Heteroskedasticity-Autocorrelation Robust Covariance Estimation Under Non-stationary Covariance Processes

Tiantian Qiu
tianqiu@fas.harvard.edu
March 12, 2008

Abstract

The need to estimate variance-covariance matrix in a time series regression arises often in economic applications involving macroeconomic or finance data. In this paper, we study the behavior of two most popular covariance matrix estimators, namely the Kiefer, Vogelsang and Bunzel kernel estimator without truncation (Kiefer, Vogelsang and Bunzel 2000, KVB thereafter) and standard consistent kernel estimators as in Andrews (1991), under the assumption of permanent changes in the covariance process of the regressor and error term. We show that under general form of time-varying covariance structure, KVB statistics no longer have the asymptotic pivotal distributions as claimed. Rather they depend on the covariance process explicitly. On the other hand, the class of traditional heteroskedasticity-autocorrelation consistent covariance estimators are still consistent under time-varying covariance. In other words, t-statistic using this class of covariance estimators will have standard normal distribution, at least asymptotically. As more and more researchers switch to KVB from traditional HAC covariance estimators, due to their simplicity and better finite sample performance, this paper point out a source of problems which may arise from using their method. It is not robustness to certain form of heteroskedasticity, namely covariance that changes persistently. We provide brief reviews on empirical evidence of such structures in many macroeconomic and finance data.
1 Introduction

In time series regression models, regressors and error terms often exhibit unknown forms of heteroskedasticity and autocorrelation. A traditional approach is to estimate the variance-covariance matrix non-parametrically using consistent kernel estimators and use t-statistic normalized by the HAC (heteroskedasticity autocorrelation consistent) standard errors, under the results that it converges to standard normal distribution asymptotically. An alternative test statistic proposed by Kiefer, Vogelsang and Bunzel (2000) completely avoid the estimation of the variance-covariance matrix. Their non-singular data dependent stochastic transformation to the OLS estimates are shown to be exactly equivalent to using Bartlett kernel HAC standard error without truncations. (Kiefer and Vogelsang 2002) Similar to traditional HAC estimators, the KVB statistics are robust to serial correlation and certain forms of heteroskedasticity. Furthermore, they are much easier to implement in practice since they do not require the choice of kernel and bandwidth. They are also shown to provide better approximations in finite sample. Most researches have since switched to this new type of statistic for inference.

In this paper, we study the behavior of KVB statistics under more general forms of time varying variance-covariance process. We assume the covariance process $\sigma(t)$ satisfies $\sigma([sT]) = \omega(s)$, where $s \in [0,1]$ and $\omega(.)$ is a non-stochastic function that is semi-positive definite at each $s$. We allow the process to be discontinuous, but only at finitely many points and everywhere else it satisfies the first order Lipschitz condition. One example that would fit into such variance-covariance processes is the volatility decline of many macroeconomic variables in the mid 80’s, now known as the Great Moderation. We also extend the assumptions to stochastic volatility set-ups, in which case, volatility converges to a non-negative function of jump-diffusion processes. We show that under these kinds of non-stationary time-varying variance-covariance structure, the KVB statistics no longer have the asymptotic distributions as claimed. The structure of the variance-covariance matrix will in general show up in the distribution of the statistics. We give examples of specific time-varying covariance structures and demonstrate the substantial size distortions as a result. While the popular and simple KVB statistics fail to perform under our more general heteroskedasticity assumptions, the class of traditional HAC estimators stick to the name. We show that the kernel estimators considered in Andrews (1991) are indeed robust to general forms of (deterministic) time-varying covariance structures. Asymptotically it gives the consistent estimate of the covariance and hence t-statistic normalized by HAC standard errors has standard normal distribution.
In this paper, we focus on time series models with time-varying covariance process. This type of structure has received much attention in the econometrics literature. Many Macroeconomic and financial variables, observed over a long time span, display changes in variances over time. Watson (1999), for example, shows that the variance of US long term interest rate has significantly increased over the 1965-1998 period, while variance of the short term rate has decreased. McConnell and Perez Quiros (2000) document a decrease in US GDP volatility over the past twenty years. Sensier and Van Dijk (2004) find that approximately 80% of the 214 U.S. monthly macroeconomic variables have experienced a significant break in volatility. They also argue that most of the breaks occurred as a sudden decline in volatility which happened after 1980. Stock and Watson (2003) provide an comprehensive review on this great moderation phenomenon, which spread across different U.S. sectors and among other developed countries.

In the time series literature, there have been many studies on inference problems of various unit root tests under time-varying volatilities. Hamori and Tokihisa (1997) consider the inference of Dickey-Fuller unit root tests with a single increase in innovation variance. They report a moderate spurious rejection of the unit root hypothesis. Kim et. al. (2002) study the more general Dicky-Fuller tests with constant and trend, allowing for a once change in innovation variance (either decrease or increase). They find severe size distortion only when there is a relative early decrease in the variance, contrast to what has been reported by Hamori and Tokihisa. Cavaliere (2004) provides a more general result in which volatility is some non-stochastic function of time with possibly finite number of discontinuities. He shows that this type of volatility process can both inflate or deflate the rejection of commonly used unit root tests.

In this paper, we consider variance-covariance estimation under standard assumption of no unit roots in regressor and error terms. The assumption of deterministic time-varying covariance process is similar to that of Cavaliere (2004) and Cavaliere and Taylor (2005). We also extend the assumption to stochastic volatility processes that vary persistently. In particular, we consider volatility processes that converge to functions of jump-diffusion processes. Hansen (1995) studies regression asymptotics under stochastic volatilities that is a continuous function of local to unit root processes, which in turn converge to functionals of Ornstein-Uhlenbeck processes as sample sizes gets large. The key difference between his and this paper is that he assumes martingale difference error term, hence no need to estimate long run covariance matrix, which is the focus of this paper. We here extend his

---

1 Also see Kim and Nelson 1999; Koop and Potter 2000; Busetti and Taylor 2003
continuous diffusion volatility processes to include jumps. And then claim that this type of jump-diffusion volatility is general enough to capture many empirical properties of economics data. We finally show that all the results about KVB statistics can be transferred from the deterministic volatility case to the stochastic ones.

The paper is organized as the follows. In section 2, we formally introduce the model and assumptions needed to derive the asymptotic results. In the section 3, we derive asymptotic distributions of KVB $t$ and $F$ statistics under these assumptions and provide simple examples and conduct Monte-Carlo simulations to demonstrate the size distortion that may arise. We also extend the assumption of deterministic covariance structure to stochastic ones in the one-dimensional case and show that similar results apply. In section 4 we show the consistency of traditional HAC estimators under fairly general conditions. Section 5 provides a brief review on empirical evidence of time-varying covariance structures in economic data. Finally we conclude.

2 The Model and Assumptions

Consider time series regression of the following form

$$y_t = X'_t \beta + u_t,$$

where $t = 1, 2, \ldots, T$. $\beta$ is a $k \times 1$ vector of regression coefficients and $X_t$ is a $k \times 1$ vector of regressors that may include constant. Conditioning on $X_t$, $u_t$ is a zero mean random variable. It is assumed in this paper that neither $X_t$ nor $u_t$ contains a unit root, but they may be serially correlated and have general form of heteroskedasticity.

The OLS estimator $\hat{\beta}$ is defined as

$$\hat{\beta} = \left( \sum_{t=1}^{T} X_tX'_t \right)^{-1} \sum_{t=1}^{T} X_t y_t,$$

with the long-run variance of $\sqrt{T}(\hat{\beta} - \beta)$ being

$$Var(\sqrt{T}(\hat{\beta} - \beta)) = \left( \frac{1}{T} \sum_{t=1}^{T} X_tX'_t \right)^{-1} \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} EX_s u_s (X_tu_t)' \left( \frac{1}{T} \sum_{t=1}^{T} X_tX'_t \right)^{-1}.$$

Since $X_t$ is observable, the only interesting term in the expression above is $\frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} EX_s u_s (X_tu_t)'$, which is the long-run covariance matrix of $X_tu_t$. Throughout the paper, expect in section 3.2 where stochastic version of the assumptions are used, we assume the following assumptions hold.
**Assumption 1**

\[ V_t \equiv X_t u_t = \delta_t \epsilon_t, \]

1. \( \delta_{[sT]} = \omega(s) \) is a \( k \times k \) matrix of functions defined on \( s \in [0, 1] \). \( \sup ||\omega(s)|| = c < \infty \) and satisfies the first order Lipschitz condition, except perhaps at a finite number of discontinuities.
2. \( \epsilon_t \) is a \( k \times 1 \) second order stationary process with \( E(\epsilon_t) = 0 \) and \( E(\epsilon_t \epsilon_t') = I \) by normalization. For some \( p > r > 2 \), it has the mixing coefficient \( \alpha_m \) of size \( -pr/(p-r) \) and

\[
\sup_{t \geq 1} ||\epsilon_t||_p = e < \infty.
\]

The long-run covariance matrix of \( \sum_{t=1}^{T} \epsilon_t \)

\[
\lim_{T \to \infty} E(T^{-1}(\sum_{t=1}^{T} \epsilon_t)(\sum_{t=1}^{T} \epsilon_t)') \equiv \Sigma_e,
\]

is well-defined and positive definite.

**Assumption 2**

\[ X_t = p_t \epsilon_t, \]

1. \( p_{[sT]} = \lambda(s) \), which satisfies same conditions as \( \omega(s) \) above.
2. \( \epsilon_t \) is a strong mixing sequence satisfying same conditions as \( \epsilon_t \), with long run positive definite covariance matrix denoted as \( \Sigma_e \).

Here we assume that the variance-covariance process evolves as a deterministic function, which is smooth except at finitely many discontinuous points. Examples of such processes include permanent shifts covariance processes, trending covariance processes and periodical covariance processes. Cavaliere (2004) gives a detailed description of them. Alternatively, we can allow the covariance to converge to some stochastic processes as sample size gets large, such as a diffusion process in Hansen (1995) or a jump process as introduced by Georgiev (2002). In the next section, we will state this alternative stochastic set-up formally with examples.

A special case of assumptions 1 and 2 is to let \( \omega(s) \) and \( \lambda(s) \) be constants over time. This corresponds to the assumptions stated in KVB. We will see later from Theorem 1 and 2 that, our general asymptotic results will simplify to the same distributions as derived in KVB under this special case.

We assume that the identity covariance matrix processes \( \epsilon_t \) and \( \epsilon_t \) are weakly stationary and strong mixing. This rules out unit roots in either the regressor and regression error term, but allows autocorrelations as long
as they vanish fast enough between two distant terms. It is worth noting
that conditional heteroskedasticity that is not very persistent and averages
out in the large sample will be allowed in the weakly stationary sequence
\( \varepsilon_t \) and \( \epsilon_t \). It does not affect the asymptotic results stated in KVB. Only
when heteroskedasticity becomes persistent enough, which results in non-
constant functions \( \omega(s) \) and \( p(s) \), we need to pay special attention to the
asymptotic distributions. Similar conclusions have been made on various
unit root tests by previous studies. For example, if volatility process fol-
low a stationary GARCH type specification, with well-defined and constant
unconditional variance, Dicky-Fuller tests remain valid asymptotically (Kim
and Schmidt 1993; Boswijk 2001). Similarly, stationary Markov switching
volatility does not result in much trouble in large sample of the unit root
test either (Nelson et. al. 2001; Cavaliere 2003). It is only when condi-
tional heteroskedasticity becomes persistent enough that problems start to
arise. Hansen (1995) derives non-standard regression asymptotics when con-
ditional volatility is a positive transformation of diffusion processes. Boswijk
(2001) finds severe size distortion of Dickey Fuller test when volatility follows
a nearly integrated GARCH. Kim et. al. (2002) and Cavaliere (2004) assume
deterministic non-stationary volatility and study the behavior of various unit
root tests.

3 Kiefer-Vogelsang Covariance Estimator Without Truncation

3.1 Some Asymptotic Results

Traditional inference technique on \( \beta \) requires a consistent estimate of

\[
\Omega \equiv \text{Var}(\sqrt{T}(\hat{\beta} - \beta)),
\]

and then apply standard t-statistic asymptotic result

\[
\hat{\Omega}^{-1/2} \sqrt{T}(\hat{\beta} - \beta) \sim N(0, I),
\]

where \( \hat{\Omega}^{-1/2} \) is the Cholesky decomposition of \( \hat{\Omega} \).

Kiefer, Vogelsang and Bunzel (2000) take a different approach. They
transform \( \sqrt{T}(\hat{\beta} - \beta) \) using a moment matrix constructed from the data that
does not require an estimate of \( \Omega \). The nuisance parameter \( \Omega \) is canceled out

\(^2\)For asymptotic results in time series regression with persistent regressor and time-
varying covariance structure, see Qiu (2007).
in the transformation and the asymptotic distribution of their test statistic
does not contain any nuisance parameter.

We will now re-state KVB test statistics and their asymptotic distribu-
tions under our non-stationary covariance assumptions 1 and 2. We will
denote $S_t = \sum_{j=1}^{t} X_j u_j = \sum_{j=1}^{t} V_j$ and $\tilde{S}_t = \sum_{j=1}^{t} X_j \hat{u}_j = \sum_{j=1}^{t} \hat{V}_j$. For a
positive definite matrix $M$, we let $M^{1/2}$ denote the (unique) lower triangular
Cholesky decomposition of $M$.

3.1.1 Testing Individual $\beta$ using $t^*$ Statistic

We firstly study statistic that tests for individual hypothesis. The statistic
proposed by KVB in this case is similar to standard $t$-statistic. But instead
of converging to standard normal, it converges to more complicated distribu-
tion based on Brownian motions. We look at this simple case to gain some
intuition before getting into more complicated statistic that tests general
linear hypothesis.

**Definition 1.** The statistic $t^*$, testing for individual hypothesis on $\beta$, is
defined as 
\[ t^* = \tilde{M}^{-1} T^{1/2} (\tilde{\beta} - \beta), \]
where $\tilde{M} = (T^{-1} \sum_{t=1}^{T} X_t X_t')^{-1} \tilde{C}^{1/2}$ and $\tilde{C} = T^{-2} \sum_{t=1}^{T} \tilde{S}_t \tilde{S}_t'$.

**Theorem 1.** Under Assumptions 1 and 2, we have
\[ t^* \xrightarrow{\text{d}} P^{-1} A(1), \]
where $P = \left[ \int_0^r \{ A(\alpha) - b(\alpha) b(1)^{-1} A(1) \} \{ A(\alpha) - b(\alpha) b(1)^{-1} A(1) \}' d\alpha \right]^{1/2}$.

$A(r) = \int_0^r \omega(s) \tilde{\Sigma}^{1/2} dW_k(s)$ and $b(r) = \int_0^r \lambda(s) \lambda(s)' ds$ with $W_k(s)$ being a
$k$-dimensional standard Brownian motion.

Recall that under the assumptions stated in KVB, the asymptotic distribu-
tion of $t^*$ is independent of the nuisance parameters. In particular,
\[ t^* \xrightarrow{\text{d}} Z^{-1} W_k(1), \]
where $Z = (\int_0^1 (W_k(r) - r W_k(1))(W_k(r) - r W_k(1))', r W_k(1))^1/2$. In order to better
compare our convergence result in Theorem 1 with that derived in KVB, we
rewrite the asymptotic distribution of $t^*$ in the following corollary.

**Corollary 1.** Alternatively, we can write the asymptotic distribution as
\[ t^* \xrightarrow{\text{d}} Q^{-1} W_k(1) \]
where \( Q = \int_0^1 \{G(r) - F(r)W_k(1)\} \{G(r) - F(r)W_k(1)\}'dr \)^{1/2}. \( G(r) = h(1)^{-1}A(r) \), is a Gaussian process with quadratic variation 
\[
[G(r), G(r)] = h(1)^{-1}h(r)h(r)'(h(1)^{-1})'. 
\]
\( F(r) \) is a \( k \times k \) function of \( r \) such that \( F(r) = h(1)^{-1}b(r)b(1)^{-1}h(1) \) and 
\[\begin{align*}
h(r) &= \left( \int_0^r \omega(s)\Sigma_s\omega(s)'ds \right)^{1/2}. 
\end{align*}\]

Corollary 1 gives us an expression directly comparable to the pivotal distribution in KVB. It is easy to see that when \( \omega(s) \) and \( \lambda(s) \) are constant functions, we get back their results. We will show an example in the one-dimensional set-up.

Example
Consider the following simple model
\[ y_t = \beta x_t + u_t, \]
where \( \beta \) is a constant and \( x_t \) is a one-dimensional random variable. We are interested in testing the hypothesis of
\[ H_0 : \beta = \beta_0. \]
Under the assumptions in KVB, under the null hypothesis, the statistic \( t^* \) has convergence result
\[
t^* \sim \frac{1}{\sqrt{T}} \sum_{i=1}^T u_t x_t \implies \frac{W(1)}{\sqrt{\sum_{i=1}^T \hat{S}_i^2}} \implies \sqrt{\int_0^1 (W(r) - rW(1))^2dr}. 
\]
If instead we assume that assumption 1 and 2 hold here for sequences of random variables \( \{u_t x_t\} \) and \( \{x_t\} \), then
\[
t^* \implies \frac{W(1)}{\sqrt{\int_0^1 (W_g(r) - f(r)W(1))^2dr}}, 
\]
where \( W_g(r) \) is a transformed Brownian motion as defined in Davidson (1994) with finite dimensional distribution \( W_g(r) \sim W(g(r)) \). \( g(r) \) and \( f(r) \) are functions of \( r \) defined as \( g(r) = \int_0^r \omega^2(s)ds \) and \( f(r) = \int_0^r \lambda^2(s)ds \). Notice that when \( \omega(s) \) and \( \lambda(s) \) are constant, \( g(r) = f(r) = r \), we are back to the result derived in KVB. But in the more general case of time varying volatilities, the standard KVB result does not hold.

\[\begin{align*}
\text{We can also represent it as} \\
dG(s) &= h(1)^{-1}\omega(s)\Sigma_s^{1/2}dW_k(s). 
\end{align*}\]
3.1.2 Testing General Linear Hypothesis using Statistic $F^*$

Recall the following multivariate regression model

$$y_t = X_t' \beta + u_t,$$

where $\beta$ is a $k \times 1$ vector of coefficients and $X_t$ is a $k \times 1$ random vector. Assume now instead of testing hypothesis on individual $\beta'$, we want to test general linear hypothesis of the form

$$H_0 : R\beta = r$$

$$H_1 : R\beta \neq r.$$

Here $R$ is a $q \times k$ matrix with rank $q$ and $r$ is a $q \times 1$ vector.

KVB propose test statistic $F^*$ for general linear hypothesis. It is similar to the classic $F$ test except that a matrix transformation replaces the HAC estimate of covariance process. We will now restate the $F^*$ test and derive its asymptotic distribution under more general conditions.

**Definition 2.** The statistic $F^*$ testing the general linear hypothesis above is defined as

$$F^* = T(R\hat{\beta} - r)'[R\hat{B}R'](R\hat{\beta} - r)/q,$$

where $\hat{B} = (T^{-1} \sum_{t=1}^{T} X_tX_t')^{-1}\hat{C}(T^{-1} \sum_{t=1}^{T} X_tX_t')^{-1}$ and $\hat{C}$ is defined as in Definition 1.

**Theorem 2.** Under Assumptions 1 and 2, we have

$$F^* \implies \frac{1}{q}(Rb(1)^{-1}A(1))'[Rb(1)^{-1}PP'b(1)^{-1}R']^{-1}(Rb(1)^{-1}A(1))$$

Where $A(r)$, $b(r)$ and $P$ are defined as in Theorem 1.

Notice that here $Rb(1)^{-1}$ is a $q \times k$ matrix of rank $q$. when $q = k$, the asymptotic distribution takes the particular form of

$$\frac{1}{q}A(1)'[P^{-1}A(1)]'[P^{-1}A(1)],$$

which is the square of $t^*$ in Theorem 1.

Again we want to compare the asymptotic distribution in Theorem 2 with the pivotal distribution of $F^*$ stated in KVB. According to Theorem
1 of their result, under the null hypothesis, statistic $F^*$ has the asymptotic distribution

$$F^* \longrightarrow \frac{1}{q} W_q(1) [\int_0^1 (W_q(r) - rW_q(1))(W_q(r) - rW_q(1))' dr]^{-1} W_q(1).$$

We now can rewrite the asymptotic distribution in Theorem 2 in a comparable fashion. As before, the expression in Corollary 2 simplifies to the KVB result under constant $\omega(s)$ and $\lambda(s)$.

**Corollary 2.** We assume that assumption 1 and 2 hold, then

$$F^* \longrightarrow \frac{1}{q} W_q(1)' O^{-1} W_q(1),$$

where $O$ is a $q \times q$ process defined as

$$O = \int_0^1 \{H(r) - I(r)W_q(1)\} \{H(r) - I(r)W_q(1)\}' dr,$$

$$H(r) = \Lambda^{-1} R b(1)^{-1} A(r)$$ is a $q$ dimensional Gaussian process with Quadratic variation

$$[H(r), H(r)] = \Lambda^{-1} R b(1)^{-1} h(r) h(r)' b(1)^{-1} R' (\Lambda^{-1})'.$$

$I(r)$ is $q \times q$ function of $r$, defined as

$$I(r) = \Lambda^{-1} [R b(1)^{-1} h(1) h(1)' b(1)^{-1} b(r) b(1)^{-1} R']^{1/2}.$$

Here $\Lambda = [R b(1)^{-1} h(1) h(1)' b(1)^{-1} R']^{1/2}$, $h(r)$ and $b(r)$ are the same as defined before.

### 3.2 Stochastic Covariance Structure

We assume in assumptions 1 and 2 that covariance processes are some deterministic functions that change persistently. Alternatively we can model them as stochastic variables. Hansen (1995) studies asymptotics of time series regressions with martingale difference error term and stationary regressor. In particular, he models the conditional volatilities of regressor and regression error term as continuous functions of local to unit root processes that converge to diffusion processes. Boswijk (2001) chooses nearly integrated GARCH(1,1) framework in his study of unit root tests, which also

---

4We can also represent it as $dH(s) = \Lambda^{-1} R b(1)^{-1} \omega(s) \Sigma^{1/2} dW_k(s)$.  

10
converges to diffusion processes as sample size gets large. Both of them obtain non-standard asymptotic distributions of test statistics. Here we will borrow Hansen’s framework and extend it to include a more general class of stochastic volatility processes. We show that Theorem 1 and 2 will continue to hold under our stochastic covariance set-up. For simplicity, we will restrict ourselves to the one dimensional world, i.e. \( k = 1 \).

We want to include stochastic non-stationary variations in the volatility processes. To do this, we can either allow the volatility to have a local to unit root like in Hansen (1995), or we can allow rare permanent mean shifts. To model jumps in a stochastic fashion, we adopt a particular form of stochastic structure break model by Georgiev (2002) and Leipus and Viano (2003). The breaks in the form of permanent level shifts arrive randomly each period and they are asymptotically rare. In particular, jumps in volatilities occur with probability \( l/T \) each period, for some constant \( l \). Notice that they are inversely proportional to length of the time \( T \), so that breaks become relatively rare as \( T \to \infty \). Since the expected number of breaks is a fixed number as \( T \to \infty \), we say that breaks occur asymptotically rare. This particular kind of stochastic break model is the natural generalization (by a conditioning argument) of the deterministic finite break models. It allows multiple breaks but requires the expected number of breaks to be bounded as sample size \( T \) goes to infinity, by making the probability of a break occurrence proportional to \( 1/T \) at each period. By making the jumps rare in the asymptotic sense, this model fits our intuition of rare structural changes in a otherwise stable system. The weak limit of such process is a compound Poisson process as shown in Georgiev (2002).

We now state the non-stationary stochastic volatility model formally. Recall the model of time series regression

\[
y_t = \beta x_t + u_t.
\]

Instead of Assumptions 1 and 2, we have the following stochastic volatility processes.

**Assumption 1’ and 2’**

\[
x_t = \delta_{1t} \epsilon_{1t},
\]

\[
u_t = \delta_{2t} \epsilon_{2t},
\]

where

\[
\delta_{1t} = f(d_1 + \tau_1 S_{1t} + \sum_{s=1}^{t} \pi_s \eta_s),
\]
\[ \delta_{2t} = f(d_2 + \tau_2 S_{2t} + \sum_{s=1}^{t} \nu_s \lambda_s), \]

Here \( f \) is a real positive continuous function. \( d_1 \) and \( d_2 \) are constants. \( \tau_1 = \theta_1/\sqrt{T} \), \( \tau_2 = \theta_2/\sqrt{T} \) with \( \theta_1, \theta_2 \) being constants.

\[
S_{1t} = (1 - c_1/T)S_{1(t-1)} + a_{1t}, \\
S_{2t} = (1 - c_2/T)S_{2(t-1)} + a_{2t},
\]

where \((a_{1t}, a_{2t})\) is a martingale difference sequence with conditional covariance \( \Sigma_\nu \). \( c_1 \) and \( c_2 \) are constants. \((\sum_{s=1}^{t} \pi_s \eta_s, \sum_{s=1}^{t} \nu_s \lambda_s)\) is a two dimensional jump processes defined as the following. \( \pi_t \) and \( \nu_t \) are i.i.d. Bernoulli variables which take the value 1 with probability \( l_\pi/T \) and \( l_\nu/T \) respectively. \((\lambda_t, \eta_t) \sim i.i.d. N(0, S)\). \( \pi_t, \nu_t \) and \((\lambda_t, \eta_t)\) are independent sequences of variables.

(2) \( \epsilon_{1t} \) and \( \epsilon_{1t} \epsilon_{2t} \) are both second order stationary processes, independent of \( \delta_{it} \). \( E(\epsilon_{1t}) = E(\epsilon_{1t} \epsilon_{2t}) = 0 \) and \( E(\epsilon_{1t}^2) = E(\epsilon_{1t} \epsilon_{2t})^2 = 1 \) by normalization. For some \( p > r > 2 \), \( \epsilon_{1t} \) and \( \epsilon_{1t} \epsilon_{2t} \) have mixing coefficient \( \alpha_m \) of size \( -pr/(p-r) \) and finite long run covariance matrices.

\[
\sup_{t \geq 1} ||\epsilon_{1t}||_p < \infty, \\
\sup_{t \geq 1} ||\epsilon_{1t} \epsilon_{2t}||_p < \infty.
\]

Under the alternative assumptions 1' and 2', it follows from Lemma 2 in the Appendix that

\[
\delta_{1[sT]} \xrightarrow{} f(d_1 + \theta_1 W^c_1(s) + J_\pi(s)), \\
\delta_{2[sT]} \xrightarrow{} f(d_2 + \theta_2 W^c_2(s) + J_\nu(s)),
\]

where \( W^c_1 \) and \( W^c_2 \) are diffusion processes. \( J_\pi(s) \) and \( J_\nu(s) \) are compound Poisson processes. As we will see later, this weak limit distribution is general enough to account for many interesting empirical properties of the volatility process, one of which is the long memory property that received much attention in the literature.

We now state the asymptotic distribution of \( t^* \) statistic under stochastic volatility assumptions.

**Corollary 3.** Under assumptions 1' and 2', we have

\[
t^* \xrightarrow{} \frac{W(1)}{\sqrt{\int_0^1 (K(r) - L(r)W(1))^2 dr}},
\]

12
where \( K(r) \) and \( L(r) \) are stochastic processes defined as

\[
K(r) = \frac{\int_0^r f(d_1 + \theta_1 W_1^c(s) + J_\pi(s)) f(d_2 + \theta_2 W_2^c(s) + J_\nu(s)) dW(s)}{\sqrt{\int_0^1 f(d_1 + \theta_1 W_1^c(s) + J_\pi(s))^2 f(d_2 + \theta_2 W_2^c(s) + J_\nu(s))^2 ds}}
\]

\[
L(r) = \frac{\int_0^r f(d_1 + \theta_1 W_1^c(s) + J_\pi(s))^2 ds}{\int_0^1 f(d_1 + \theta_1 W_1^c(s) + J_\pi(s))^2 ds}.
\]

Here \( W_1^c \) and \( W_2^c \) are diffusion processes and \( J_\pi \) and \( J_\nu \) are compound Poisson processes as defined in Lemma 2.

Notice that since the jump processes are independent of the Brownian motion in our set-up, \( K(r) \) has marginal distribution following variance mixture of normal distributions. In particular, we can write the marginal distribution of \( K(r) \) for each \( r \in [0, 1] \) as the following form

\[
K(r) \sim \int N(0, u) dP(u),
\]

where

\[
u = \frac{\int_0^r f(d_1 + \theta_1 W_1^c(s) + J_\pi(s))^2 f(d_2 + \theta_2 W_2^c(s) + J_\nu(s))^2 ds}{\int_0^1 f(d_1 + \theta_1 W_1^c(s) + J_\pi(s))^2 f(d_2 + \theta_2 W_2^c(s) + J_\nu(s))^2 ds}
\]

and \( P(u) \) is the probability measure of \( u \).

We can think of models satisfying assumptions 1’ and 2’ as extensions to standard discrete-time stochastic volatility (SV) models. Not only does it includes stationary time varying heteroskedasticity in \( \epsilon_{it} \), it also allows for non-stationary variations of the stochastic volatilities in \( \sigma_{it} \). In the classic SV models, the logarithm of the squared innovation is assumed to follow a stationary AR(1) process. (Taylor 1986) Assumption 1’ and 2’ allow jumps to be added to this class of autoregressive SV models to better capture some of the empirical properties of economic data, in particular the long range dependence. We now give an example.

**Example: A Log-linear ARSV with Jumps Model**

\[
r_t = e^{h_t/2} \xi_t,
\]

\[
h_t = \alpha + z_t + \sum_{s=1}^t \pi_s \eta_s,
\]
where $z_t = \phi z_{t-1} + \sigma_a a_t$. $\xi_t, a_t$ and $\eta_t$ are uncorrelated standard normal white noise shocks. $|\phi| < 1$, $\alpha$ and $\sigma_a$ are constant parameters. $\sum_{s=1}^{t} \pi_s \eta_s$ is a jump process as described in assumption 1’.

The model described above introduces asymptotically infrequently stochastic jumps to standard autoregressive stochastic volatility framework. Much of the empirical research on discrete-time stochastic volatility has focused on the latter. It has nice applications in the finance literature, serving as discrete version of the continuous-time asset pricing models. (Hull and White 1987; Chesney and Scott 1989; Andersen 1994) It ensure that the volatility process is non-negative without any restrictions on the coefficients. It also allows volatility to be estimated by a linear state-space filter by simply taking the logarithm of the squared innovation. Our generalization keeps the convenient log-linear from, but added jumps to better describe the data. By taking the logarithm of the squared innovation, we get an AR(1) process contaminated by jumps.

$$\log(r_t^2) = \alpha + z_t + \sum_{s=1}^{t} \pi_s \eta_s + log(\xi_t^2).$$

The Jump component can play an important role in explaining the “long memory” property discovered in volatilities of most financial returns. There is now a huge literature on modeling and forecasting volatility using the long memory GARCH or long memory stochastic volatility models. But an alternative explanation for the observed long lasting dependence is structural breaks. Perron and Qu(2006) explains in details how rare breaks can lead to slow decay of the sample autocorrelation function and bias up the log-periodogram estimate of the fractional integration parameter $d$. They also provide test that reject the fractionally integrated process in favor of autoregressive SV model contaminated with jumps. With the evidence provided in their paper, we propose the log-linear ARSV with jump model as a compelling alternative to the more complicated long memory class of models.

3.3 Monte-Carlo Simulation

In the previous section, we gave simple examples to demonstrate how the asymptotic distributions in Theorem 1 and 2 can be substantially different from the pivotal distributions claimed in KVB (2000). Here we will perform Monte-Carlo simulations to study quantitatively the impact of non-stationary covariance structure on the regression inferences, in particular deterministic permanent one-break covariances and stochastic jump-diffusion type covariances.
3.3.1 \( t^* \) statistic

Here we restrict to simple one dimensional model of the form

\[ y_t = \beta x_t + u_t. \]

Let \( u_t = \sigma_t \varepsilon_t \) and \( x_t = p_t \varepsilon_t \), where \( \varepsilon_t \) and \( \varepsilon_t \) are AR(1) processes defined as

\[ \varepsilon_t = \rho_1 \varepsilon_{t-1} + a_t \quad \text{and} \quad \varepsilon_t = \rho_2 \varepsilon_{t-1} + b_t \]

with constant \( \rho \) and \((a_t, b_t) \in i.i.d. N(0, I)\).

We will study both deterministic changes in volatilities and stochastic jump-diffusion type volatilities. We define several volatility processes as the following. \( J_a(s) \) is a compound Poisson process with jump density \( k = 1 \) and sizes random draws from \( N(0, 1) \). \( J_b(s) \) is a compound Poisson process with \( k = 3 \) and sizes \( N(0, 3) \). \( W^c_a(s) \) is an OU process with \( c = -50 \) and \( W^c_b(s) \) is an Brownian motion (OU with \( c = 0 \)). The simulation is conducted with \( T = 1000 \) and \( N = 10000 \) replications.

\[ \text{break0} \quad p(s) = 1, \sigma(s) = 1 \]

\[ \text{break1} \quad p(s) = 2I_{[0,0.5]}(s) + 1I_{[0.5,1]}(s), \sigma(s) = 1 \]

\[ \text{break2} \quad p(s) = 2I_{[0,0.5]}(s) + 1I_{[0.5,1]}(s), \sigma(s) = 2I_{[0,0.5]}(s) + 1I_{[0.5,1]}(s) \]

\[ \text{break3} \quad p(s) = 5I_{[0,0.5]}(s) + 1I_{[0.5,1]}(s), \sigma(s) = 1 \]

\[ \text{break4} \quad p(s) = 5I_{[0,0.5]}(s) + 1I_{[0.5,1]}(s), \sigma(s) = 5I_{[0,0.5]}(s) + 1I_{[0.5,1]}(s) \]

\[ J_1 \quad p(s) = \exp(1 + J_{a_1}(s)), \sigma(s) = \exp(1 + J_{a_2}(s)) \]

\[ J_2 \quad p(s) = \exp(1 + J_{b_1}(s)), \sigma(s) = \exp(1 + J_{b_2}(s)) \]

\[ JD_1 \quad p(s) = \exp(1 + J_{a_1}(s) + W^c_{a_1}(s)), \sigma(s) = \exp(1 + J_{a_2}(s) + W^c_{a_2}(s)) \]

\[ JD_2 \quad p(s) = \exp(1 + J_{b_1}(s) + W^c_{b_1}(s)), \sigma(s) = \exp(1 + J_{b_2}(s) + W^c_{b_2}(s)) \]
<table>
<thead>
<tr>
<th>$\rho = $</th>
<th>-0.5</th>
<th>-0.3</th>
<th>0</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.9</th>
<th>0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>break0</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>break1</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.07</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>break2</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.08</td>
<td>0.08</td>
<td>0.09</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>break3</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.12</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>break4</td>
<td>0.13</td>
<td>0.13</td>
<td>0.12</td>
<td>0.13</td>
<td>0.13</td>
<td>0.15</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>J1</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.07</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>J2</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
<td>0.10</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>JD1</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>JD2</td>
<td>0.09</td>
<td>0.08</td>
<td>0.08</td>
<td>0.09</td>
<td>0.09</td>
<td>0.10</td>
<td>0.12</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Empirical rejection rate using 5% two-sided critical value of the $t^*$ asymptotic distribution derived in KVB. $T = 1000$ Based on 10000 simulations.

### 3.3.2 $F^*$ statistic

We generate data according to the following model with $k = 5$

$$y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t,$$

$$x_{it} = p_{it} e_{it}$$

$$u_t = \sigma_t \epsilon_t.$$

We consider the AR(1) model as well as MA(1) model for the regressors and error term. In the AR(1) model, $e_{it} = \rho e_{it(t-1)} + a_{it}$ and $\epsilon_t = \rho \epsilon_{t-1} + b_t$ with $(a_{it}, b_t) \sim i.i.d.N(0, (1 - \rho^2)I)$. In the MA(1) model, $e_{it} = a_{it} + \theta a_{it(t-1)}$ and $\epsilon_t = b_t + \theta b_{t-1}$ with $(a_{it}, b_t) \sim i.i.d.N(0, \frac{1}{1 - \theta^2}I)$. For each model, we test the hypotheses of $H_0 : \beta_1 = 0; H_0 : \beta_1 = \beta_2 = 0; H_0 : \beta_1 = \beta_2 = \beta_3 = 0; H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$. We label the hypotheses according to the number of restrictions being tested, i.e. $q = 1, 2, 3, 4$ respectively.

As for the $t^*$ statistic, we study both deterministic changes in volatilities and stochastic jump-diffusion type volatilities. The simulation is conducted with $T = 1000$ and $N = 10000$ replications.

**break0**  \( p_i(s) = \sigma(s) = 1 \)

**break1**  \( p_i(s) = \sigma(s) = 2I_{[0,0.5]}(s) + 1I_{[0.5,1]}(s), \)

**break2**  \( p_i(s) = \sigma(s) = 5I_{[0,0.5]}(s) + 1I_{[0.5,1]}(s), \)
\[ p_i(s) = \exp(1 + J_{ai}(s) + W_{ai}^c(s)), \quad \sigma(s) = \exp(1 + J_{a0}(s) + W_{a0}^c(s)) \]

\[ p_i(s) = \exp(1 + J_{bi}(s) + W_{bi}^c(s)), \quad \sigma(s) = \exp(1 + J_{b0}(s) + W_{b0}^c(s)) \]

<table>
<thead>
<tr>
<th>( \rho = )</th>
<th>-0.5</th>
<th>-0.3</th>
<th>0</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.9</th>
<th>0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1) ( q = 1 ) break0</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
<td>0.09</td>
</tr>
<tr>
<td>break1</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.08</td>
<td>0.09</td>
<td>0.13</td>
</tr>
<tr>
<td>break2</td>
<td>0.13</td>
<td>0.12</td>
<td>0.12</td>
<td>0.13</td>
<td>0.14</td>
<td>0.14</td>
<td>0.16</td>
<td>0.20</td>
</tr>
<tr>
<td>AR(1) ( q = 2 ) break0</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.06</td>
<td>0.08</td>
<td>0.11</td>
</tr>
<tr>
<td>break1</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.08</td>
<td>0.11</td>
<td>0.15</td>
</tr>
<tr>
<td>break2</td>
<td>0.17</td>
<td>0.16</td>
<td>0.15</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>0.22</td>
<td>0.27</td>
</tr>
<tr>
<td>AR(1) ( q = 3 ) break0</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.09</td>
<td>0.13</td>
</tr>
<tr>
<td>break1</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.08</td>
<td>0.12</td>
<td>0.17</td>
</tr>
<tr>
<td>break2</td>
<td>0.19</td>
<td>0.18</td>
<td>0.17</td>
<td>0.18</td>
<td>0.18</td>
<td>0.20</td>
<td>0.26</td>
<td>0.33</td>
</tr>
<tr>
<td>AR(1) ( q = 4 ) break0</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.06</td>
<td>0.10</td>
<td>0.16</td>
</tr>
<tr>
<td>break1</td>
<td>0.07</td>
<td>0.06</td>
<td>0.06</td>
<td>0.07</td>
<td>0.06</td>
<td>0.08</td>
<td>0.14</td>
<td>0.21</td>
</tr>
<tr>
<td>break2</td>
<td>0.20</td>
<td>0.19</td>
<td>0.19</td>
<td>0.18</td>
<td>0.20</td>
<td>0.22</td>
<td>0.31</td>
<td>0.40</td>
</tr>
</tbody>
</table>

Table 2: Empirical rejection rate using 5% critical value of the \( F^* \) asymptotic distribution derived in KVB. \( T = 1000 \) Based on 10000 simulations.

4 Traditional HAC Estimators and t-statistic

We show in previous sections that test statistics proposed in KVB have non-standard asymptotic distributions under non-stationary time varying covariance assumption. These results may limit the application of their tests. When there are structural breaks in the covariance process or volatility appears persistent enough to have a trend or unit root, test statistics proposed by KVB experience size distortions asymptotically. In this section we will show that, unlike KVB statistics, the traditional HAC estimators are indeed robust to our general form of heteroskedasticity. In particular, they are consistent under non-stationary covariance assumptions and hence the asymptotic normality of the traditional t-statistic remains valid. One important reason for the popularity of the KVB statistics is their better finite sample
performance compared to the traditional HAC estimators. Our results show that the advantage of using KVB rather than traditional HAC might be offset, even in a small sample, when there is some degree of non-stationarity in the covariance structure.

We now define the class of traditional HAC estimators and give conditions under which they will be consistent when covariance process is time varying. The conditions on the kernel and bandwidth are actually identical to those stated in Hansen (1992). This shows that the general class of traditional HAC estimators are very robust, at least asymptotically.

**Definition 3.** We define the long-run covariance matrix and its kernel estimator as the following

\[ J_T = \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} E X_s u_s (X_t u_t)' = \sum_{j=-T+1}^{T-1} \Gamma_T(j), \]

where \( \Gamma_T(j) = \frac{1}{T} \sum_{t=j+1}^{T} EV_t V_{t-j} \) for \( j \geq 0 \) and \( \Gamma_T(j) = \frac{1}{T} \sum_{t=-j+1}^{T} EV_{t+j} V_t' \) for \( j < 0 \).

\[ \hat{J}_T = \sum_{j=-T+1}^{T-1} k(j/s_T) \hat{\Gamma}(j), \]

where \( \hat{\Gamma}(j) = \frac{1}{T} \sum_{t=j+1}^{T} \hat{V}_t \hat{V}_{t-j}' \) for \( j \leq 0 \), and \( \frac{1}{T} \sum_{t=-j+1}^{T} \hat{V}_{t+j} \hat{V}_t' \) for \( j < 0 \). \( k(.) \) is the kernel function with bandwidth \( s_T \).

**Assumption 3** The long run covariance matrix of \( \epsilon_t \), defined as

\[ J_T = \sum_{j=-T+1}^{T-1} \Gamma_\epsilon(j) \]

satisfies the following condition

\[ \sum_{j=-T+1}^{T-1} j||\Gamma_\epsilon(j)|| \rightarrow e < \infty. \]

We adopt the assumptions on the kernel weights \( k(.) \) and bandwidth parameter \( s_T \) from Hansen (1992). Kernels and bandwidth satisfy assumption 4 and 5 include most of those considered in the literature.

**Assumption 4** For all \( x \in R \), \( |k(x)| \leq 1 \) and \( k(x) = k(-x); k(0) = 1; k(x) \) is continuous at zero and for almost all \( x \in R ; \int_R |k(x)|dx < \infty. \)

\[ ^5 \text{Except we require the serial correlation to vanish faster, as in assumption 3.} \]
Assumption 5 \( s_T \to \infty \), and for some \( q \in (1/2, \infty) \), \( s_T^{1+2q}/T = O(1) \).

Theorem 3. Under assumptions 1 to 5, we have

\[
\hat{J}_T - J_T \to 0,
\]

\[
t \Rightarrow N(0, I).
\]

5 Evidence of Non-stationary Covariance Structures

5.1 Macroeconomic Volatility

There has been a large body of empirical research focusing on time varying volatilities of macroeconomic time series. The most famous example is the Great Moderation in the United States and the rest of the developed world. Great Moderation refers to the fact that, over the past 20 years, the volatility of aggregate economic activity has fallen dramatically in most of the industrialized countries. In the United States, a sharp reduction of output volatility was firstly documented by Kim and Nelson (1999) and McConnell and Perez-Quiros (2000). Both of them estimate 1984 as the time of this break occurrence. Research using international data has found similar behavior in GDP of other developed economies, even though the time and nature of this change are different across them. (van Dijk, Osborn and Sensier 2002; Del Negro and Otrok 2003; Stock and Watson 2003)

When studying individual sectors and time series, decline in volatilities is also found in most of them. Stock and Watson (2003) look at 22 major US series consisting of the main NIPA aggregates and selected other macro series. They found that the volatility of 21 out of 22 series fell in the post 84 period. Van Dijk and Sensier (2004) find that approximately 80% of the 214 U.S. monthly macroeconomic variables have experienced a significant break in volatility.

Given all the evidence of the substantial decline in volatility of macroeconomic time series, when using these series in regression analysis, special attention should be paid when using KVB statistics in inference. Even with a large data set, the empirical rejection rate can still be significantly different from designed. On the other hand, traditional HAC estimator gives a consistent estimate of the covariance structure and hence provide asymptotically valid inference.
5.2 Asset Return Volatility

Asset return volatility is found to be very persistent. Modeling this strong persistence has constituted a major research agenda in the econometrics literature for a number of years. One approach is to look at the so called long memory property and build models that generate this property. The fractionally integrated GARCH (FGARCH) and fractionally integrated stochastic volatility models have gain enormous popularity in modeling and forecasting of financial asset volatility. This type of models generate long range dependence but still maintain the stationarity of the unconditional volatility process.

On the other hand, research has discovered that even though long memory may be a genuine feature of the volatility, it can also occur when there are structural breaks in the process. A short memory process, such as a standard ARMA, contaminated by occasional level shifts can produce long memory properties such as slowly decaying autocorrelation function. It may also bias up the coefficient of fractional integration. Early reference on this includes Granger and Ding (1996), Teversovsky and Taqqu (1997), Lobato and Savin (1998), Diebold and Inoue (2001).

Mikosch and Starica (2004) demonstrate, in theoretical ways, how non-stationarity of the data (deterministic change in variance here) can produce spurious long memory property, namely the slow decaying of sample autocorrelation function and integrated GARCH (IGARCH) effect. Their assumption of finitely many changes in variance level is a special case of assumptions 1 and 2 in this paper. On the other hand, Perron and Qu (2006) applied the stochastic jump model of Georgiev (2002) and show that fractional integration coefficient $d$ is biased up when a short memory process is contaminated by stochastic rare jumps. They also provide empirical evidence showing that it is jumps rather than fractional integration underlying the volatility of stock returns. If it is indeed the non-stationarity in volatility that generates the long range dependence, regression inference with KVB type statistics applying to financial data is also questionable.

5.3 Non-stationary Correlation

We saw that volatilities of many individual time series change over time. It is natural to think that relationship between two time series may also subject to change. One particular example in the financial market is that correlations between international stock markets are changing over time. During market turbulence, correlation seems to rise. King and Wadhwani (1990), Bertero and Mayer (1990), Lee and Kim (1993) claimed that correlations increased
significantly after the 1987 US market crash. Applying more complicated GARCH type models, Longin and Solnik(1995) reject the null hypothesis of constant conditional correlation. They discover an increasing correlation between markets over the past thirty years. They also find that correlation rises in period of high volatility. Similarly, Ramchand and Susmel(1997) find that correlations between the US and other world markets are on average 2 to 3.5 times higher when the US market is in a turbulent (high variance) state as compared to the stable (low variance) state.

6 Conclusion

In this paper, we study the behavior of two most popular covariance matrix estimates under the assumption of non-stationary covariance structures. We assume that the unconditional covariance matrix is deterministic and time varying and show that under this more general assumption, KVB statistics no longer have the pivotal asymptotic distributions as claimed. Rather, the distributions depend on the time-varying covariance structure explicitly. We extend the deterministic assumption to stochastic jump diffusion ones and show that similar conclusion holds. On the other hand, we prove in this paper that traditional HAC estimators are robust even under the non-stationary covariance structure. It provides a consistent estimate of the variance-covariance matrix and hence t-statistic applying HAC estimates is asymptotically normal. At the end we provide a brief review of empirical evidence on non-stationary covariance structure. Since this type of non-stationarity often appears in macroeconomic and financial data, extra care should be taken when we do regression inference with KVB statistics. The empirical size of the tests can be substantially biased when volatility or correlation of the regressor and error term are changing over time.

In this paper, we only focus on the non-stationary covariance structure and the problems it may cause to the LS regression inference. In particular, we only relax the conditional heteroskedasticity assumption to allow for unconditional change of the second moment, but make no change to the autocorrelation part of the assumption. Of course, increasing the degree of autocorrelation could also affect the behavior of various test statistics. For example, unit roots in either the regressor or regression error term will break down the standard asymptotic convergence results. So will a long memory, but still stationary, standard error. These are interesting questions for future research, but not the focus of this paper.
Appendix

Lemma 1. Under Assumptions 1 and 2, for \( r \in [0, 1] \)

\[
T^{-1/2} \sum_{t=1}^{[Tr]} X_t u_t \Rightarrow \int_0^r \omega(s) \Sigma_\epsilon^{1/2} dW(s).
\]

\[
T^{-1} \sum_{t=1}^{[Tr]} X_t X_t' \Rightarrow \int_0^r \lambda(s) \lambda(s)' ds.
\]

Proof of Lemma 1 Firstly, following joint convergence result follows from Phillips and Durlauf (1986), under assumptions 1 and 2.

\[
(\sigma_{[Tr]} \Sigma_\epsilon^{1/2}, \frac{1}{\sqrt{T}} \Sigma_\epsilon^{-1/2} \sum_{t=1}^{[Tr]} \epsilon_t) \Rightarrow (\omega(r) \Sigma_\epsilon^{1/2}, K(r)).
\]

We can rewrite the sum in term of integral, under assumption 1

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} X_t u_t = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \sigma_t \epsilon_t = \int_0^r \sigma_{[Tr]} \Sigma_\epsilon^{1/2} d(\frac{1}{\sqrt{T}} \Sigma_\epsilon^{-1/2} \sum_{t=1}^{[Tr]} \epsilon_t).
\]

Applying Hansen (1992) Theorem 3.1, we have the following

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \sigma_t \epsilon_t - \Lambda_{[Tr]}^* \Rightarrow \int_0^r \omega(s) \Sigma_\epsilon^{1/2} dW_k(s),
\]

where

\[
\Lambda_{[Tr]}^* = \frac{1}{\sqrt{T}} \sum_{i=1}^t (\sigma_i \Sigma_\epsilon^{1/2} - \sigma_{i-1} \Sigma_\epsilon^{1/2}) z_i' - \frac{1}{\sqrt{T}} \sigma_t \Sigma_\epsilon^{1/2} z_{t+1}'
\]

\[
z_i = \sum_{k=1}^{\infty} E_i (\Sigma_\epsilon^{-1/2} \epsilon_{i+k}).
\]

We will now show that for all \( r \in [0, 1] \), \( \Lambda_{[Tr]}^* = o_p(1) \).

\[
\sup_{t \geq 1} ||\Lambda_t^*|| \leq \sup_{t \geq 1} \frac{1}{\sqrt{T}} \sum_{i=1}^t (\sigma_i \Sigma_\epsilon^{1/2} - \sigma_{i-1} \Sigma_\epsilon^{1/2}) z_i' + \sup_{t \geq 1} \frac{1}{\sqrt{T}} \sigma_t \Sigma_\epsilon^{1/2} z_{t+1}'
\]

\[\text{Since } \epsilon_t \text{ is not a martingale difference sequence, we cannot directly apply the well-known stochastic integral convergence result in Kurtz and Protter (1991).} \]
\[
sup_{t \geq 1} \left| \frac{1}{\sqrt{T}} \sigma_t \Sigma^{1/2} z_{t+1} \right| \leq (\sup_{t \geq 1} \left| \sigma_t \right|) \left| \Sigma^{1/2} \right| (\frac{1}{\sqrt{T}} \sup_{t \geq 1} ||z_t||) \to 0,
\]

since \( \frac{1}{\sqrt{T}} \sup_{t \geq 1} ||z_t|| \to 0 \) by proof of Theorem 4.1 in Hansen (1992) and \( \sup_{t \geq 1} ||\sigma_t|| \) is bounded by assumption 1.

\[
\sup_{t \geq 1} \left| \frac{1}{\sqrt{T}} \sum_{i=1}^{t} (\sigma_i \Sigma^{1/2} - \sigma_{i-1} \Sigma^{1/2}) z_i \right|
\leq (\sup_{t \geq 1} \sum_{i=1}^{t} ||\sigma_i - \sigma_{i-1}||) \left| \Sigma^{1/2} \right| (\frac{1}{\sqrt{T}} \sup_{t \geq 1} ||z_t||)
\leq \left( \sum_{i=1}^{T} ||\sigma_i - \sigma_{i-1}|| \right) \left| \Sigma^{1/2} \right| \sup_{t \geq 1} ||z_t|| \to 0
\]

The last line follows from uniform Lipschitz continuity of function \( \omega(r) \) and \( \frac{1}{\sqrt{T}} \sup_{t \geq 1} ||z_t|| \to 0 \) as before. Hence we have the first convergence result

\[
T^{-1/2} \sum_{i=1}^{T} X_t u_t \Rightarrow \int_0^r \omega(s) \Sigma_s^{1/2} dW(s).
\]

To show the second convergence result, notice that

\[
T^{-1} \sum_{t=1}^{T} X_t X_t' = T^{-1} \sum_{t=1}^{T} (p_t e_t' p_t') = T^{-1} \sum_{t=1}^{T} p_t p_t' + T^{-1} \sum_{t=1}^{T} (p_t (e_t' - I) p_t').
\]

clearly \( T^{-1} \sum_{t=1}^{T} p_t p_t' \to \int_0^r \lambda(s) \lambda(s)' ds \) under assumption 2. It remains to show that \( T^{-1} \sum_{t=1}^{T} (p_t (e_t' - I) p_t') = o_p(1) \).

\[
\sup_r \left| \frac{1}{T} \sum_{t=1}^{T} p_t (e_t' - I) p_t' \right| \leq (\sup_r ||\lambda(r)||)^2 \sup_r \left| \frac{1}{T} \sum_{t=1}^{T} (e_t' - E(e_t')) \right| \to 0,
\]

under bounded \( \lambda(s) \) and second order stationary series \( e_t \).

**Proof of Theorem 1**

\[
t^* \equiv \hat{M}^{-1/2} (\hat{\beta} - \beta) = Z^{-1} T^{-1/2} \sum_{t=1}^{T} X_t u_t,
\]

where \( \hat{C} = ZZ' = T^{-2} \sum_{t=1}^{T} \hat{S}_t \hat{S}_t' \) with \( Z \) being the lower triangular Cholesky decomposition of positive definite matrix \( \hat{C} \).
\[
\hat{C} = T^{-2} \sum_{t=1}^{T} \hat{S}_t \hat{S}_t' = \frac{1}{T} \sum_{t=1}^{T} (T^{-1/2} \hat{S}_t)(T^{-1/2} \hat{S}_t')
\]

with

\[
T^{-1/2} \hat{S}_{[rT]} = T^{-1/2} S_{[rT]} - (T^{-1} \sum_{t=1}^{[rT]} X_tX_t') (T^{-1} T^{-1/2} S_T).
\]

Applying continuous mapping theorem to results in Lemma 1 we have

\[
T^{-1/2} \hat{S}_{[rT]} \rightarrow A(r) - b(r)b(1)^{-1} A(1),
\]

Hence

\[
Z \rightarrow \int_0^1 \{A(r) - b(r)b(1)^{-1} A(1)\} \{A(r) - b(r)b(1)^{-1} A(1)\}' dr \]^{1/2}.
\]

Combining this with Lemma 1, we get Theorem 1.

To get the alternative expression in Corollary 1, we denote

\[
h(r)h(r)' = \int_0^r \omega(s) \Sigma \omega(s)' ds
\]

with \(h(r)\) being the lower diagonal matrix Cholesky decomposition. We can then rewrite the stochastic integrals in Theorem 1 as the following

\[
P^{-1}A(1) = P^{-1}h(1)W_k(1) = (h(1)^{-1} P)^{-1}W_k(1).
\]

Notice that because \(h(1)^{-1}\) is lower diagonal

\[
h(1)^{-1} P \equiv h(1)^{-1} \left[ \int_0^r \{A(r) - b(r)b(1)^{-1} A(1)\} \{A(r) - b(r)b(1)^{-1} A(1)\}' dr \right]^{1/2}
\]

\[
= \left[ \int_0^r \{h(1)^{-1}(A(r) - b(r)b(1)^{-1} A(1))\} \{h(1)^{-1}(A(r) - b(r)b(1)^{-1} A(1))\}' \right]^{1/2}
\]

Since \(h(1)\) and \(\omega(s) \Sigma \) are deterministic, \(h(1)^{-1} A(r) = h(1)^{-1} \int_0^r \omega(s) \Sigma^{1/2} dW_k(s)\) is a Gaussian process with quadratic variation

\[
h(1)^{-1} \int_0^r \omega(s) \Sigma \omega(s)' ds (h(1)^{-1})' = h(1)^{-1} h(r) h(r)' (h(1)^{-1})'.
\]

Similarly, \(h(1)^{-1} b(r) b(1)^{-1} A(1) = h(1)^{-1} b(r) b(1)^{-1} h(1) W_k(1) \equiv F(r) W_k(1)\).

These gives us the expression in Corollary 1.
Proof of Theorem 2 and Corollary 2

\[ F^* = T(R\hat{\beta} - r)'[R\hat{\beta}R']^{-1}(R\hat{\beta} - r)/q \]

Firstly, under the null hypothesis of \( R\beta = r \), we have

\[ \sqrt{T}(R\hat{\beta} - r) = R\sqrt{T}(\hat{\beta} - \beta) = R\left(T^{-1}\sum_{t=1}^{T}X_tX'_t\right)^{-1/2}\sum_{t=1}^{T}X_tu_t. \]

By Lemma 1 and 2,

\[ \sqrt{T}(R\hat{\beta} - r) \xrightarrow{D} Rb(1)^{-1}A(1). \]

Secondly,

\[ R\hat{\beta}R' = R\left(T^{-1}\sum_{t=1}^{T}X_tX'_t\right)^{-1}\hat{\beta}C\left(T^{-1}\sum_{t=1}^{T}X_tX'_t\right)^{-1}R' \]

By Lemma 1, 2 and proof of Theorem 1, we have

\[ R\hat{\beta}R' \xrightarrow{D} Rb(1)^{-1}PP'b(1)^{-1}R'. \]

Combining the results above, we have Theorem 2.

To get the alternative expression in Corollary 2, notice that \( Rb(1)^{-1}h(1)A(1) \) can be rewritten as \( \Lambda W_q(1) \), where \( \Lambda \) is the \( q \times q \) square root of \( Rb(1)^{-1}h(1)h(1)'b(1)^{-1}R' \). This is true because \( Rb(1)^{-1}h(1) \) is a matrix of rank \( q \) and \( W_q(1) \) is a vector of independent Brownian motions.

\[ F^* \xrightarrow{D} \frac{1}{q} (Rb(1)^{-1}A(1))'[Rb(1)^{-1}PP'b(1)^{-1}R']^{-1}(Rb(1)^{-1}A(1)) \]

\[ = W_q(1)'\Lambda'[Rb(1)^{-1}PP'b(1)^{-1}R']^{-1}\Lambda W_q(1) \]

\[ = W_q(1)'[\Lambda^{-1}Rb(1)^{-1}PP'b(1)^{-1}R'(\Lambda^{-1})]'^{-1}W_q(1) \]

\( \Lambda^{-1} \) exists because \( Rb(1)^{-1}h(1)h(1)'b(1)^{-1}R' \) is symmetric and of full rank.

\[ \Lambda^{-1}Rb(1)^{-1}PP'b(1)^{-1}R'(\Lambda^{-1})' \]

\[ = \int_0^1 \{\Lambda^{-1}Rb(1)^{-1}(A(r) - b(r)b(1)^{-1}A(1))\}\{\Lambda^{-1}Rb(1)^{-1}(A(r) - b(r)b(1)^{-1}A(1))\}'dr \]

\( \Lambda^{-1}Rb(1)^{-1}A(r) \) can be rewritten as a \( q \) dimensional Gaussian process with quadratic variation

\[ \Lambda^{-1}Rb(1)^{-1}h(r)h(r)'b(1)^{-1}R'(\Lambda^{-1})'. \]

\( \Lambda^{-1}Rb(1)^{-1}b(r)b(1)^{-1}A(1) \) can be rewritten as \( I(r)W_q(1) \) as stated in Corollary 2.
Lemma 2. Under assumptions 1' and 2', we have the following weak convergence results.

\[ \delta_{1[sT]} \Rightarrow f(d_1 + \theta_1 W_1^c(s) + J_\pi(s)), \]
\[ \delta_{2[sT]} \Rightarrow f(d_2 + \theta_2 W_2^c(s) + J_\nu(s)). \]

Here \( W_1^c \) and \( W_2^c \) are Ornstein-Uhlenbeck processes defined as

\[ dW_1^c(s) = -c_1 W_1^c(s) + dW_1(s), \]
\[ dW_2^c(s) = -c_2 W_2^c(s) + dW_2(s). \]

\((W_1, W_2)\) is a vector Brownian motion with covariance matrix \( \Omega_z \). \((J_\pi, J_\nu)\) is a vector compound Poisson process as defined in Georgiev (2002) with jump density \( k_\pi \) and \( k_\nu \) respectively.

Proof of Corollary 3

\[ T^{-1/2} \sum_{t=1}^{[T]} u_t x_t = T^{-1/2} \sum_{t=1}^{[T]} \sigma_{1t} \sigma_{2t} \epsilon_{1t} \epsilon_{2t} \]

Under assumptions 1' and 2', applying Lemma 2 together with Hansen (1992) Theorem 3.1,

\[ T^{-1/2} \sum_{t=1}^{[T]} u_t x_t \Rightarrow a \int_0^r f(d_1 + \theta_1 W_1^c(s) + J_\pi(s)) f(d_2 + \theta_2 W_2^c(s) + J_\nu(s)) dW(s), \]

where \( a \) is the long run variance. By similar argument as in Lemma 1, we also have

\[ T^{-1} \sum_{t=1}^{[T]} x_t^2 \Rightarrow \int_0^r f(d_1 + \theta_1 W_1^c(s) + J_\pi(s))^2 ds. \]

Hence we have

\[ t^* = \frac{T^{-1/2} \sum_{t=1}^{T} u_t x_t}{\sqrt{T^{-2} \sum_{t=1}^{T} \hat{S}_t^2}} \]
\[ \Rightarrow \frac{K^*(r)}{\sqrt{\int_0^1 (K^*(r) - L(r) K^*(1))^2 dr}} \]

where

\[ K^*(r) = a \int_0^r f(d_1 + \theta_1 W_1^c(s) + J_\pi(s)) f(d_2 + \theta_2 W_2^c(s) + J_\nu(s)) dW(s) \]
\[ L(r) = \frac{\int_0^r f(d_1 + \theta_1 W_1^c(s) + J_x(s))^2 ds}{\int_0^1 f(d_1 + \theta_1 W_1^c(s) + J_x(s))^2 ds}. \]

To get the final expression in corollary 3, we divide both the numerator and denominator by \( \sqrt{\int_0^1 f(d_1 + \theta_1 W_1^c(s) + J_x(s))^2 f(d_2 + \theta_2 W_2^c(s) + J_y(s))^2 ds} \).

Since the jump process is independent of the Brownian motion, we have

\[ \sqrt{\int_0^r f(d_1 + \theta_1 W_1^c(s) + J_x(s))^2 f(d_2 + \theta_2 W_2^c(s) + J_y(s))^2 ds} \sim N(0, 1). \]

The rest follows from here.

**Proof of Theorem 3** Define \( \tilde{J}_T = \sum_{j=1}^{T-1} k(j/S_T) \tilde{\Gamma}(j) \), where

\[ \tilde{\Gamma}(j) = \frac{1}{T} \sum_{t=j+1}^T V_t V_{t-j} \text{ for } j \leq 0, \]

\[ \tilde{\Gamma}_j = \frac{1}{T} \sum_{t=-j+1}^T V_{t+j} V'_t \text{ for } j < 0. \]

We will prove Theorem 3 in three steps.

**Step One** \( \hat{J}_T - \tilde{J}_T \to 0 \)

\[ \hat{J}_T - \tilde{J}_T = \sum_{j=0}^{T-1} k(j/S_T) \frac{1}{T} \sum_{t=j+1}^T (\hat{V}_t \hat{V}'_{t-j} - V_t V'_{t-j}) + \sum_{j=-T}^{1} k(j/S_T) \frac{1}{T} \sum_{t=-j+1}^T (\hat{V}_{t+j} \hat{V}'_t - V_{t+j} V'_t). \]

We will show that

\[ A = \frac{\sqrt{T}}{S_T} \left\| \sum_{j=0}^{T-1} k(j/S_T) \frac{1}{T} \sum_{t=j+1}^T (\hat{V}_t \hat{V}'_{t-j} - V_t V'_{t-j}) \right\| = O_p(1), \]

and exactly the same arguments leads to that

\[ B = \frac{\sqrt{T}}{S_T} \left\| \sum_{j=-T}^{1} k(j/S_T) \frac{1}{T} \sum_{t=-j+1}^T (\hat{V}_{t+j} \hat{V}'_t - V_{t+j} V'_t) \right\| = O_p(1). \]

Since \( \frac{\sqrt{T}}{S_T} \left| \hat{J}_T - \tilde{J}_T \right| \leq A + B = O_p(1) \) and \( \frac{T}{S_T} \to \infty \), it follows that

\[ \hat{J}_T - \tilde{J}_T \to 0. \]
\[
A = \frac{\sqrt{T}}{S_T} \left\| \sum_{j=0}^{T-1} k(j/S_T) \frac{1}{T} \sum_t \left( \hat{V}_t \hat{V}_{t-j} - V_t V'_{t-j} \right) \right\|
\]

\[
= \frac{\sqrt{T}}{S_T} \left\| \sum_{j=0}^{T-1} k(j/S_T) \frac{1}{T} \sum_t (-X'_{t-j}(\hat{\beta} - \beta)V_t X'_{t-j} - X'_{j}(\hat{\beta} - \beta)X_t V'_{t-j} + X'_{j}(\hat{\beta} - \beta)X_{t-j}(\hat{\beta} - \beta)X_t \right\|
\]

\[
\leq \frac{\sqrt{T}}{S_T} \left\| \sum_{j=0}^{T-1} k(j/S_T) \sum_t \frac{1}{T} \left[ X'_{t-j}(\hat{\beta} - \beta) \right] V_t X'_{t-j} \right\|
\]

\[
+ \frac{\sqrt{T}}{S_T} \left\| \sum_{j=0}^{T-1} k(j/S_T) \sum_t \frac{1}{T} \left[ X'_{j}(\hat{\beta} - \beta) \right] X_t V'_{t-j} \right\|
\]

\[
+ \frac{\sqrt{T}}{S_T} \left\| \sum_{j=0}^{T-1} k(j/S_T) \sum_t \frac{1}{T} \left[ X'_{j}(\hat{\beta} - \beta)X_{t-j}(\hat{\beta} - \beta) \right] X_t X'_{t-j} \right\|
\]

Firstly,

\[
\frac{\sqrt{T}}{S_T} \left\| \sum_{j=0}^{T-1} k(j/S_T) \sum_t \frac{1}{T} \left[ X'_{t-j}(\hat{\beta} - \beta) \right] V_t X'_{t-j} \right\|
\]

\[
\leq \frac{\sqrt{T}}{S_T} \sum_{j=0}^{T-1} |k(j/S_T)| \sup_t X'_t(\hat{\beta} - \beta) \left\| \sum_t \frac{1}{T} V_t X'_{t-j} \right\|
\]

\[
\leq | \sup_t X'_t \sqrt{T}(\hat{\beta} - \beta) | \left( \sum_{j=0}^{T-1} \frac{|k(j/S_T)|}{S_T} \left( \left\| \sum_j \sum_t V_t X'_{t-j} \right\| \right) \right)
\]

\[
= | \sup_t X'_t \sqrt{T}(\hat{\beta} - \beta) | \left( \sum_{j=0}^{T-1} \frac{|k(j/S_T)|}{S_T} \left( \left\| \sum_j \sum V_t \right\| \left( \left\| \sum_j \sum X'_s \right\| \right) \right) \right)
\]

\[
= O_p(1)
\]

This is because \( \sqrt{T}(\hat{\beta} - \beta) \), \( \sup_t X_t \), \( \frac{1}{\sqrt{T}} \sum V_t \) and \( \frac{1}{\sqrt{T}} \sum X'_s \) are all \( O_p(1) \) under assumptions 1 and 2. \( \sum_{j=0}^{T-1} \frac{|k(j/S_T)|}{S_T} \to \int_0^\infty |k(s)| ds < \infty \) by assumption 3.

Similarly,

\[
\frac{\sqrt{T}}{S_T} \left\| \sum_{j=0}^{T-1} k(j/S_T) \sum_t \frac{1}{T} \left[ X'_{j}(\hat{\beta} - \beta) \right] X_t V'_{t-j} \right\| = O_p(1).
\]
\[
\frac{\sqrt{T}}{S_T} \left\| \sum_{j=0}^{T-1} k(j/S_T) \sum_t \frac{1}{T} [X_i'(\hat{\beta} - \beta)X_{t-j}(\hat{\beta} - \beta)]X_tX'_{t-j} \right\| \\
\leq \frac{1}{\sqrt{T}} (\sup_t X'_i \sqrt{T}(\hat{\beta} - \beta))^2 \sum_{j=0}^{T-1} \left\| \frac{k(j/S_T)}{S_T} \right\|(\left\| \frac{1}{\sqrt{T}} \sum_t X_t \right\|)(\left\| \frac{1}{\sqrt{T}} \sum_s X'_s \right\|) \\
= \frac{O_p(1)}{\sqrt{T}} = o_p(1)
\]

**Step Two:** \( \tilde{J}_T - J_T \to 0 \)

\[
\tilde{J}_T - J_T \\
= \sum_{j=0}^{T-1} k(j/S_T)[\frac{1}{T} \sum_{t=j+1}^T (V_t V'_{t-j} - E V'_{t-j})] + \sum_{j=-T+1}^{-1} k(j/S_T)[\frac{1}{T} \sum_{t=-T+1}^{j} (V_{t+j} V'_t - E V_{t+j} V'_t)] \\
+ \sum_{j=0}^{T-1} (k(j/S_T) - 1)(\frac{1}{T} \sum_{t=j+1}^T E V'_{t-j}) + \sum_{j=-T+1}^{-1} (k(j/S_T) - 1)(\frac{1}{T} \sum_{t=-j+1}^{T} E V'_{t+j}) \\
S_T^{-1} T^{1-2/r} \left\| \sum_{j=0}^{T-1} k(j/S_T)[\frac{1}{T} \sum_{t=j+1}^T (V_t V'_{t-j} - E V'_{t-j})] \right\|_{r/2} < \infty,
\]

follows exactly from the proof of Theorem 1 in Hansen (1992), due to the fact that \( V_i = \sigma_t \epsilon_t \) where \( \sigma_t \) is uniformly bounded by assumption 1 and \( \epsilon_t \) satisfies all conditions stated in Hansen’s arguments.

\[
\left\| \frac{1}{T} \sum_{t=j+1}^T E V'_{t-j} \right\| = \left\| \frac{1}{T} \sum_{t=j+1}^T \sigma_t E(\epsilon_t \epsilon'_{t-j}) \sigma'_{t-j} \right\| \leq \frac{T-j}{T} (\sup_t \sigma_t)^2 \| \Gamma_\epsilon(j) \|
\]

where \( \Gamma_\epsilon(j) = E(\epsilon_t \epsilon'_{t-j}) \) since \( \epsilon_t \) is second order stationary.

Since \( k(j/S_T) \to 1 \) for each \( j \) as \( T \to \infty \) under assumption 3, and \( \left\| \sum_j \Gamma_\epsilon(j) \right\| < \infty \), by dominated convergence theorem, applied to the counting measure, we have

\[
\sum_{j=0}^{T-1} (k(j/S_T) - 1)(\frac{1}{T} \sum_{t=j+1}^T E V'_{t-j}) \to 0.
\]

**Step Three:** \( J_T \to \int_0^1 \omega(s) \Sigma_\epsilon \omega(s)' ds \)

\[
J_T = \sum_{j=0}^{T-1} (\frac{1}{T} \sum_{t=j+1}^T E V'_{t-j}) + \sum_{j=-T+1}^{-1} (\frac{1}{T} \sum_{t=-j+1}^{T} E V'_{t+j})
\]

29
\[
\frac{1}{T} \sum_{j=0}^{T-1} \sum_{t=j+1}^{T} \sigma_t \Gamma_\epsilon(j) \sigma'_t - \frac{1}{T} \sum_{j=0}^{T-1} \sum_{t=j+1}^{T} \sigma_t \Gamma_\epsilon(j) \sigma'_t
\]

We firstly want to show that

\[
\left\| \frac{1}{T} \sum_{j=0}^{T-1} \sum_{t=j+1}^{T} \sigma_t \Gamma_\epsilon(j) \sigma'_t - \frac{1}{T} \sum_{j=0}^{T-1} \sum_{t=j+1}^{T} \sigma_t \Gamma_\epsilon(j) \sigma'_t \right\| \to 0,
\]

and

\[
\left\| \frac{1}{T} \sum_{j=-T+1}^{-1} \sum_{t=-j+1}^{T} \sigma_{t-j} \Gamma_\epsilon(j) \sigma'_t - \frac{1}{T} \sum_{j=-T+1}^{-1} \sum_{t=-j+1}^{T} \sigma_t \Gamma_\epsilon(j) \sigma'_t \right\| \to 0.
\]

Since

\[
\frac{1}{T} \sum_{j=0}^{T-1} \sum_{t=j+1}^{T} \sigma_t \Gamma_\epsilon(j) \sigma'_t + \frac{1}{T} \sum_{j=-T+1}^{-1} \sum_{t=-j+1}^{T} \sigma_t \Gamma_\epsilon(j) \sigma'_t
\]

\[
= \frac{1}{T} \sum_{t=1}^{T} \sigma_t (\sum_{j=1-t}^{T} \Gamma_\epsilon(j)) \sigma'_t + \frac{1}{T} \sum_{t=1}^{T} \sigma_t (\sum_{j=1-t}^{-1} \Gamma_\epsilon(j)) \sigma'_t
\]

\[
= \frac{1}{T} \sum_{t=1}^{T} \sigma_t (\sum_{j=1-t}^{T} \Gamma_\epsilon(j)) \sigma'_t
\]

We have

\[
\left\| J_T - \frac{1}{T} \sum_{t=1}^{T} \sigma_t (\sum_{j=1-t}^{T} \Gamma_\epsilon(j)) \sigma'_t \right\| \to 0.
\]

For each \( j \geq 0, \)

\[
\left\| \sum_{t=j+1}^{T} \sigma_t \Gamma_\epsilon(j) \sigma'_t - \sum_{t=j+1}^{T} \sigma_t \Gamma_\epsilon(j) \sigma'_t \right\|
\]

\[
\leq \sum_{t=j+1}^{T} \left\| \sigma_t \Gamma_\epsilon(j) (\sigma'_t - \sigma'_t) \right\|
\]

\[
\leq \sup_t \left\| \sigma_t \right\| \left\| \Gamma_\epsilon(j) \right\| \sum_{t=j+1}^{T} \left\| \sigma'_t - \sigma'_t \right\|
\]

\[
= d \left\| \Gamma_\epsilon(j) \right\| \sum_{t=j+1}^{T} \left\| \omega((t - j)/T) - \omega(t/T) \right\|
\]

\[
\leq d \left\| \Gamma_\epsilon(j) \right\| (T - j) j / T
\]

30
The last inequality follows from uniform Lipschitz Continuity. Hence we have, under assumption 3,
\[
\frac{1}{T} \sum_{j=0}^{T-1} \sum_{t=j+1}^{T} \sigma_t \Gamma_\epsilon(j) \sigma'_{t-j} - \frac{1}{T} \sum_{j=0}^{T-1} \sum_{t=j+1}^{T} \sigma_t \Gamma_\epsilon(j) \sigma'_t \leq d \left( \frac{1}{T} \sum_{j=0}^{T-1} j ||\Gamma_\epsilon(j)|| \right) \rightarrow 0.
\]

Exactly the same arguments hold for \( j < 0 \) case.

Finally, we want to show that
\[
\frac{1}{T} \sum_{t=1}^{T} \sigma_t \left( \sum_{|\epsilon| \geq t} \Gamma_\epsilon(j) \right) \sigma'_t \rightarrow 0,
\]
hence we have
\[
||J_T - \int_0^1 \omega(s) \Sigma_\epsilon(s) \omega'(s)ds|| \rightarrow 0,
\]
as \[
\frac{1}{T} \sum_{t=1}^{T} \sigma_t (\sum_{|\epsilon| \geq t} \Gamma_\epsilon(j)) \sigma'_t = \int_0^1 \omega(s) \Sigma_\epsilon(s) \omega'(s)ds.
\]

\[
||\frac{1}{T} \sum_{t=1}^{T} \sigma_t (\sum_{|\epsilon| \geq t} \Gamma_\epsilon(j)) \sigma'_t || \leq \frac{1}{T} (\sup_{t} ||\sigma_t||)^2 \left( \sum_{t=1}^{T} \sum_{|\epsilon| \geq t} ||\Gamma_\epsilon(j)|| \right)
\]

\[
= d^2 \frac{1}{T} \sum_{j=-\infty}^{\infty} j ||\Gamma_\epsilon(j)|| \rightarrow 0.
\]
References


[37] Teverovsky V. and Taqqu M. (1997): “Testing for long-range dependence in the presence of shifting means or a slowly declining trend,