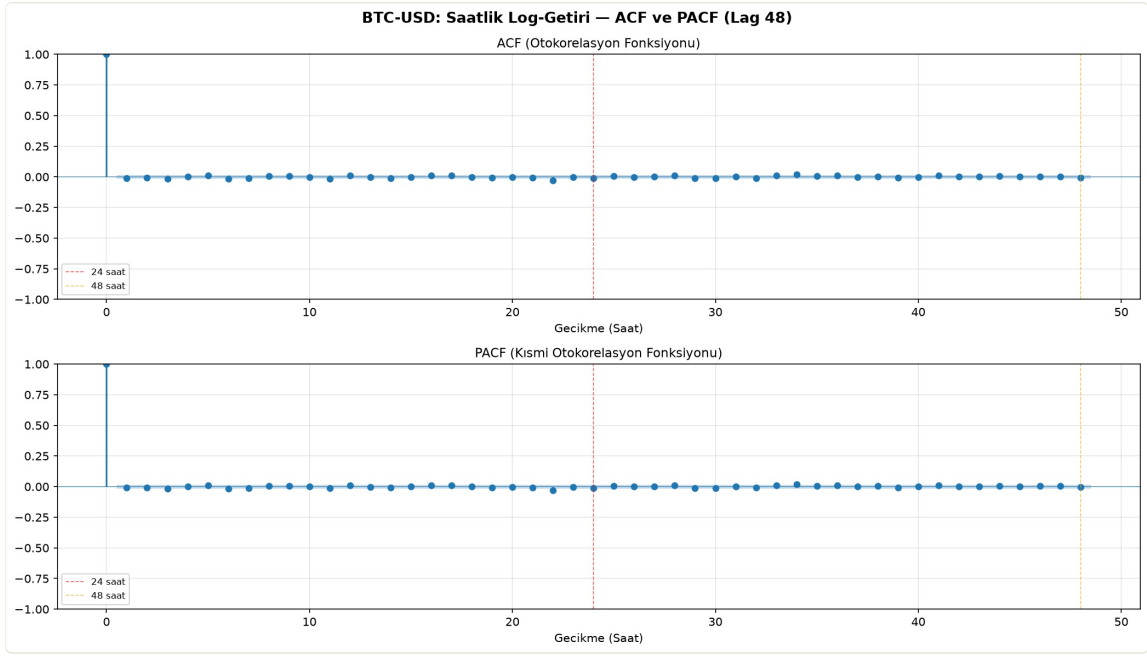
Metrik	Değer	Yorum
Ortalama	0,001970	Sıfıra yakın
Standart Sapma	0,4793	Saatlik ölçekte
Çarpıklık	-0,1669	Hafif sola çarpık
Fazladan Kurtosis	9,8225	Kalın kuyruklu (Normal = 0)
Jarque-Bera	69.679 (p≈0)	Normallik kesin biçimde reddedilir

Kurtosis 9,82 demek, aşırı büyük getirilerin normal dağılımın öngördüğünden çok daha sık yaşandığı demek. Bitcoin'de "beklenmedik büyük hareket" aslında pek de beklenmedik değil.



Gerçek getiri dağılımı (histogram), Normal ve Student-t dağılımlarıyla karşılaştırılıyor (solda) ve Normal Q-Q grafiği (sağda) — kuyruklarda teorik çizgiden sapma açıkça görülüyor.

Durağanlık testleri de (ADF: -133,20, p≈0; KPSS: 0,26, p=0,10) serinin durağan olduğunu, yani ortalama ve varyansının zaman içinde sistematik olarak kaymadığını doğruladı — bu, GARCH modellemesi için önkoşul niteliğinde bir bulgu.

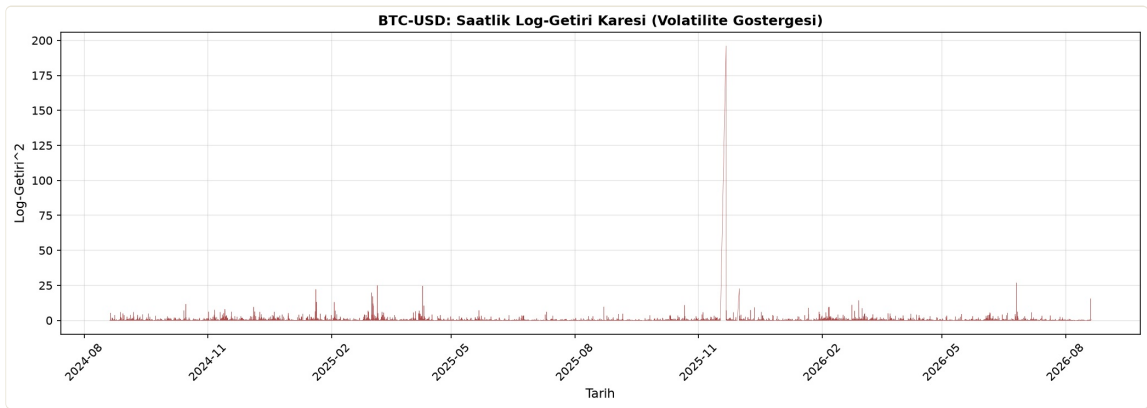


Peki getirilerin kendisi durağan ve "temiz" görünüyorsa, GARCH modellemesine neden ihtiyacımız var? Cevap, getirilerin **karelerinde** saklı.



7. GARCH'ın Kalbi: Volatilite Kümelenmesi

Getirilerin kendisi rastgele görünse de, getirilerin *büyüklüğü* (yönünden bağımsız olarak) rastgele değil. Haftalık ortalama mutlak getiriyi çizdiğimde şunu gördüm:



Bu görsel, GARCH ailesinin var olma nedenini özetliyor. Fikir aslında çok sade:

1. **Bugünün volatilitesi, dünkü volatiliteye bakar.** Sakin dönemler sakın kalma, fırtınalı dönemler fırtınalı kalma eğiliminde.

2. **Kötü haberler, iyi haberlerden daha çok sarsar.** Buna "kaldıraç etkisi" deniyor (birazdan detaylandıracağım).
3. **Uç değerler için kalın kuyruklu bir dağılım gerekir.** Normal dağılım yerine Student-t veya Skewed-t.

Bu üç fikri farklı derecelerde yakalayan dört model ailesini test ettim: **ARCH** (Engle, 1982), **GARCH** (Bollerslev, 1986), **EGARCH** (Nelson, 1991) ve **GJR-GARCH** (Glosten-Jagannathan-Runkle, 1993). Her birini üç farklı hata dağılımıyla (Normal, Student-t, Skewed-t) kurarak toplam 15 model elde ettim.



8. 15 Model, Tek Kazanan

Tüm modelleri aynı ortalama denklemleriyle (AR(2), Bölüm 9'da değineceğim) ve aynı veri setiyle kurduktan sonra, AIC (Akaike Bilgi Kriteri) değerlerine göre sıraladım:

Sıra	Model	AIC	BIC	Kalıcılık	Durum
1	EGARCH × Skewed-t	175.063,06	175.132,90	0,948	Durağan
2	EGARCH × t	175.067,80	175.129,88	0,948	Durağan
3	GJR × t	175.287,49	175.349,57	0,983	Durağan
4	GARCH × Skewed-t	175.288,83	175.350,91	1,000	Sınırdaki
5	Approx. IGARCH × t	175.293,30	175.347,62	1,000	Patlayıcı

EGARCH × Skewed-t, en yakın rakibi olan GJR modelinden yaklaşık 219 puan daha iyi bir AIC ile açık ara kazandı. Kalıcılık parametresi ($\beta=0,948$) 1'in altında, yani model **durağan** — patlayıcı olmayan, matematiksel olarak sağlıklı bir süreç. Ve evet, bu sonuç Katsiampa'nın (2017) 8 yıl önce Bitcoin için bulduğu sonuçla birebir örtüşüyor.

Normal dağılımlı tüm varyantlar (AIC $\approx 178.800+$) tabloya girmeye bile yaklaşamadı — bu da kalın kuyruklu dağılımların bu veri setinde ne kadar kritik olduğunu gösteriyor.

Bir de "Approx. IGARCH" adında üçüncü bir varyant test ettim: $\alpha + \beta$ toplamı 1'e çok yakın ($\approx 0,05$ tolerans) çıkan modelleri, entegre GARCH (IGARCH) süreçlerine yakın kabul ettim. Bu varyantların çoğu "patlayıcı" (kalıcılığı 1 veya üzerinde) çıktı — yani şokların etkisi zamanla sönümlenmiyor, kalıcı hale geliyor. Pratikte bu, uzun vadeli volatilité tahmini için güvenilir olmayan bir yapı anlamına geliyor ve bu yüzden nihai seçimde tercih edilmedi.



9. Ortalama Denklemi ve Model Denklemi

GARCH modelinin varyans kısmına geçmeden önce, getirinin ortalama (mean) kısmını da modellemek gerekiyor. Sabit ortalama, AR(1) ve AR(2) modellerini karşılaştırdım; AR(2) en düşük AIC ile öne çıktı ve tutarlılık için tüm 15 modelde sabit tutuldu.

Nihai modelin tam matematiksel formu (Nelson, 1991) şöyle:

$$\begin{aligned} \log(h_t) &= \omega + \alpha(|z_{t-1}| - E|z|) + \gamma \cdot z_{t-1} + \beta \cdot \log(h_{t-1}) \\ z_t &= c + \phi_1 \cdot z_{t-1} + \phi_2 \cdot z_{t-2} + e_t \quad [AR(2)] \\ e_t &\sim \text{Skewed-}t(\nu, \lambda) \end{aligned}$$

Tahmin edilen katsayılar:

- ω (sabit) = 0,063
- α (boyut etkisi) = 0,365
- **β (kalcılık) = 0,948**
- **γ (kaldıraç) = -0,027**
- ν (serbestlik derecesi) = 3,29
- λ (çarpıklık) = -0,026

Bu sayılar soyut görünebilir, ama her biri bir hikâye anlatıyor. β 'nin 1'e yakın ama altında olması, bir şokun etkisinin uzun süre (ama sonsuza kadar değil) sürdüğü anlamına geliyor. γ 'nın negatif ve anlamlı olması ise bir sonraki bölümün konusu.



10. Kötü Haberler Neden Daha Çok Sarsıyor?

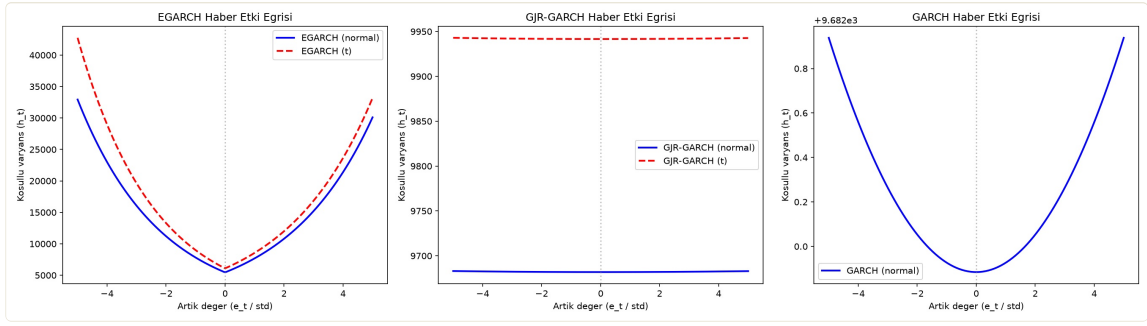
Finans literatüründe iyi bilinen bir olgu var: piyasada kötü bir haber geldiğinde (fiyat düşünce), volatilité aynı büyüklükteki iyi bir haberden (fiyat yükselince) daha fazla artıyor. Buna **kaldıraç etkisi** (leverage effect) deniyor.

Bunu iki farklı model üzerinden test ettim:

Model	Dağılım	Gamma (γ)	p-değeri	Durum
EGARCH	Normal	-0,0090	0,4642	Anlamsız
GJR	Normal	+0,0189	0,4114	Anlamsız
EGARCH	Student-t	-0,0253	0,0026	Anlamlı
GJR	Student-t	+0,0453	0,0074	Anlamlı

İki bulgu dikkat çekici. Birincisi: kaldıraç etkisi yalnızca kalın kuyruklu dağılımlarda (Student-t, Skewed-t) istatistiksel olarak anlamlı çıkıyor — Normal dağılımda görünmez bile. Bu, dağılım seçiminin sadece "istatistiksel bir detay" olmadığını, bulguların kendisini şekillendirdiğini gösteriyor.

İkincisi: EGARCH'ta γ negatif, GJR'de γ pozitif — işaretler ters ama ekonomik anlam aynı, çünkü iki model matematiksel olarak farklı kurulmuş.



EGARCH, GJR-GARCH ve GARCH modelleri için "haber etki eğrileri" (news impact curves). EGARCH ve GJR'nin asimetric (eğri merkeze göre kayık), GARCH'ın ise simetrik (parabol) tepki verdiğiine dikkat edin.

İlginç bir not: Bouri vd. (2017) Bitcoin'de kaldıraç etkisinin zayıf olduğunu bulmuştu; benim bulgularım bu noktada literatürden biraz ayrıştı — etki hem EGARCH hem GJR'de istatistiksel olarak anlamlı çıktı. Bilim böyle işliyor zaten: her çalışma bir öncekini tam olarak doğrulamak zorunda değil.



11. Krizi İkiye Bölmek: Alt Dönem Analizi

Kasım 2025 çöküşü o kadar belirgindi ki, veriyi bu tarihten ikiye bölüp ("Dönem 1: öncesi", "Dönem 2: sonrası") nihai modeli her iki dönemde ayrı ayrı tahmin etmek istedim. Amacım basitti: model parametreleri kriz öncesi ve sonrasında ne kadar farklılaşıyor?

Parametre	Dönem 1 (n=10.895)	Dönem 2 (n=6.409)	Fark
alpha (α)	0,3761	0,3428	-0,0333
beta (β)	0,9389	0,9614	+0,0226
gamma (γ)	-0,0349	-0,0128	+0,0221

Gamma her iki dönemde de negatif — yani kaldıraç etkisi krizden önce de sonra da varlığını sürdürüyor, bu da bulgunun ne kadar sağlam olduğunu gösteriyor. Ama etkinin büyüklüğü kriz öncesinde (-0,035) kriz sonrasına (-0,013) göre belirgin biçimde daha güçlü. Bunu şöyle yorumluyorum: kriz öncesi dönem, piyasanın henüz "sindirmediği" haberlere karşı daha sinirli tepki veriyordu; kriz sonrasında ise piyasa bir anlamda "yorgun" ve tepkiler nispeten yumuşamış.

Elbette bu bölünmenin kendisi bir veri artefaktına (yfinance kesintisi nedeniyle doğal bir kesim noktası oluşması) dayandığı için, bulguyu kesin bir nedensellik olarak değil, tanımlayıcı bir gözlem olarak sunuyorum.



12. Modeli Sorguya Çekmek: Tanısal Testler

Bir modeli "en iyi" ilan etmeden önce, artıklarında (residuals) hâlâ yakalanmamış bir örüntü kalıp kalmadığını kontrol etmek gerekiyor. Ljung-Box ve ARCH-LM testlerini uyguladığımda, teknik olarak tüm modeller reddedildi ($p \approx 0$).

İlk bakışta bu kötü bir haber gibi görünüyor. Ama R^2 değerlerine baktığımda tablo değişti: ARCH-LM testinin R^2 değeri sadece %1,97-2,49 civarında — yani model, varyansın **%97,5-98'ini** açıkladıktan sonra artıklarda yalnızca küçük bir kalıntı etki kalıyor. Brooks (2019)'un da belirttiği gibi, büyük örneklem hacimlerinde (benim durumumda $n=17.324$) istatistiksel testler aşırı hassas hale geliyor — **istatistiksel anlamlılık** ile **pratik önem** burada birbirinden ayrışıyor.

Bu benim için önemli bir ders oldu: bir p-değerine bakıp hemen "model kötü" demek yerine, etkinin büyüklüğüne (effect size) de bakmak gerekiyor.



13. Modeli Gerçek Dünyaya Çıkarmak: OOS, VaR ve Dürüst Bir Kapanış

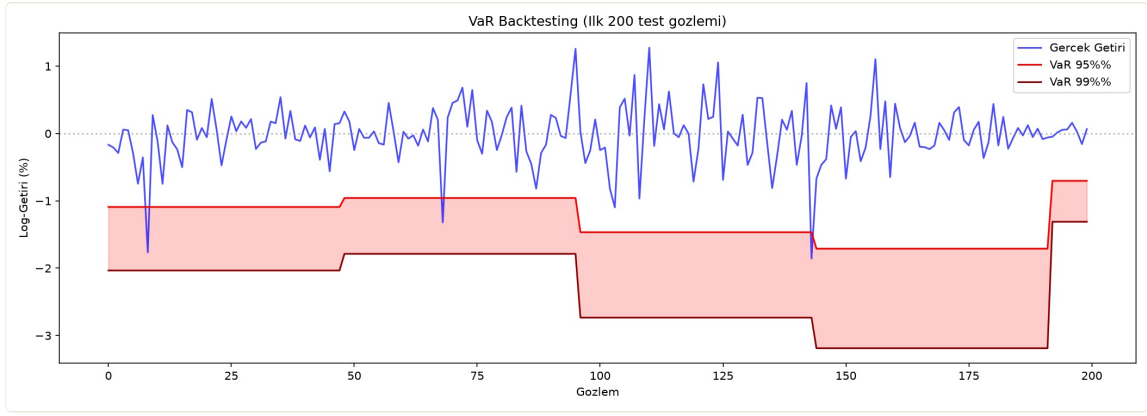
Son adım, modelin sadece eğitim verisinde değil, hiç görmediği veride de işe yarayıp yaramadığını test etmektir. Veriyi %80 eğitim / %20 test olarak ayırdım; 1.000 gözlemlik kayan pencereyle, her 24 saatte bir yeniden eğiterek örneklem dışı (out-of-sample) tahminler ürettim.

Model	RMSE	MAE	QLIKE
GARCH × Skewed-t	5,989	3,759	1,793
GJR × Student-t	5,929	3,746	1,796
GJR × Skewed-t	5,911	3,743	1,800
EGARCH × Skewed-t	5,880	3,751	1,843

Burada beklemediğim bir şey oldu: EGARCH × Skewed-t, örneklem içinde en iyisiyken ve örneklem dışında en iyi RMSE'ye sahipken, **QLIKE** ölçütünde son sıraya düştü. Bu, olası bir **aşırı uyum (overfitting)** sinyali. Ama Diebold-Mariano testi (Newey-West HAC düzeltmeli) bu farkın istatistiksel olarak anlamlı olmadığını gösterdi ($p=0,595$) — yani pratikte modeller arasında kesin bir üstünlük yok, ve bunu gizlemek yerine olduğu gibi raporladım.

Son olarak, modelin risk yönetimindeki kullanılabilirliğini Value at Risk (VaR) ile test ettim:

Seviye	İhlal	Gerçekleşen	Beklenen	Karar
%95 VaR	83/3.465	%2,40	%5,0	RED
%99 VaR	15/3.465	%0,43	%1,0	RED



VaR backtesting sonuçları (ilk 200 test gözlemi). Mavi çizgi gerçek getiriyi, kırmızı bantlar %95 ve %99 VaR sınırlarını gösteriyor. Gerçek getirinin bu sınırları aştığı nokta sayısı, beklenenden azdı.

Kupiec testi her iki seviyede de modeli reddetti — ama yön önemli: gerçekleşen ihlaller beklenenden **az**, yani model riski *olduğundan fazla* gösteriyor. Bu, tehlikeli olan senaryonun (riski olduğundan az göstermek) tam tersi; model aşırı temkinli. Kusursuz değil, ama güvenlik açısından "yanlış tarafta" bir hata.



Kapanış: Bir Modelden Fazlası

Bu projenin sonunda elimde EGARCH(1,1) × Skewed-t adında, AIC'si en düşük, kalıcılığı durağan, kaldıraç etkisini istatistiksel olarak anlamlı biçimde yakalayan bir model var. Ama benim için asıl değerli olan, bu modelin nerede iyi, nerede sınırlı olduğunu adım adım, dürüstçe ortaya koyabilmiş olmak: veri boşluklarının kriz anlarıyla örtüşmesi, olası aşırı uyum sinyali, VaR'ın aşırı temkinli kalibrasyonu — hiçbirini halının altına süpürmedim.

Gelecek için üç öneri bırakıyorum: VaR kalibrasyonunu iyileştirmek için Expected Shortfall (CVaR) denenebilir; kriz dönemlerinde veri sürekliliği için Binance/Coinbase gibi alternatif kaynaklar kullanılabilir; ve rejim değişimli GARCH (MS-GARCH) modelleri R ortamında test edilerek sağlamlık analizi genişletilebilir.

Bitcoin'in nabzını tam olarak yakalayabildim mi? Kısmen. Ama en azından artık o nabzın nerede düzenli, nerede öngörülemez attığını biraz daha iyi biliyorum.



Bu makale, "Bitcoin Saatlik Volatilite Tahmininde GARCH Ailesi Modelleri" başlıklı bitirme projesinin bulgularına dayanmaktadır. Proje kaynak kodları, veri seti ve tam rapor talep üzerine paylaşılabilir.

Anahtar kelimeler: GARCH, EGARCH, Bitcoin, volatilite, kripto para, zaman serisi, risk yönetimi, Value at Risk, finansal ekonometri