FUNDAÇÃO GETULIO VARGAS
ESCOLA DE ECONOMIA DE SÃO PAULO

JOSÉ PARREIRAS ANTUNES NETO

TESTING FOR EXTREME LONG-MEMORY COMMON FEATURES IN VOLATILITY PROCESSES

São Paulo

2020

JOSÉ PARREIRAS ANTUNES NETO

TESTING FOR EXTREME LONG-MEMORY COMMON FEATURES IN VOLATILITY PROCESSES

Dissertação apresentada à Escola de Economia de São Paulo como pré-requisito à obtenção de título de mestre em Economia de Empresas.

Orientador: Marcelo Fernandes.

.

São Paulo

2020

Antunes Neto, José Parreiras.

Testing for extreme long-memory common features in volatility processes / José Parreiras Antunes Neto. - 2020.

36 f.

Orientador: Marcelo Fernandes.

Dissertação (mestrado CMEE) – Fundação Getulio Vargas, Escola de Economia de São Paulo.

1. Volatilidade (Finanças). 2. Análise de séries temporais. 3. Autocorrelação (Estatística). 4. Agregação (Economia). I. Fernandes, Marcelo. II. Dissertação (mestrado CMEE) – Escola de Economia de São Paulo. III. Fundação Getulio Vargas. IV. Título.

CDU 519.246.8

Ficha Catalográfica elaborada por: Isabele Oliveira dos Santos Garcia CRB SP-010191/O
Biblioteca Karl A. Boedecker da Fundação Getulio Vargas - SP

JOSÉ PARREIRAS ANTUNES NETO

TESTING FOR EXTREME LONG-MEMORY COMMON FEATURES IN VOLATILITY PROCESSES

Dissertação apresentada à Escola de Economia de São Paulo como pré-requisito à obtenção de título de mestre em Economia de Empresas.

Data de aprovação: 30/04/2020

Banca examinadora:

Prof. Dr. Marcelo Fernandes
FGV-EESP (Orientador)

Prof. Dr. Gustavo Fruet Dias
University of East Anglia

Prof. Dr. Fernando Chague
FGV-EESP

Agradecimentos

Agradeço a minha família por toda a base e educação que me permitiram ao longo dos anos. A pais, avós, tios e primos pelo amor e carinho sempre.

Ao meu orientador, Marcelo Fernandes, pelo suporte durante o curso e por confiar em mim desde a primeira reunião que tivemos. Aos Professores Gustavo Fruet Dias e Fernando Chague, por terem aceitado participar da banca de dissertação e pelas contribuições feitas a este trabalho que foram de grande importância para o resultado final.

Agradeço aos Professores Sérgio Guerra e Hélio Berni, que sempre foram meus mentores e que foram fundamentais para me direcionar a fazer o meu mestrado.

A todas as amizades de Itabirito, principalmente aos *P.U.G.S.*: Eduardo Garajau, Gabriel Oliveira, Germano Crispim, Elmo Braga Jr., Marco Tulio Gurgel, Marco Wellington, Nícolas Simões, Pedro Ferreira e Victor Brito, que sempre estiveram comigo independente dos quilômetros de distância que nos separaram recentemente. Ainda agradeço aos amigos Gustavo Siqueira, Breno Fagundes e Diego Mulelos por toda a companhia de longa data.

A todos os amigos da *Salinha*: Alexandre Rabelo, Angelo Hermeto, Ariel Akerman, Caio Castro (aquele), Fernando Lopes, Gabriel Jardanovski, Gabriel Toledo, João Pugliese, Leonardo Makoto, Luis Gustavo Gomes, Nícolas Pimentel, Otávio Braga, Otávio Teixeira, Pedro Ogeda, Rafael Alves e Victor Orestes. Sem a companhia de todos eles, o mestrado seria uma tarefa impossível.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

Resumo

A agregação de processos de baixa dependência gera uma elevação na memória da variável agregada. Muitos modelos capturam esta memória longa e foram utilizados para modelar volatilidade. Destes, os que mais se destacam são baseados no modelo de integração fracionada. Entretando, dependendo do modod em que as séries são agregadas, os modelos de integração fracionada não conseguem aproximar da melhor forma a função de autocorrelação. Nós assumimos uma forma mais flexível para a função de autocorrelação de modo a acomodar a memória longa existente nos processos de volatilidade. Neste artigo, propomos um estimador para séries de memória longa e derivamos o seu comportamento assintótico. Baseado no nosso estimador, construímos um teste para a identificação de componentes comuns em várias medidas de volatilidade que apresentam alta persistência. Nós aplicamos a nossa metodologia a medidas observáveis de volatilidade de mercado (especificamente VIX e VXO), onde a memória longa é um problema frequente. Encontramos evidência a favor da existência de um componente comum de memória longa nestas séries que pode ser relacionado à volatilidade natural de mercado.

Palavras-chave: Memória Longa, Volatilidade, Componentes Comuns, Integração Fracionada, Autocorrelação, Estimador de Distância Mínima, Agregação.

Abstract

Cross-sectional aggregation of short-memory processes may generate long memory in the resulting time series. Many time-series models capture long-memory and have already been used to model volatility, of which the most popular relies on fractional integration. However, depending on how aggregation occurs, fractional integration does not provide a good approximation for the autocorrelation function. Accordingly, we assume a very flexible form for the autocorrelation function in order to accommodate long-memory in volatility processes. In this article, we propose an estimator for long-memory processes and derive its asymptotic behavior. We construct a test for the identification of a common feature among multiple volatility measures that exhibit high persistence. We apply our methodology to observed market volatility (namely VIX and VXO), where long-memory is a frequent concern, and find evidence suggesting that the natural market volatility appears as a long-memory feature in our data.

Keywords: Long memory, Volatility, Common Features, Fractional Integration, Autocorrelation, Minimum Distance Estimator, Aggregation.

List of Figures

Figure 1 – Autocorrelation Function of ACT 15

Figure 2 – Histograms of Monte Carlo Study 21

Figure 3 – Evolution of VIX and VXO over time 23

Figure 4 – Autocorrelation Functions for VIX, VXO and VIX-VXO 24

List of Tables

Table 1 – Data description 23
Table 2 – AR(1) Estimation 24
Table 3 – ACT Estimation 25
Table 4 – ACT Model 26

Contents

1 Introduction **10**

2 Aggregation and Persistence **13**

3 Common Features **16**

4 Inference **18**

4.1 Estimation 18

4.2 Testing 19

4.3 Power against fractional integration 20

5 Common Features in the CBOE Volatility Indexes **22**

6 Conclusion **27**

Bibliography **28**

1 Introduction

Volatility series commonly exhibit a high degree of correlation with very distant observations. For this reason, various models have been derived for long memory such as Fractionally Integrated and Heterogeneous Autoregressive models (Granger (1980) and Corsi (2009), respectively). However, when dealing with market volatility, the classical models may be insufficient. Market volatility aggregates individual stocks volatility, which are usually modeled with exponential processes. This aggregation combines the cross-correlations of many stocks and generates a high level of persistence in the aggregated measure. These nonlinearities generate a more complex structure in the autocorrelations, which classical models fail to approximate. For this reason, in order to incorporate this long memory and study the common features in volatility, we use a more flexible model. Our model nests classical shorter-memory models and allows for the more persistent pattern seen in the data. Neglecting to consider the characteristics of the long memory can lead to a large loss of information and make estimation and forecasting less efficient.

Aggregation of multiple time series is known to enhance persistence in stochastic processes. This elevated persistence has induced econometricians to construct advanced models capable of handling a more complex set of variables. The pioneering work of Granger (1980) and Granger and Joyeux (1980) motivates and presents the commonly used fractionally integrated (FI) model - a specific autoregressive (AR) model that seems to work well for most long-memory series. Another kind of long-memory model, introduced by Corsi (2009), constrains the structure of the AR model for specific lags. It seems to be well-suited in some cases. Fernandes, Medeiros, and Scharth (2014) give applied examples for this model.

Changing the aggregation type, however, causes the classical models to fail to fully capture the persistence of the resulting time series. The result of this nonlinear type of aggregation is studied by Abadir and Talmain (2002, 2012) and Abadir, Caggiano, and Talmain (2013) (ACT). They demonstrate that the autocorrelation function (ACF) implied by Granger's models does not capture the persistence well in such cases. They show that when aggregating variables that follow geometric autoreggressive moving average ($ARMA$) processes, the behavior of the ACF does not match the predicted ACFs of fractional integrated model. They derive an ACF approximation that seems better-suited

to these cases than the current econometric models. We build on their work to construct a novel approach to model aggregated series and construct a new model to deal with the long-memory common feature.

We propose a method to identify the presence of long-memory common features following ACT model. We explain how to estimate the parameters of ACT and derive the asymptotic behavior of our estimators. We estimate our parameters using a minimum distance estimator (MDE) approach which approximates the empirical autocorrelations to the ones implied by the ACT model. Under simple assumptions, we show that our estimator is consistent and asymptotically normally distributed. This asymptotic normality allows us to construct a test for common features, a single component that seems to drive the dynamics of multiple series (Vahid and Engle (1993)).

We apply our model in aggregated volatility data using the Chicago Board Options Exchange volatility indexes: VIX and VXO. These are two of the most-used volatility measures which aim to proxy for unobserved market volatility. Both indexes analyze a wide range of stocks in order to approximate the market and therefore possess a high level of persistence (the stocks constituting the S&P 100 and the S&P 500 indexes, respectively). As measures of market volatility may rely on an aggregation function similar to these studied by Abadir and Talmain (2002), we can detect the presence of this kind of long-memory process in financial data. Also, since both variables aim to capture the same market, they may present a common long-memory feature of this kind (the natural volatility process). We explore this topic in Section 5. We collect evidence suggesting the existence of an ACT common feature in volatility, possibly related to the natural market volatility process that observed measures tend to proxy. This is suggested by the fact that, when looking at individual variables, we can detect the presence of ACT process. However, when taking linear combinations of our variables, this long-memory pattern seems to disappear, indicating that this long memory may be generated by a common component.

This work is organized as follows. In Section 2 we present a new type of aggregation based on Abadir and Talmain (2002). We show that, when considering a simple model for financial data, market volatility follows the ACT model. In Section 3 we analyze the long-memory common feature and how it can be identified given the correct test statistic. In Section 4 we explain the estimation of this long-memory process and the implied hypothesis testing to detect the common feature in the data. In Section 5 we apply our findings to market volatility data. We show that when looking at aggregated volatility in

different markets, a common efficient process can be detected to drive the dynamics of the random variables. This process relates to those described in the early sections and can be used to exemplify our method.

2 Aggregation and Persistence

Aggregation of multiple time series causes an increase in the persistence of the final variable — as we combine the cross-correlation of many variables into the aggregated measures, we generate the long-memory dynamics usually seen in the data. However, Abadir and Talmain (2002) argue that, depending on the kind of aggregation underlying the objective variable, usual econometric models (such as the fractionally integrated $FI(d)$ models) are unable to model the dynamics of persistent variables.

The aggregation of multiple $ARMA$ processes generates a hyperbolic decay on the autocorrelation function of the aggregated variable as shown by Granger (1980). However, in economic formulations we usually assume exponential processes. These nonlinearities imply that aggregated variables cannot be modeled by the classical methods, because the correlations between current and past values do not exhibit a hyperbolic decay. This is shown on the works of Abadir and Talmain (2002, 2012) and Abadir, Caggiano, and Talmain (2013). As an example, take a simple model used in finance. Assume a set $\mathcal{S}$ of stocks following independent stochastic volatility models:

$$r_{i,t} = \exp\left(\frac{h_{i,t}}{2}\right)\xi_{i,t} \qquad i \in \mathcal{S}, t \in \mathbb{Z} \tag{2.1}$$

$$h_{i,t} = \Psi_q(B)\eta_{i,t} \tag{2.2}$$

with $\xi_{i,t} \sim WN(0, \sigma^2_{\xi,i})$ and $\eta_{i,t} \sim WN(0, \sigma^2_{\eta,i})$, where $\Psi_q(B) = \sum_{j=1}^{q} \psi_j B^j$ is a small-order polynomial. It then implies that the logarithm of the returns can also be modeled by finite-order moving average:

$$\log r^2_{i,t} = \Psi_q(B)\eta_{i,t} + \log \xi^2_{i,t} \tag{2.3}$$

which involves a moving average process of order q. Market volatility aggregates each individual process in 2.3:

$$V_t \coloneqq \sum_{i \in \mathcal{S}} w_i \mathbb{E}_t \left[r^2_{i,t+1}\right] \tag{2.4}$$

so that it is the sum of "geometric-AR" processes. As shown by Abadir and Talmain (2002), this kind of log aggregation, although shown to be stationary and persistent, cannot be well-modeled by Granger's fractionally-integrated models. The decay of its autocorrelation function does not present the exponential pattern implied by the $FI(d)$ process. Abadir,

Caggiano, and Talmain (2013) show that the autocorrelation function of this kind of "log-aggregated" variables is well-approximated by[1]:

$$\rho_\theta(\tau) \coloneqq \frac{1 - a(1 - \cos(\omega\tau))}{1 + b\tau^c} \qquad \theta \coloneqq (a, b, c, \omega). \tag{2.5}$$

We can interpret the above function as follows. The sinusoidal behavior of the numerator controls for the shape of this new type of persistent process. The parameter a controls for the amplitude of oscillations. The denominator controls for the decay of memory. Taking $a = 0$, we are back to the classical ACF of AR and $ARFI$ processes. The parameters b and c are given to be strictly positive. We base our model on the work of Abadir, Caggiano, and Talmain (2013). For a time series Y_t with autocorrelation function as in (2.5) and "true" parameter vector θ_0, we write $Y_t \sim ACT(\theta_0)$.

We examine the ACT model and consider a set of different values of θ to show the many forms that $\rho_\theta(\tau)$ can assume. The results are shown in Figure 1. We start with a vector of parameters $(1, 0.05, 0.5, 0.002)$. As indicated by Abadir, Caggiano, and Talmain (2013), for persistent variables, we might have $a \approx 1$, which results in the pattern described in Panel (a). The values of ω also tend to be quite low, to promote the long cycles in the ACF. The values of b and c control for the decay of the ACF. In Panel (b) we see what happens when $a = 0$. As indicated earlier, when that is the case, we get a function that resembles those generated by fractional integration. When taking either b or c to 0, Panel (c) shows that we do not have an ACF that decays (in absolute value) as the lag grows. Such a case cannot be possible for stationary time series, so we consider only strictly positive values of these parameters. Also, when ω is too high (here equal to 0.01), we have the pattern in Panel (d), which restricts the cycles of the ACF to really short lags. Therefore we see that the parameter ω controls for the periodicity of the ACF.

[1] In accordance with the authors, we assume identification of each parameter in (2.5).

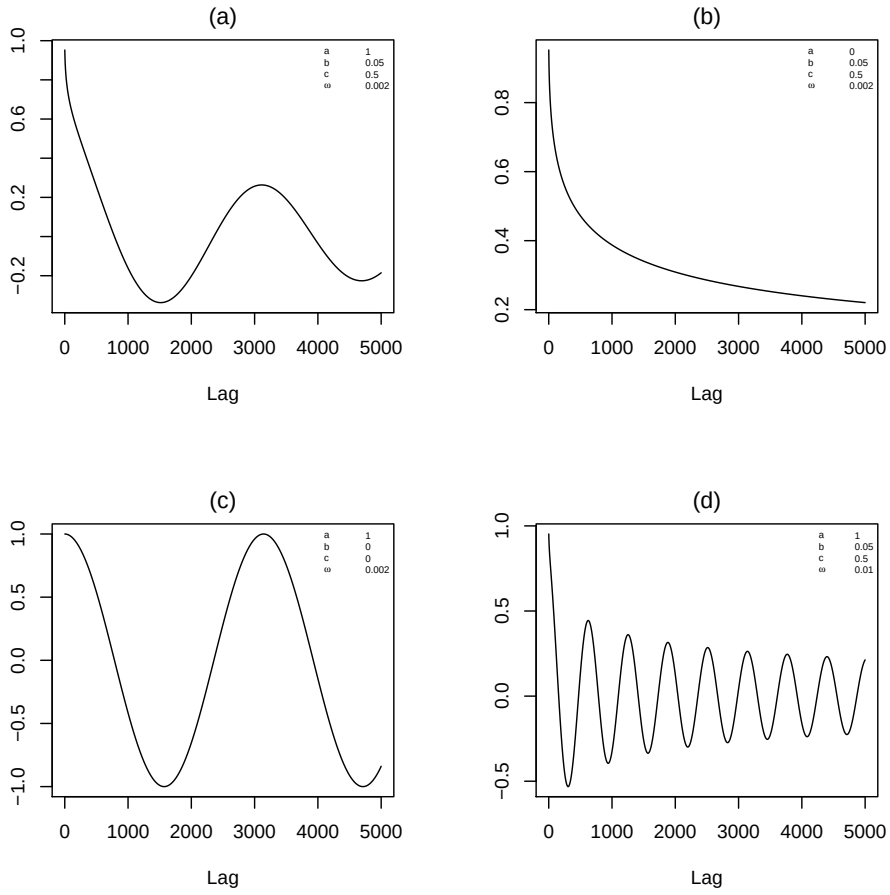

Figure 1: Autocorrelation Function of ACT

3 Common Features

We follow the work of Vahid and Engle (1997) and Engle and Kozicki (1993) to specify a common feature. For a K-dimensional vector stochastic process $Y_t \coloneqq (Y_{1,t}, \ldots, Y_{K,t})$, the feature $f_t \in \mathbb{R}^r$, (with $k > r$) is a common feature:

1. If $Y_{i,t}$ has (does not has) the feature, then $cY_{i,t}$ also has (does not has) the feature for any $c \neq 0$.
2. If $Y_{i,t}$ and $Y_{j,t}$ both do not have the feature, then $X_t \coloneqq Y_{i,t} + \lambda Y_{j,t}$ does not has the feature for any $\lambda \in \mathbb{R}$.
3. If $Y_{i,t}$ has the feature, but $Y_{j,t}$ does not, then $X_t \coloneqq Y_{i,t} + \lambda Y_{j,t}$ has the feature for any $\lambda \in \mathbb{R}$.

As Engle and Kozicki (1993) write, " a feature will be said to be common if a linear combination of the series fails to have the feature even though each of the series individually has the feature".

Under the hypothesis that our common feature $f_t \sim ACT(\theta_0)$, we can write

$$Y_t = \mu + Af_t + \varepsilon_t \qquad \varepsilon_t \sim VMA(q) \tag{3.1}$$

with q possibly infinite. As $f_t \sim ACT(\theta_0)$, the increased degree of persistence of f_t contaminates the dynamics of Y_t. The elements in ε_t are considered of shorter memory. This means that, by eliminating f_t, we take the high persistence out of our variables and are left with usual processes. As the elements of Y_t cannot drift apart in the long run, accounting for the common feature can improve estimation and predictions. We explain how can we test for this long-memory common feature.

From the definition of (2.5), taking $a = 0$ we have that

$$\rho_\theta(\tau) = \frac{1}{1 + b\tau^c}.$$

According to Hosking (1981) the autocorrelation function for a fractionally-integrated process of order d can be approximated by $K\tau^{2d-1}$ for some positive constant K as $\tau \to \infty$. When $a = 0$ it is easily shown that, as $\tau \to \infty$,

$$\rho_\theta(\tau) \approx \frac{1}{b}\tau^{-c}$$

so the ACT and the FI models do look close enough. Testing for this kind of long-memory feature in Y_t is equivalent to testing for the significance of a. If $a = 0$, then the usual $FI(d)$ do are capable of modeling these variables, because the implied ACF approximates the ACF of the ACT model. However, if $a \neq 0$, information is lost if the consequences of the ACT model are not considered as the FI and ACT models diverge.

We then test for the hypothesis that this long memory is common to every variable in Y_t by searching for a vector $\alpha \in \mathbb{R}^{k\times 1}$ such that the parameter $a = 0$ in the ACF of $\alpha' Y_t$. We next show how to estimate the parameters of the ACT model and explain how we can test for a long-memory common feature on the process of Y_t.

4 Inference

We now show how to estimate the parameters of the ACT autocorrelation. We derive the asymptotic distribution of these estimators and then construct the procedure to test for the long-memory common feature on the vector random variable Y_t.

4.1 Estimation

Take Y_t following the ACT model defined as in (2.5) with "true" parameter vector θ_0. For a given sample $(Y_t)_{t=1}^T$ and an arbitrary set $\mathcal{H} \subset \mathbb{N}$, we take $\widehat{\rho}(h)$ defined as the sample autocorrelation of order h, i.e.:

$$\widehat{\rho}_T(h) \coloneqq \frac{\sum_{t=h}^T (Y_t - \bar{Y}_T)(Y_{t-h} - \bar{Y}_T)}{\sum_{t=1}^T (Y_t - \bar{Y}_T)^2} \qquad \bar{Y}_T \coloneqq \frac{1}{T}\sum_{t=1}^T Y_t. \tag{4.1}$$

We have that

$$\widehat{\rho}_T(h) = \rho_\theta(h) + \xi_T^h \qquad \underset{T\to\infty}{pLim}\ \xi_T^h = 0,\ \forall h \in \mathcal{H}. \tag{4.2}$$

From this framework, we can estimate θ_0 by approximating $\widehat{\rho}(h)$ by the deterministic function $\rho_\theta(h)$ using a minimum distance estimator (MDE). We now establish our main theorem.

Theorem 1. *Let $Y_t \sim ACT(\theta_0)$ and fix an arbitrary set $\mathcal{H} \subset \mathbb{N}$. Assume that Y_t has finite fourth moment and its autocorrelations are absolutely summable. Assume that θ_0 is identified and Θ is a compact and convex subset of $\mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}$. Define the MDE estimator $\widehat{\theta}_T$ as the minimizer of*

$$\widehat{Q}_T(\theta) \coloneqq \widehat{g}_T(\theta)' W \widehat{g}_T(\theta) \tag{4.3}$$

for $g_T(\theta) \coloneqq (\widehat{\rho}_T(h) - \rho_\theta(h))_{h\in\mathcal{H}}$ and W an arbitrary weighting matrix. Then as $T \to \infty$, $\widehat{\theta}_T \xrightarrow{p} \theta_0$.

Also, define $G(\theta) \coloneqq \nabla g_T(\theta)$ and assume that $G(\theta_0)WG(\theta_0)'$ is invertible. Then

$$\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) \xrightarrow{d} N(0, \Omega) \tag{4.4}$$

for $\Omega \coloneqq [G(\theta_0)WG(\theta_0)']^{-1} G(\theta_0)WVW'G(\theta_0)' [G(\theta_0)WG(\theta_0)']^{-1}$ *where* $V \coloneqq \mathbb{E}[\widehat{g}_T(\theta_0)\widehat{g}_T(\theta_0)']$ *with*

$$v_{i,j} \coloneqq \sum_{k\in\mathbb{N}} \{\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)\} \times \{\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)\} \quad (4.5)$$

The proofs are exhibited in the Appendix A.

Given the asymptotic normality of the estimator $\widehat{\theta}_T$, we can proceed to test the presence of the long-memory common feature in the vector Y_t by testing for the significance of the estimated $\widehat{a}$.

4.2 Testing

Consider the representation of variables in terms of the common feature (f_t) and their idiosyncratic components (ε_t) as in (4.6)

$$Y_t = \mu + Af_t + \varepsilon_t \qquad \varepsilon_t \sim VMA(q) \quad (4.6)$$

where f_t is the common feature defined as in Section 3. Under the null hypothesis, there exists a vector $\alpha_0 \in \mathbb{R}^p$ such that $\alpha_0'A = 0$. Engle and Kozicki (1993) show that testing for the presence of this common feature is equivalent to testing for a Scalar Component Model (SCM) as in Tiao and Tsay (1989). The SCM implies the existence of $\alpha_0 \in \mathbb{R}^{N\times 1}$ such that $\alpha_0'Y_t$ does not have the feature. In our case, it is equivalent to testing for the existence of α_0 such that the parameter θ for the ACT on the linear combination $\alpha_0'Y_t$ has $a = 0$. From (4.4), we can test the significance of a using the standard *Wald test*.

First, we need to identify the feature in each variable $Y_{i,t}$ in Y_t. For each i we estimate $\widehat{\alpha}_i$ and test whether this value is statistically different from 0. According to (4.4), we use the test statistic $\frac{\widehat{a}_i^2}{\frac{1}{T}\widehat{\mathbb{V}(a_i)}}$ which is asymptotically distributed as $\chi^2(1)$. If each one of our variables presents estimations of a statistically different from zero, we have identified that each has the feature. Then we are able to test whether this feature is indeed common.

Denote $X_t(\alpha) \coloneqq \alpha' Y_t$ a linear combination of the variables in Y_t and the test statistic $s(\alpha)$:

$$s(\alpha) \coloneqq \frac{\widehat{a}_\alpha^2}{\frac{1}{T}\widehat{\mathbb{V}(a_\alpha)}}$$

where $\widehat{a}_\alpha$ is the respective estimation of the ACT parameter a for the series $X_t(\alpha)$. We consider the null hypothesis $\mathcal{H}_0 : \inf_\alpha |a_\alpha| = 0$ of the existence of common feature and the

alternative $\mathcal{H}_1 : \inf_\alpha |a_\alpha| \neq 0$ of the nonexistence of a common feature. We take a set of testable vectors $\mathcal{A}$ and evaluate $\{s(\alpha) : \alpha \in \mathcal{A}\}$. Under the hypothesis of the existence of the long-memory common feature f_t, the vector α_0 is such that $s(\alpha_0) \leq \min_{\alpha\in\mathcal{A}} s(\alpha)$ and from (4.4), $s(\alpha_0) \xrightarrow{d} \chi^2(1)$. The procedure then is to compare the set of test statistics to some critical value c of the $\chi^2(1)$ table in order to infer the "true" α_0. If $\min_{\alpha\in\mathcal{A}} s(\alpha) \leq c$, then we do have sufficient evidence that the common feature exists. According to Engle and Kozicki (1993), the test statistics using the minimum have size less than or equal to the nominal size of the test over $s(\alpha_0)$.

4.3 Power against fractional integration

We have shown that the ACT is flexible enough to assume many forms. One of them is the $FI(d)$ pattern, which is represented by a slow decay of the ACF on the order of τ^{-d}. We now show that, assuming that our series follows an FI process, the estimated parameters of the ACT will likely be able to identify it. We consider a Monte Carlo study simulating $2,000$ repetitions of a fractional integrated process with parameter $d = 0.49$. We set the number of observations at $10,000$. For each simulation, we estimate our model and store the values of the parameters. Figure 1 shows the distribution of the estimated values of θ_0. We present only values in a 3-standard-deviation range.

We identify if the ACT model is indeed capable of identifying no sinusoidal pattern in the ACF and indicate that the series follow the correct FI process. According to our specification, when considering this case, the estimations of a must converge to 0 as the sinusoidal pattern in (2.5) does not appears in a fractional integrated model. When $a = 0$, as shown in Figure 1, the sinusoidal behavior of the ACF is extinguished and the decay pattern returns those described by FI.

Figure 2 shows that the estimated a are very close to zero, with mean of -0.06, median of 0.08 and standard deviation 1.06. This suggests that ACT model does have enough power to distinguish between a "true" ACT process (one with $a \neq 0$) and a fractional integrated process.

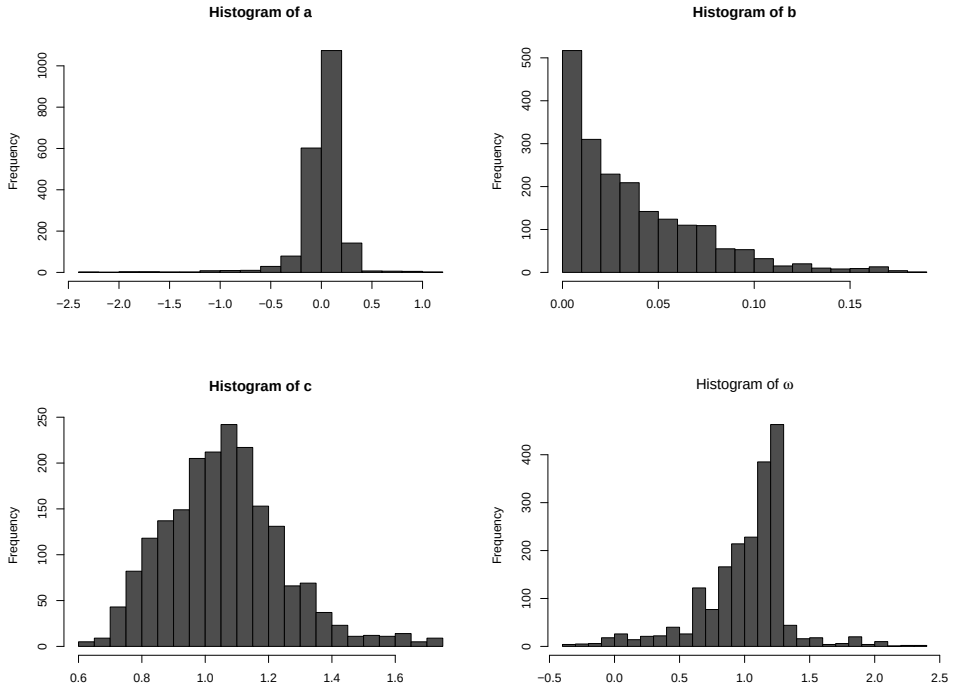

Figure 2: Histograms of Monte Carlo Study

5 Common Features in the CBOE Volatility Indexes

When we consider market volatility in financial models, we take the aggregated measure of many stocks. As shown in Section 2, this aggregation may not be well modeled by the usual $ARFIMA$ models. As financial models often rely on the assumption of geometric-AR processes (as in the stochastic volatility model presented in Section 2), the implied market volatility may be stated as following the $ACT(\theta_0)$ process.

The presence of common features in finance motivates the study of topics such as price and volatility discovery, in which one seeks to analyze the information arrival on different markets. When looking directly at prices, the integrated property allows one to represent variables (Gonzalo and Granger (1995)) as a sum of a $I(1)$ component (usually seen as the efficient price - the common feature) and a stationary noise variable. Hasbrouck (1995) and Baillie, Booth, Tse, and Zabotina (2002) elaborate on that topic. For individual volatility, where the fractionally-integrated models seem to be suitable, an analog decomposition can be stated for this case, in which the vector volatility is a sum of a $FI(d)$ natural volatility and an $I(0)$ noise Dias, Scherrer, Papailias, et al. (2016). However, as shown in Section 2, for aggregated measures of volatility, the assumption of a $FI(d)$ process may not be adequate. This new common feature on volatility data must be identified before investigating market volatility.

Volatility is traded for stock markets and many observed measures can be obtained. All these measures have the same goal to approximate the natural market volatility process. Then we can model them as driven by a latent common feature $V_t \sim ACT(\theta_0)$. For a vector of volatility measures $Y_t = (Y_{1,t}, Y_{2,t}, \ldots, Y_{k,t})$, we then write

$$Y_t = AV_t + \varepsilon_t \tag{5.1}$$

with $\varepsilon_t \coloneqq (\varepsilon_{1,t}, \varepsilon_{2,t}, \ldots, \varepsilon_{k,t})$ representing the vector of idiosyncratic components.

For our example, we consider the data on two volatility indexes. Both are from the Chicago Board Options Exchange (CBOE): the VIX and the VXO. These measure the volatility of the stocks constituted in the S&P 500 and in the S&P 100 indexes, respectively. Both indexes are calculated as functions of option strike prices and are proxies for the

natural market volatility. The data consists of daily observations from January 2nd, 1990 to December 28th, 2018 (7306 observations). Results are shown for the logarithm of each variable.

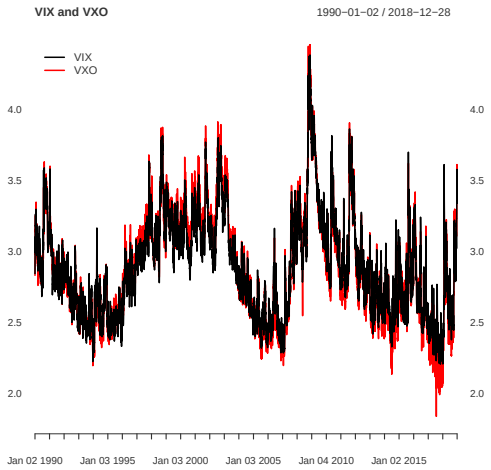

Figure 3: Evolution of VIX and VXO over time

	VIX	VXO
Min	2.213	1.844
1st. Qt	2.609	2.590
Median	2.856	2.854
Mean	2.893	2.895
3rd Qt.	3.121	3.160
Max	4.393	4.469
Std. Dev.	0.349	0.386
Kurtosis	0.642	0.471
Skewness	3.332	3.030

Table 1: Data description

Figure 3 shows that these series tend to follow a very similar path, indicating that the common factor might be present in this situation. We then analyze the behavior of their autocorrelations.

The autocorrelation function of both variables (Figure 4), show high level of persistence, as expected. Autocorrelations are statistically different from zero up to lag 2000. Moreover, the decay pattern does not seems to fit those predicted by AR and FI models. Indeed, when estimating an $AR(1)$ and a $FI(d)$ for both series, the coefficients appear to be on the boundary of the parameter space as shown in Table 2. The estimation of $FI(d)$ report estimators $\widehat{d}$ equal to 0.499 for both variables again near the boundary of the parameter space of $(-0.5, 0.5)$[1].

As these variables are thought to follow the same natural volatility, we examine the dynamics using an arbitrary vector $\alpha = (1, -1)$, i.e., we study the dynamics of the difference between VIX and VXO. The autocorrelation function is shown in Figure 4.

The pattern of the above ACF do look closer to the predicted dynamics of the usual AR models. Also, the estimations of an $AR(1)$ leave the boundary of the parameter space as shown in Table 2.

[1] We only consider values for d which are relative to stationary variables.

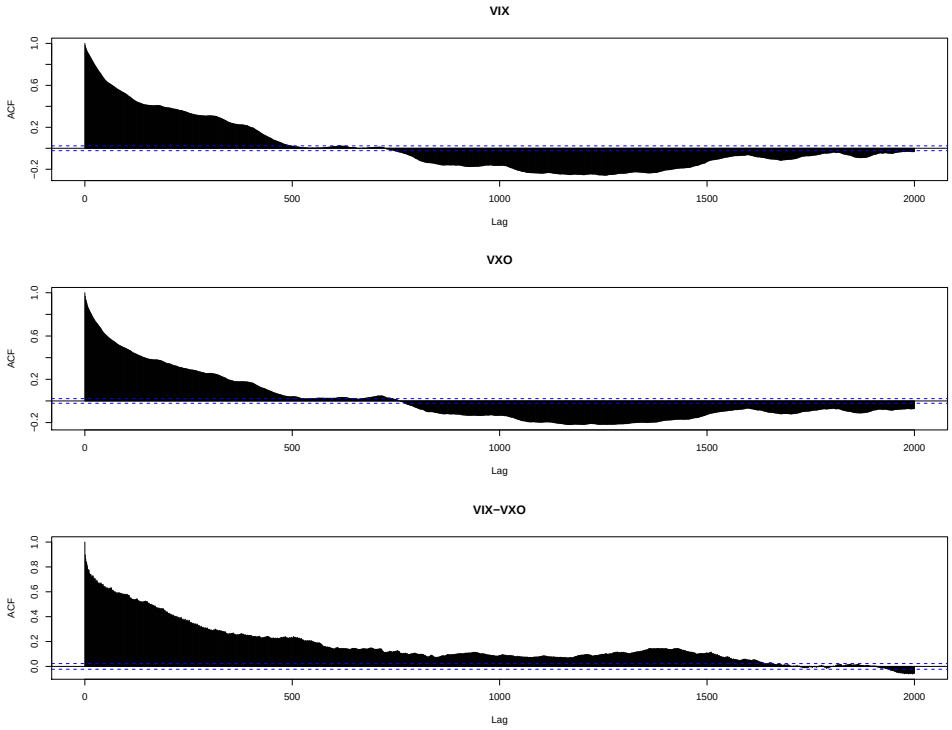

Figure 4: Autocorrelation Functions for VIX, VXO and VIX-VXO

Table 2: AR(1) Estimation

	Dependent variable		
	VIX	VXO	VIX-VXO
	(1)	(2)	(3)
ρ	0.983^{***}	0.983^{***}	0.893^{***}
	(0.002)	(0.002)	(0.005)
Intercept	2.893^{***}	2.895^{***}	-0.002
	(0.043)	(0.049)	(0.004)
Observations	7,306	7,306	7,306
σ^2	0.004	0.005	0.001
Note:	*: p< 0.1; **: p< 0.05; ***: p< 0.01		

We then proceed to estimate the ACT model for both series.[2] As seen in Table 3,

[2] Every estimation in this section considers an efficient estimator of θ_0 where the loading matrix of the MDE is constructed using a first-step estimation of ACT. The set of autocorrelations used in this exercise is $\mathcal{H} := \{1, 5, 10, \ldots, 5000\}$ and the parameter $M = 100$.

both series present estimations of a that are statistically different from 0, as the Wald test indicates.

Table 3: ACT Estimation

	Dependent variable	
	VIX	VXO
	(1)	(2)
a	1.058^{***}	1.054^{***}
	(0.001)	(0.001)
b	0.053^{***}	0.04^{***}
	(0.014)	(0.011)
c	0.537^{***}	0.59^{***}
	(0.001)	(0.001)
ω	0.002^{**}	0.002^{**}
	(0.001)	(0.001)
Note:	*: p<0.1; **: p<0.05; ***: p<0.01	

We proceed to test for a long-memory common feature in these variables. As both intend to capture market volatility, we can imagine that the pattern shown to this point is a characteristic of this unobservable process. If this is true, then we are able to eliminate this characteristic by taking some linear combination. We take a set of vectors

$$\mathcal{A} \coloneqq \{(1,-1),(1,-0.9),\ldots,(1,-.1)\} \bigcup \{(-1,1),(-0.9,1),\ldots,(-.1,1)\}$$

and test the significance of a for each linear combination $\alpha' Y_t$ for $\alpha \in \mathcal{A}$.

The results are shown in Table 4. As we can see, taking the vector $\alpha = (1,-0.8)$ generates the smaller test statistic, which is not statistically different from zero. This suggests that there exists some linear combination that eliminates the long-memory common feature, as the estimated parameter a does not seem to be significantly different from zero at a 5% level. As we can eliminate this kind of feature in the data, we can infer that it is a common feature. As seen in Section 2, we can relate this common feature to the natural market volatility measure, which aggregates individual stock volatility in a non-linear way and generates the ACT model. Since both measures analyzed tend to proxy for this variable, the pattern shown by their ACF does seem to be generated by the “natural market volatility" common feature. Moreover, when eliminating this common

feature, we are left with the idiosyncratic components of each variable. These components can be related to market microstructure noise (MMN) and have shown to be of shorter memory than the common feature.

This identification of an *ACT* common feature has implications to future works. According to our results, one cannot ignore the presence of this long-memory feature in volatility data. Disconsidering this pattern can lead to inferior estimations and forecasts due to the loss of information.

Table 4: ACT Model

ω	$s(\omega)$	ω	$s(\omega)$
$(1, -1.0)$	292.690^{***}	$(-1.0, 1)$	292.690^{***}
$(1, -0.9)$	51.466^{***}	$(-0.9, 1)$	5096.532^{***}
$(1, -0.8)$	2.745^{*}	$(-0.8, 1)$	72.758^{***}
$(1, -0.7)$	383.695^{***}	$(-0.7, 1)$	597.007^{***}
$(1, -0.6)$	12.069^{***}	$(-0.6, 1)$	13.638^{***}
$(1, -0.5)$	38.624^{***}	$(-0.5, 1)$	34.313^{***}
$(1, -0.4)$	237.964^{***}	$(-0.4, 1)$	3.604^{*}
$(1, -0.3)$	4823.614^{***}	$(-0.3, 1)$	53.149^{***}
$(1, -0.2)$	1457.984^{***}	$(-0.2, 1)$	$4.54e{+}21^{***}$
$(1, -0.1)$	350.579^{***}	$(-0.1, 1)$	61.723^{***}

Note: *: p<0.1; **: p<0.05; ***: p<0.01

6 Conclusion

We have presented a method to estimate the model of Abadir, Caggiano, and Talmain (2013) and test for a long-memory common feature following their model. The ACT model represents an alternative to long-memory modeling when the usual fractionally-integrated models are unable to capture the dynamics of the proposed data. It is shown that this long memory emerges when one changes the way variables are aggregated. This new long-memory process presented in the literature has shown to improve the fit, especially when analyzing aggregated variables. Our method is based on a minimum distance estimator and consistency and asymptotic normality are provided with minor assumptions.

Given the asymptotic behavior of our results, we show how to test for the common features according to the methodology of Vahid and Engle (1997) and Engle and Kozicki (1993) using a simple Wald test. We test for the presence of the long-memory common features in volatility data. Our exercise shows that, when modeling aggregated volatility measures (such as the VIX and VXO), one must account for the ACT dynamics. We also show that the source of this ACT pattern- can be traced to the natural market volatility which VIX and other measures aim to proxy. Moreover, market microstructure noise seems to be of shorter memory than the common feature. This identification of an ACT common feature has implications for future works. According to our results, ignoring the presence of this long-memory feature in volatility data leads to considerably inferior estimations and forecasts due to the loss of information capture in the past observations.

Bibliography

Abadir, K. M., G. Caggiano, and G. Talmain (2013). Nelson–plosser revisited: the acf approach. *Journal of Econometrics 175*(1), 22–34.

Abadir, K. M. and G. Talmain (2002). Aggregation, persistence and volatility in a macro model. *The Review of Economic Studies 69*(4), 749–779.

Abadir, K. M. and G. Talmain (2012). Beyond co-integration: Modelling co-movements in macro finance. *Available at SSRN 1989417*.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica 71*(2), 579–625.

Baillie, R. T., G. G. Booth, Y. Tse, and T. Zabotina (2002). Price discovery and common factor models. *Journal of Financial Markets 5*(3), 309–321.

Brockwell, P. J., R. A. Davis, and S. E. Fienberg (1991). *Time series: theory and methods*. Springer Science & Business Media.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics 7*(2), 174–196.

Dias, G. F., C. M. Scherrer, F. Papailias, et al. (2016). Volatility discovery. *CREATES Research Papers 113*, 114.

Engle, R. F. and S. Kozicki (1993). Testing for common features. *Journal of Business & Economic Statistics 11*(4), 369–380.

Fernandes, M., M. C. Medeiros, and M. Scharth (2014). Modeling and predicting the cboe market volatility index. *Journal of Banking & Finance 40*, 1–10.

Gonzalo, J. and C. Granger (1995). Estimation of common long-memory components in cointegrated systems. *Journal of Business & Economic Statistics 13*(1), 27–35.

Granger, C. W. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics 14*(2), 227–238.

Granger, C. W. and R. Joyeux (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis 1*(1), 15–29.

Hannan, E. J. (1974). The uniform convergence of autocovariances. *The Annals of Statistics*, 803–806.

Hasbrouck, J. (1995). One security, many markets: Determining the contributions to price discovery. *The Journal of Finance 50*(4), 1175–1199.

Hosking, J. (1981). Fractional differencing. biometrika 68 165–176. *Mathematical Reviews (MathSciNet): MR614953 Zentralblatt MATH 464.*

LeBaron, B. et al. (2001). Stochastic volatility as a simple generator of apparent financial power laws and long memory. *Quantitative Finance 1*(6), 621–631.

Newey, K. and D. McFadden (1994). Large sample estimation and hypothesis. *Handbook of Econometrics, IV, Edited by RF Engle and DL McFadden*, 2112–2245.

Stock, J. H. and M. W. Watson (1988). Testing for common trends. *Journal of the American statistical Association 83*(404), 1097–1107.

Tiao, G. C. and R. S. Tsay (1989). Model specification in multivariate time series. *Journal of the Royal Statistical Society: Series B (Methodological) 51*(2), 157–195.

Tsay, W.-J. (2010). Maximum likelihood estimation of stationary multivariate ARFIMA processes. *Journal of Statistical Computation and Simulation 80*(7), 729–745.

Vahid, F. and R. F. Engle (1993). Common trends and common cycles. *Journal of Applied Econometrics*, 341–360.

Vahid, F. and R. F. Engle (1997). Codependent cycles. *Journal of Econometrics 80*(2), 199–221.

Appendix

Asymptotic Results

Asymptotic Results of the ACT Estimator

Let $\{Y_t\}$ be a strictly stationary series:

$$Y_t = \sum_{j\in\mathbb{Z}} \psi_j Z_{t-j} \qquad Z_t \sim i.i.d.(0,\sigma^2)$$

satisfying $\sum_{j\in\mathbb{Z}} |\psi_j| < \infty$ and $\mathbb{E}[Z_t^4] = \eta\sigma^4 < \infty$. Assume that $Y_t \sim ACT(\theta_0)$, i.e.,

$$\rho_0(h) \coloneqq \frac{\mathbb{C}(X_t, X_{t+h})}{\mathbb{V}(X_t)} = f(h;\theta_0)$$

$\theta \coloneqq (a,b,c,\omega) \in \Theta \subset \mathbb{R}\times\mathbb{R}_+\times\mathbb{R}_+\times\mathbb{R}$ and $f(h;\theta) \coloneqq \frac{1-a(1-\cos(\omega h))}{1+bh^c}$.

Consider a sample of T observations $\{Y_t\}_{t=1}^T$ and define

$$\widehat{\gamma}_T(h) \coloneqq T^{-1}\sum_{t=1}^{T-h}(Y_t - \bar{Y}_T)(Y_{t+h} - \bar{Y}_T)$$

with $\widehat{\rho}_T(h) \coloneqq \widehat{\gamma}_T(h)/\widehat{\gamma}_T(0)$.

Take a set of selected lags $\mathcal{H} \subset \mathbb{N}$. Define the vector:

$$g(\theta) \coloneqq (\rho(h) - f(h;\theta))_{h\in\mathcal{H}}$$

then θ_0 is the unique solution for $g(\theta) = 0$. The minimum distance estimator for θ_0 would then be the minimizer of the sample objective function

$$\widehat{Q}_T(\theta) \coloneqq -\widehat{g}_T(\theta)' W \widehat{g}_T(\theta)$$

where $\widehat{g}_T(\theta) \coloneqq (\widehat{\rho}_T(h) - f(h;\theta))_{h\in\mathcal{H}}$ and W is any positive semi-definite matrix. Let $\widehat{\theta}_T$ be such a minimizer. We define its asymptotic properties below.

Proposition 1. *(Consistency of $\widehat{\theta}_T$)*

Assume that θ_0 is identified and Θ is a compact and convex subset of $\mathbb{R}\times\mathbb{R}_+\times\mathbb{R}_+\times\mathbb{R}$. Then as $T\to\infty$, $\widehat{\theta}_T \xrightarrow{p} \theta_0$.

Proof. Since the coefficients $(\psi_j)_{j\in\mathbb{Z}}$ are absolutely summable and Z_t has finite fourth moment, from Brockwell, Davis, and Fienberg (1991) we have that $\widehat{\rho}(h) \xrightarrow{p} \rho(h)\ \forall h \in \mathbb{N}$. Therefore

$$\widehat{Q}_T(\theta) \xrightarrow{p} Q(\theta) \coloneqq -g(\theta)' W g(\theta)$$

from the Continuous Mapping Theorem. Since Θ is compact, $\mathbb{E}\left[\sup_{\theta\in\Theta} \|g(\theta)\|\right] < \infty$. The continuity of g and the compactness of Θ allow us to use *Theorem 2.6* from Newey and McFadden (1994) and conclude that $\widehat{\theta}_T \xrightarrow{p} \theta_0$. □

Proposition 2. *(Asymptotic Normality of $\widehat{\theta}_T$)*

Assume that θ_0 is identified and Θ is a compact subset of $\mathbb{R}\times\mathbb{R}_+\times\mathbb{R}_+\times\mathbb{R}$. Let $G(\tilde{\theta}) \coloneqq \nabla_\theta g(\tilde{\theta})$ and assume that $G(\theta_0)WG(\theta_0)'$ is invertible. Then

$$\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) \xrightarrow{d} N(0, \Omega)$$

with

$$\Omega \coloneqq \left[G(\theta_0)WG(\theta_0)'\right]^{-1} G(\theta_0)WVW'G(\theta_0)' \left[G(\theta_0)WG(\theta_0)'\right]^{-1}. \tag{1}$$

Proof. By definition, the estimator $\widehat{\theta}_T$ satisfies the first order condition

$$\nabla_\theta \widehat{Q}_T(\widehat{\theta}_T) = -2\widehat{G}_n(\widehat{\theta}_T)' W \widehat{g}_T(\widehat{\theta}_T) = 0 \tag{2}$$

for $G(\theta) \coloneqq \nabla_\theta \widehat{g}_T(\theta)$, i.e.:

$$\begin{aligned} G(\theta) &= -\left[\nabla_\theta f(h;\theta)'\right]_{h\in\mathcal{H}} \\ &= -\left[\left(\tfrac{\partial}{\partial a}f(h;\theta) \quad \tfrac{\partial}{\partial b}f(h;\theta) \quad \tfrac{\partial}{\partial c}f(h;\theta) \quad \tfrac{\partial}{\partial \omega}f(h;\theta)\right)'\right]_{h\in\mathcal{H}} \\ \nabla_\theta f(h;\theta) &= -\frac{1}{1+bh^c}\left[1 - cos(\omega h) \quad h^c f(h;\theta) \quad h^c \log h f(h;\theta) \quad a\sin(\omega h).\right] \end{aligned} \tag{3}$$

From the convexity of Θ and the continuity of g, we can expand g around θ_0 in (2):

$$\begin{aligned} 0 &= G(\widehat{\theta}_T)WG(\theta_0) + G(\widehat{\theta}_T)WG(\theta_0)'\left(\widehat{\theta}_T - \theta_0\right) + o_p(1) \\ &= G(\theta_0)WG(\theta_0) + G(\theta_0)WG(\theta_0)'\left(\widehat{\theta}_T - \theta_0\right) + o_p(1) \\ \left(\widehat{\theta}_T - \theta_0\right) &= \left[G(\theta_0)WG(\theta_0)'\right]^{-1} G(\theta_0)WG(\theta_0) + o_p(1) \\ \sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) &= \left[G(\theta_0)WG(\theta_0)'\right]^{-1} G(\theta_0)W\sqrt{T}G(\theta_0) + o_p(1) \end{aligned} \tag{4}$$

From *Theorem 7.2.1* in Brockwell, Davis, and Fienberg (1991) we have that as $n\to\infty$:

$$\begin{aligned} \sqrt{T}\widehat{g}_T(\theta_0) &= \sqrt{T}\left(\widehat{\rho}_T(h) - f(h;\theta_0)\right)_{h\in\mathcal{H}} \\ &= \sqrt{T}\left(\widehat{\rho}_T(h) - \rho(h)\right)_{h\in\mathcal{H}} \xrightarrow{d} N(0, V) \end{aligned}$$

where $V \coloneqq [v_{i,j}]_{i,j\in\mathcal{H}}$ with

$$\begin{aligned} v_{i,j} \coloneqq \sum_{k\in\mathbb{N}} &\{\rho(k+i)+\rho(k-i)-2\rho(i)\rho(k)\}\times \\ &\{\rho(k+j)+\rho(k-j)-2\rho(j)\rho(k)\}. \end{aligned} \tag{5}$$

Therefore we have that

$$\sqrt{T}\left(\widehat{\theta}_T-\theta_0\right)\xrightarrow{d} N(0,\Omega) \tag{6}$$

with $\Omega \coloneqq \left[G(\theta_0)WG(\theta_0)'\right]^{-1}G(\theta_0)WVW'G(\theta_0)'\left[G(\theta_0)WG(\theta_0)'\right]^{-1}$.

□

It is also clear from the above definition of Ω that the optimal choice of W is the inverse of the covariance matrix of $\widehat{g}_T(\theta_0)$, i.e., $W = V^{-1}$ or any of its consistent estimators (either using empirical ACF or using a first-step estimator of θ_0).

Estimation of V

Proposition 3. *(Uniform Convergence of ρ)*

For any consistent estimator $\widehat{\theta}$ of θ_0, the function $\rho(h,\widehat{\theta})$ converges uniformly in h for $\rho(h,\theta_0)$.

Proof. As $\rho(h,\theta)$ is continuous in θ we have that $\forall h\in\mathbb{N}$, $\rho(h,\widehat{\theta})\xrightarrow{p}\rho(h,\theta_0)$ by the Continuous Mapping Theorem. For uniform convergence, fix $h\in\mathbb{N}$ and perform a first-order expansion of $\rho(h,\widehat{\theta})$ around θ_0:

$$\begin{aligned} \rho(h,\widehat{\theta}) &= \rho(h,\theta_0)+\nabla_\theta\rho(h,\theta_0)'(\tilde{\theta}-\theta_0) \\ \left|\rho(h,\widehat{\theta})-\rho(h,\theta_0)\right| &= \left|\nabla_\theta\rho(h,\theta_0)'(\tilde{\theta}-\theta_0)\right| \end{aligned} \tag{7}$$

for some $\tilde{\theta}\in\Theta$. Recall the expression of $\nabla_\theta\rho(h,\theta_0)$:

$$\nabla_\theta\rho(h,\theta_0) = -\left[\tfrac{1-cos(\omega_0 h)}{1+b_0h^{c_0}} \quad \tfrac{h^{c_0}\rho(h,\theta_0)}{1-b_0h^{c_0}} \quad \tfrac{h^{c_0}\log h\rho(h,\theta_0)}{1+b_0h^{c_0}} \quad \tfrac{a_0\sin(\omega_0 h)}{1+b_0h^{c_0}}\right] \tag{8}$$

We have that $\rho(h,\theta_0)$ is bounded; $h^{c_0}/(1+b_0h^{c_0})\to 1$ and $\log h/(1+b_0h^{c_0})\to 0$. Therefore, there exists $K<\infty$ such that $\sup_{h\in\mathbb{N}}|\nabla_\theta\rho(h,\theta_0)|\le K\iota$, for $\iota \coloneqq \begin{pmatrix}1 & 1 & 1 & 1\end{pmatrix}'$. So

$$\sup_{h\in\mathbb{N}}\left|\rho(h,\widehat{\theta})-\rho(h,\theta_0)\right| = \sup_{h\in\mathbb{N}}\left|\nabla_\theta\rho(h,\theta_0)'(\tilde{\theta}-\theta_0)\right|\xrightarrow{p}0. \tag{9}$$

□

Recall the definition of V:

$$\begin{aligned} V &\coloneqq [v(i,j)]_{i,j\in\mathcal{H}} \\ v(i,j) &\coloneqq \sum_{k=1}^{\infty} \{\rho(k+i)+\rho(k-i)-2\rho(i)\rho(k)\} \times \{\rho(k+j)+\rho(k-j)-2\rho(j)\rho(k)\}. \end{aligned} \tag{10}$$

For ease of notation, define $u_{i,j}(k)$ as an element of the sum in $v(i,j)$ and its corresponding input estimator as $\widehat{u}_{i,j}(k)$. For any $i,j \in \mathcal{H}$, consider the truncation of $v_{i,j}$ defined by $\tilde{v}_{i,j}$ defined by replacing ρ with $\widehat{\rho}_T$ in (10) and the feasible estimator $\widehat{v}_{i,j}$ of $\tilde{v}_{i,j}$:

$$\tilde{v}(i,j) \coloneqq \sum_{k=1}^{M(T)} u_{i,j}(k) \qquad \widehat{v}_T(i,j) = \sum_{k=1}^{M(T)} \widehat{u}_{i,j}(k) \tag{11}$$

where M(T) is an increasing function in T.

Proposition 4. *(Consistency of $\widehat{v}_T(i,j)$)*

Assume that the autocorrelations are absolutely summable, i.e., $\sum_{h\in\mathbb{N}} \rho(k) \leq C < \infty$ for some $C \in \mathbb{R}$. Assume that $M(T) = O(T^{\kappa})$ for some $\kappa < 1/2$. Then as $T \to \infty$ the feasible estimator $\widehat{v}_T(i,j)$ converges in probability to $v(i,j)$.

Proof. Recall the definition of $\widehat{v}_T(i,j)$:

$$\begin{aligned} \widehat{v}_T(i,j) &\coloneqq \sum_{k=1}^{M(T)} \widehat{u}_{i,j}(k) \\ \widehat{v}_T(i,j) - v(i,j) &= \widehat{v}_T(i,j) \pm \tilde{v}(i,j) - v(i,j) \\ |\widehat{v}_T(i,j) - v(i,j)| &\leq |\widehat{v}_T(i,j) - \tilde{v}(i,j)| + |\tilde{v}(i,j) - v(i,j)| \\ &\leq \sum_{k=1}^{M(T)} |\widehat{u}_{i,j}(k) - u_{i,j}(k)| + \sum_{k>M(T)} |u_{i,j}(k)|. \end{aligned} \tag{12}$$

From Proposition 3 we have that $\exists B \in \mathbb{R}$ s.t. $\sup_{k\in\mathbb{N}} |\widehat{u}_{i,j}(k) - u_{i,j}(k)| \leq B\left(\tilde{\theta} - \theta_0\right)$ for some $\tilde{\theta}$ consistent for θ_0 as $T \to \infty$. Therefore

$$\begin{aligned} |\widehat{v}_T(i,j) - v(i,j)| &\leq \sum_{k=1}^{M(T)} B\left(\tilde{\theta} - \theta_0\right) + \sum_{k>M(T)} |u_{i,j}(k)| \\ &\leq M(T)B\left(\tilde{\theta} - \theta_0\right) + \sum_{k>M(T)} |u_{i,j}(k)|. \end{aligned} \tag{13}$$

Since the autocovariances are absolutely summable, $u_{i,j}(k)$ is absolutely summable. Therefore the second term on the RHS is $o(1)$. Moreover, as $\widehat{\theta} - \theta_0$ is $O_p(T^{-1/2})$. Finally, taking $M = O(T^{\kappa})$ and assuming $\kappa < 1/2$

$$\begin{aligned} |\widehat{v}_T(i,j) - v(i,j)| &\leq M(T)O(T^{-1/2}) + o(1) \\ &\leq O_p(T^{\kappa-1/2}) + o(1). \end{aligned} \tag{14}$$

Then, as $\kappa - 1/2 < 0$ the RHS converges (in probability) to zero and $\widehat{v}_T(i,j)$ is consistent for $v(i,j)$. □