The Nonlinear Iterative Least Squares (NL-ILS) Estimator: An Application to Volatility Models

Gustavo Fruet Dias*
Queen Mary University of London

This version: November 21, 2012

Job Market Paper

Abstract

The paper proposes a new robust estimator for GARCH-type models: the nonlinear iterative least squares (NL-ILS). This estimator is especially useful on specifications where errors have some degree of dependence over time or when the conditional variance is misspecified. I illustrate the NL-ILS estimator by providing algorithms that consider the GARCH(1,1), weak-GARCH(1,1), GARCH(1,1)-in-mean and RealGARCH(1,1)-in-mean models. I establish the consistency and asymptotic distribution of the NL-ILS estimator, in the case of the GARCH(1,1) model under assumptions that are compatible with the QMLE estimator. The consistency result is extended to the weak-GARCH(1,1) model and a further extension of the asymptotic results to the GARCH(1,1)-in-mean case is also discussed. A Monte Carlo study provides evidences that the NL-ILS estimator is consistent and outperforms the MLE benchmark in a variety of specifications. Moreover, when the conditional variance is misspecified, the MLE estimator delivers biased estimates of the parameters in the mean equation, whereas the NL-ILS estimator does not. The empirical application investigates the risk premium on the CRSP, S&P500 and S&P100 indices. I document the risk premium parameter to be significant only for the CRSP index when using the robust NL-ILS estimator. The paper argues that this comes from the wider composition of the CRPS index, resembling the market more accurately, when compared to the S&P500 and S&P100 indices. This finding holds on daily, weekly and monthly frequencies and it is corroborated by a series of robustness checks.

JEL classification numbers: C13, C15, C18, C22, G12

Keywords: Risk premium; GARCH-type models; Iterative estimators; Contraction mapping; Realized volatility

Acknowledgments: I am especially grateful to George Kapetanios for his comments and suggestions in this project. I am also indebted to Marcelo Fernandes, Emmanuel Guerre, Ana Galvão, Tim Bollerslev, Andrew Patton and Liudas Giraitis for valuable comments as well as the seminar participants at Queen Mary (Econometrics Reading Group) and Duke Financial Econometrics Lunch Group. The latest version of this paper can be downloaded from http://webspace.qmul.ac.uk/gfruedias. The usual disclaimer applies.

*School of Economics and Finance, Queen Mary University of London, London, E1 4NS, UK. E-mail: g.fruetdias@qmul.ac.uk. Tel: +44 (0)20 7882 8846. Fax: +44 (0)20 8983 3580
1 Introduction

Measuring volatility and identifying its sources is of major importance in finance and economics. Investors are concerned about asset return volatility because it plays crucial role on asset pricing, risk management and portfolio allocation. As a result, the task of modeling the conditional variance has been a central topic in econometrics following the seminal papers of Engle (1982) and Bollerslev (1986). Since then, different specifications and frameworks, such as GARCH-type models, stochastic volatility, realized volatility and combinations of these approaches have been adopted, trying to capture the very specific stylized facts observed in financial returns. A natural extension that emerges from modeling the conditional variance is the relation between risk and return. The intertemporal capital asset pricing model (ICAPM) of Merton (1973) establishes a positive relation between the conditional excess returns and the conditional variance, implying that investors should be remunerated for bearing extra risk. Engle, Lilien, and Robins (1987) provide the first econometric specification that relates the conditional second moment to the first moment, allowing to test the ICAPM model. Following them, several attempts have been undertaken to estimate the risk premium parameter, however empirical evidences on the sign and significance of this parameter are blurred. Two potential causes are: firstly, as Bollerslev, Chou, and Kroner (1992) point out, quasi-maximum likelihood (QMLE) estimates of the risk premium parameter using the GARCH-in-mean framework may be inconsistent if the conditional variance is misspecified. Secondly, as Drost and Nijman (1993) discuss, sampling frequency impacts the validity of the assumptions governing the QMLE estimator and as a consequence of time aggregation, estimates of the risk premium parameter may be inconsistent

In this paper, I address the two above-mentioned issues, by proposing a novel fully parametric iterative estimator, the non-linear iterative least squares estimator (NL-ILS). The NL-ILS estimator nests the GARCH(1,1), weak-GARCH(1,1), GARCH(1,1)-in-mean and RealGARCH(1,1)-in-mean models. I derive the consistency and asymptotic distribution for the GARCH(1,1) model under mild assumptions. The asymptotic results for the NL-ILS estimator do not depend on the correct specification of the stochastic term distribution, allowing, therefore, the NL-ILS estimator to compete against the QMLE estimator. Moreover, I extend the consistency result to the weak-GARCH(1,1) case, which as far as my knowledge goes, is only covered by the estimator proposed by Francq and Zakoian (2000). Furthermore, I show through Monte Carlo exercises that the NL-ILS estimator is more robust to misspecification of the conditional variance than the QMLE

1Linton and Perron (2003), Linton and Sancetta (2009), Conrad and Mammen (2008), Christensen, Dahl, and Iglesias (2012) point a third issue. They find strong evidences that the relation between risk and return is non-linear, indicating that the mixed results obtained with the full parametric GARCH-in-mean models could be the results of misspecification of the mean equation.

2Under certain assumptions briefly discussed in Section 2.1, the NL-ILS is also valid for ARMA(1,1) and weak-ARMA(1,1) models.
estimator when considering the GARCH(1,1)-in-mean case. This result is particularly important when investigating the existence of the risk-return tradeoff, since the true data generation process (DGP) of the conditional variance is unknown in practise. I find evidences that bias on the QMLE estimates of the risk premium parameter leads to false significant risk premium estimates in a full parametric GARCH(1,1)-in-mean model.

The literature on GARCH-type models is extremely extensive, with a wide range of specifications aiming to capture different stylized facts (see Francq and Zakoian (2010) and Bollerslev (2008)). In this paper, I will focus on the following models: GARCH(1,1), weak-GARCH(1,1), GARCH(1,1)-in-mean and RealGARCH(1,1)-in-mean originally proposed by Bollerslev (1986), Drost and Nijman (1993), Engle, Lilien, and Robins (1987) and Hansen, Huang, and Shek (2011), respectively. Another important branch of the GARCH literature examines the asymptotic properties of the QMLE estimator. Research on this topic has mainly focused on relaxing moment assumptions as a way to accommodate heavy-tailed marginal distributions (see Francq and Zakoian (2008) for a survey on this topic). I address this issue by establishing the asymptotic theory for the GARCH(1,1) model under assumptions that are compatible with the QMLE estimator. Apart from Christensen, Dahl, and Iglesias (2012) of which work nests the full parametric GARCH(1,1)-in-mean and which is based on the profile log-likelihood approach, there has not been so far a proper QMLE asymptotic theory covering this model. This paper discusses the extension of the NL-ILS asymptotic results for the GARCH(1,1)-in-mean and the RealGARCH(1,1)-in-mean models.

Recently, the abundant availability of high frequency data has triggered a new class of volatility models: the realized volatility (see Mcaleer and Medeiros (2008) for an extensive survey on the different estimators available in the literature). Jointly with that, models that combine GARCH-type structure with realized measures, such as GARCH-X in Engle (2002), HEAVY in Shepard and Sheppard (2010) and RealGARCH in Hansen, Huang, and Shek (2011), have also become popular. These “turbo”\(^3\) models have the nice property of adjusting much faster to shocks in volatility, providing a better forecasting performance than GARCH-type models. By extending the NL-ILS algorithm to the RealGARCH(1,1)-in-mean model, I am able to assess whether, by augmenting the volatility equation with realized variance measures, the risk premium parameter estimate improves. Moreover, the theoretical framework I use to establish the asymptotic theory for the GARCH(1,1) model can also be extended to accommodate exogenous regressors in the variance equation as in the RealGARCH(1,1)-in-mean model. Another important advantage of the NL-ILS framework emerges from its robustness to disturbances that possess some nonlinear dependence. From the RealGARCH framework, the measurement equation relates the conditional variance to the realized variance. Hansen, Huang, and

\(^3\)This is an expression used by Shepard and Sheppard (2010), and it illustrates, in a very good way, the enhanced properties of this class of augmented models.
Shek (2011) assume the stochastic term in the measurement equation is an independent and identically distributed (iid) process evolving on daily basis. I argue the conditional and realized variance evolve at different frequencies. The former one evolves on a daily basis, whereas the latter evolves intraday. Following that, modeling the stochastic term in the measurement equation as an iid process might turn out to be a far too strong assumption. Hence, it makes necessary the adoption of estimators that can cope with disturbances possessing dependence on higher moments, such as linear projections, as discussed in Drost and Nijman (1993) and Drost and Werker (1996).

In the empirical section I investigate the existence of the risk premium in the spirit of the ICAPM model proposed by Merton (1973). To do so, I adopt the GARCH(1,1)-in-mean specification. The main question is whether the risk premium parameter is significant and presents the correct sign by using an estimator which is robust to misspecification of the conditional variance and also to dependence on the errors. I assess this question in two dimensions: temporal frequency and market proxy. To evaluate the former one I estimate the model on daily, weekly and monthly basis. To appraise the latter dimension I adopt three market indices: CRSP, S&P500 and S&P100. The choice of comparing different indices emerges from the different compositions they have. The CRSP data set is known to be the best proxy for the market. When I implement the NL-ILS, the risk premium is significant only in the CRSP data set. A different picture arises when I use the QMLE estimator: the risk premium is significant in all frequencies and indices. This result holds across all three frequencies. Following the consistency issue of the QMLE estimator, I perform robustness checks using RealGARCH-in-mean (with both NL-ILS and QMLE estimators), EGARCH-in-mean, GJR-GARCH-in-mean and APARCH-in-mean models. These exercises deliver results for the risk premium estimates which are in line with the ones found when using the robust NL-ILS estimator. I argue that the NL-ILS estimator is able to capture the “true” risk premium, since its results reflect the wider composition of the CRPS index, resembling the market more accurately, when compared to S&P500 and S&P100 indices.

This paper is organized as follows. Section 2 introduces the NL-ILS estimator and establishes the asymptotic theory for the GARCH(1,1) case. I start the discussion with a generic model nesting the GARCH, GARCH-in-mean and RealGARCH models. I then illustrate the specific cases of GARCH(1,1), weak-GARCH(1,1) and GARCH(1,1)-in-mean. Section 3 presents the NL-ILS algorithm for the GARCH(1,1), weak-GARCH(1,1), GARCH(1,1)-in-mean and RealGARCH(1,1)-in-mean. Section 4 displays an extensive Monte Carlo study, assessing the finite sample performance of the NL-ILS compared to the QMLE benchmark with respect to consistency, efficiency and forecast accuracy. This section also discusses the robustness of the NL-ILS estimator when the conditional variance is misspecified. In Section 5, I assess the risk-return tradeoff considering different indices at three sampling frequencies. Section 6 concludes. The Appendix contains all
2 Asymptotic theory: main results

This section provides theoretical results regarding the consistency and asymptotic distribution of the NL-ILS estimator. I start with a generic model nesting three of the models I discuss throughout this paper: GARCH(1,1), GARCH(1,1)-in-mean and RealGARCH(1,1)-in-mean models. Firstly, I derive the consistency and asymptotic distribution for this generic model under high level assumptions. Secondly, I relax some of these assumptions focusing on these models and analyzing them in greater depth in the following sections. The theory developed in this section is based on the work of Dominitz and Sherman (2005). Following their work, the crucial point on showing consistency and asymptotic distribution for this class of iterative estimators lies on proving that the population mapping is an Asymptotic Contraction Mapping (ACM)\(^4\). If the population mapping is an ACM, then it has a fixed point. This allows the use of the fixed point theorem to derive the consistency of the iterative estimator. The asymptotic theory is derived using the population mapping evaluated at the true vector of parameters, allowing the use of asymptotic results obtained from the standard non-linear least squares (NL-LS) framework.

Assume a stationary stochastic process \(\{y_t\}_{t=1}^{T}\) with finite fourth moment.

\[
y_t = f(Y_{t-l}, X_{t-m}, \sigma_t, \theta_1) + \epsilon_t, \quad l \geq 1, \quad m \geq 0 \tag{1}
\]

\[
\epsilon_t = \sigma_t \eta_t \tag{2}
\]

\[
\sigma_t^2 = \sum_{i=1}^{p} \alpha_i \epsilon_{t-i} + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2 + \sum_{i=1}^{r} \gamma_i v_{t-i} \tag{3}
\]

\[
Z_t = \Psi_0 + U_t + \sum_{i=1}^{\infty} \Psi_i U_{t-i} \quad \Psi_i = \varrho_i(\theta_2), \quad i = 0, 1, ..., \infty \tag{4}
\]

where \(f(Y_{t-l}, X_{t-m}, \sigma_t, \theta_1)\) is a twice continuously differentiable function; \(\varrho_i(\theta_2)\) is a continuous function for all \(i\)'s; \(X_{t-m}\) is a matrix of exogenous regressors; \(Y_{t-l}\) is a vector containing lags of the dependent variable; \(\sigma_t^2\) is a latent variable (conditional variance); \(Z_t = (\epsilon_t^2, v_t)\); \(U_t\) is a vector of martingale difference sequence (m.d.s.) processes, such that \(\mathbb{E}(U_t) = 0\) and \(\text{Var}(U_t) = \Sigma_u\) with \(\Sigma_u\) being a diagonal matrix; \(\theta_1\) is a vector of free parameters in (1), \(\theta_2\) is a vector of free parameters in (3) and \(\theta = (\theta_1, \theta_2)'\). Denote \(\mathbb{B}\) as the space where \(\theta\) is defined. Equation (1) is generic enough to accommodate models that are nonlinear in the parameters, also nesting linear regressions. As in Dominitz and Sher-

\(^4\)Using the definition in Dominitz and Sherman (2005), a collection \(\{K^\omega_T(.): T \geq 1, \omega \in \Omega\}\) is an ACM on \((\mathbb{B}, d)\) if \(d(K^\omega_T(x), K^\omega_T(y)) \leq cd(x, y)\) as \(T \to \infty\), where \(c \in [0, 1)\), \((\mathbb{B}, d)\) is a metric space with \(x, y \in \mathbb{B}\), \((\Omega, \mathcal{A}, \mathbb{P})\) denoting a probability space and \(K^\omega_T(.)\) is a function defined on \(\mathbb{B}\).
Note that $\Psi_j\theta_j$ denotes the iteration which parameters are computed. The number of iterations $\theta T$ to be a function of $N$ implies that, if $tive function as $Q_T(y_t, v_i; \theta)$ and its population counterpart as $\mathbb{E}(Q_T(y_t, v_i; \theta))$. These functions are non-linear in the parameters, yielding NL-LS estimates on all iterations. Therefore, the NL-ILS estimator consists on computing NL-LS estimates using the estimates of the latent variables as regressors, updating, at each iteration, the latent variable using the NL-LS parameter estimates. This procedure is repeated until the parameters converge.

**Definition 1 Mapping:**

Define the population mapping as $N(\theta_j)$ and its sample counterpart as $\hat{N}_T(\hat{\theta}_j)$, such that at any $j$ iteration, $N(\theta_j)$ maps from $\theta_j$ to $\theta_{j+1}$ and $\hat{N}_T(\hat{\theta}_j)$ maps from $\hat{\theta}_j$ to $\hat{\theta}_{j+1}$.

$$
\theta_{j+1} = N(\theta_j) = \min_{\theta_{j+1}} \mathbb{E} \left\{ \frac{1}{T} \sum_{t=1}^{T} \left[ (y_t - f(Y_{-t}, X_{t-m}, \sigma_{j,t}, \theta_{j+1}))^2 - Q_{j+1,0} - \sum_{i=0}^{\infty} \Psi_{j+1,i} U_{j,t-1-i} \right]^2 \right\} \tag{5}
$$

$$
\hat{\theta}_{j+1} = \hat{N}_T(\hat{\theta}_j) = \min_{\hat{\theta}_{j+1}} \frac{1}{T} \sum_{t=1}^{T} \left[ (y_t - f(Y_{-t}, X_{t-m}, \hat{\sigma}_{j,t}, \hat{\theta}_{j,t+1}))^2 - \hat{Q}_{j+1,0} - \sum_{i=0}^{\bar{q}} \hat{\Psi}_{j+1,i} \hat{U}_{j,t-1-i} \right]^2 \tag{6}
$$

Note that $\Psi_{j+1,i}$ and $\hat{\Psi}_{j+1,i}$ depend on $\theta_{2,j+1}$ and $\hat{\theta}_{2,j+1}$, respectively. The subscript $j$ denotes the iteration which parameters are computed. The number of iterations $j$ is set to be a function of $T$, such that as $T \to \infty$, $j \to \infty$ at some rate satisfying $\ln(T)/j = o(1)$ and $\bar{q}$ is a truncation parameter, such that $\bar{q} \to \infty$ at a logarithmic rate of $T$. From the population mapping definition, $\theta = N(\theta)$ holds as an identification condition. This implies that, if $N(\theta_j)$ is an ACM, then $\theta$ is the fixed point of the population mapping and the following bound holds for any $j$:

$$
|\theta_{j+1} - \theta_j| = |N(\theta_j) - N(\theta_{j-1})| \leq \kappa |\theta_j - \theta_{j-1}| \tag{7}
$$

where $\kappa = [0, 1)$ is the contraction parameter. By using the Newton-Raphson (NR) procedure, the two mappings in Definition 1 have the following linear representation:

$$
\theta_{j+1} = N(\theta_j) = \theta_j - [H(\theta_j)]^{-1} G(\theta_j) \tag{8}
$$

$$
\hat{\theta}_{j+1} = \hat{N}_T(\hat{\theta}_j) = \hat{\theta}_j - \left[ \hat{H}_T(\hat{\theta}_j) \right]^{-1} \hat{G}_T(\hat{\theta}_j) \tag{9}
$$

where $\hat{G}_T(\hat{\theta}_j)$ and $\hat{H}_T(\hat{\theta}_j)$ are the sample gradient and Hessian computed from $Q_T(\hat{\theta}_j)$ whereas $H(\theta_j)$ and $G(\theta_j)$ are their population counterparts. To use the theory developed by Dominitz and Sherman (2005), I introduce assumptions which are related to the identification of classical nonlinear regression models (see Amemiya (1985) pg. 129 for more details), and assumptions governing the behavior of both population and sample
mappings.  

**Assumptions A**

1.  
\[
\mathbb{E}\left\{ \left[ f(Y_{-t}, X_{t-m}, \sigma_t, \theta) - f(Y_{-t}, X_{t-m}, \sigma_t, \tilde{\theta}) \right]^2 + \left[ \left( \varphi_0(\theta_2) + \sum_{i=1}^{\infty} \varphi_i(\theta_2) \right) \varphi_0(\tilde{\theta}_2) + \sum_{i=1}^{\infty} \varphi_i(\tilde{\theta}_2) \right] \right\} \neq 0, \quad \forall \tilde{\theta} \neq \theta
\]

2. \( \text{Cov}(f(Y_{-t}, X_{t-m}, \sigma_t, \theta) , \epsilon_t) = 0. \)

3. The disturbances \( \eta_t \) have a non-degenerate distribution such that \( \eta_t \sim \text{iid}(0, 1) \) and \( \mathbb{E}(\eta_t^4) < \infty. \)

4. \( N(\theta_j) \) is an ACM in spirit of the definition of Dominitz and Sherman (2005) for all \( \theta_j \in \mathbb{B}. \)

5. \( \sup_{\xi, \varsigma \in \mathbb{B}} \left| N(\xi) - N_T(\varsigma) \right| = o_p(1) \) for all \( \xi, \varsigma \in \mathbb{B}. \theta_j \in \mathbb{B}. \)

Assumption A1 implies the population mapping is identified, allowing the use of the NL-LS estimator to recover estimates of \( \theta. \) Note that Assumption A2 is weaker than the usual assumption presented in linear regressions with stochastic regressors. In these cases, the regressors at time \( t \) are assumed to be independent of \( \epsilon_s \) for all \( t \) and \( s \) as discussed in Hamilton (1994) chapter 8. Our setup, however, relies on relaxing this assumption in spirit of the AR(\( p \)) model (case 4 in chapter 8 of Hamilton (1994)). Assumption A4 states that population mapping in Definition 1 is an ACM. Assumptions A5 establishes uniform convergence between the sample and population mappings. Under Assumptions A1, A2, A3, A4 and A5, Theorems 2 and 4 in Dominitz and Sherman (2005) hold, yielding the consistency and asymptotic distribution of the NL-ILS algorithm for the generic model defined in (1), (3) and (4).

### 2.1 GARCH(1,1)

From the seminal papers of Engle (1982) and Bollerslev (1986), a process \( \{y_t\}_{t=1}^{\infty} \) is said to be a strong GARCH(1,1), (GARCH(1,1) hereafter), if the following structure holds:

\[
y_t = \epsilon_t = \eta_t \sigma_t \\
\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2
\]

where \( \eta_t \) is an \( \text{iid} \sim (0, 1) \) and \( \sigma_t^2 \) is the latent conditional variance. Sufficient conditions on the parameters of (11) that guarantee the process in (10) is second-order stationary are: \( \omega > 0, \alpha > 0, \beta > 0 \) and \( \alpha + \beta < 1 \) (see Francq and Zakoian (2010) for an extensive study
on stationarity solutions to GARCH(p,q) models. The conditional variance equation of
the GARCH(1,1) model allows an ARMA(1,1) representation on the form of:

\[ \epsilon_t^2 = \omega + a \epsilon_{t-1}^2 + u_t + bu_{t-1} \]  

(12)

where \( a = (\alpha + \beta) \) and \( b = -\beta \) are the autoregressive and moving average parameters, respectively. Denote \( \phi \) as \( \phi = (\omega, a, b)' \). The disturbances \( u_t = \epsilon_t^2 - \sigma_t^2 \) are m.d.s., such
that \( \mathbb{E}(u_t) = 0 \) and \( \text{Var}(u_t) = \sigma_u^2 \). If the GARCH(1,1) in (10) and (11) is covariance
stationary, then the ARMA(1,1) in (12) can be expressed as \( \text{MA}(\infty) \) as:

\[ \epsilon_t^2 = \psi_0 + \sum_{i=1}^{\infty} \psi_i u_{t-i} + u_t \]  

(13)

where \( \psi_0 = \frac{\omega}{1-a}, \psi_i = a^i(a+b) \).

I establish the consistency and asymptotic distribution for the GARCH(1,1) model,
by relaxing some of the high level assumptions I imposed to the generic model. Note that
the generic model nests the GARCH(1,1) model, by setting \( f(Y_{t-l}, X_{t-m}, \sigma_t, \theta_1) = 0, \) lags orders \( p \) and \( q \) to 1 and \( \gamma_i = 0 \) for all \( i = 1, 2, ..., r \). These imply that the VMA
in (4) reduces to the \( \text{MA}(\infty) \) depicted in (13). Using Definition 1, and setting the
sample objective function as

\[ Q_T(y_t; \hat{\phi}_{j+1}) = \frac{1}{T} \sum_{t=1}^{T} \left[ \epsilon_t^2 - \hat{\psi}_{j+1,0} - \sum_{i=0}^{q} \hat{\psi}_{j+1,i} \hat{u}_{t-i} \right]^2, \]

the population and the sample mappings are defined as:

**Definition 2 GARCH(1,1) Mapping:**

Define the population mapping for the GARCH(1,1) model as \( N(\theta_j) \) and its sample coun-
terpart as \( \hat{N}_T(\hat{\theta}_j) \), such that at any \( j \) iteration, \( N(\theta_j) \) maps from \( \theta_j \) to \( \theta_{j+1} \) and \( \hat{N}_T(\hat{\theta}_j) \) maps from \( \hat{\theta}_j \) to \( \hat{\theta}_{j+1} \).

\[ \phi_{j+1} = N(\phi_j) = \min_{\phi_{j+1}} \mathbb{E} \left\{ \frac{1}{T} \sum_{t=1}^{T} \left[ \epsilon_t^2 - \psi_{j+1,0} - \sum_{i=0}^{\infty} \psi_{j+1,i} u_{t-i} \right]^2 \right\} \]  

(14)

\[ \hat{\phi}_{j+1} = \hat{N}_T(\hat{\phi}_j) = \min_{\phi_{j+1}} \frac{1}{T} \sum_{t=1}^{T} \left[ \epsilon_t^2 - \hat{\psi}_{j+1,0} - \sum_{i=0}^{q} \hat{\psi}_{j+1,i} \hat{u}_{t-i} \right]^2 \]  

(15)

Remark Definition 2: the \( \text{MA}(\infty) \) representation satisfies assumption A1 in the generic
setup (see Lemma 2 in the appendix). I relax assumption A4 and A5 in order to establish
the consistency and asymptotic distribution of the NL-ILS estimator. To this purpose, I
state the following assumptions:
Assumptions B

1. The GARCH(1,1) model stated in (10) and (11) is second-order stationary and yields $\sigma_t^2 > 0$ for all $t$. These imply that $\omega > 0$, $\alpha > 0$, $\beta > 0$ and $\alpha + \beta < 1$. Also, assume that $\phi \in \mathbb{B}$ and $\mathbb{B}$ is compact.

2. The disturbances $\eta_t$ have a non-degenerate distribution such that $\eta_t \sim iid (0, 1)$ and $\mathbb{E}(\eta_t^4) < \infty$.

3. Define the gradient of $N(\phi_j)$ as $V(\phi_j) = \nabla_{\phi_j} N(\theta_j)$ and its sample counterpart as $\hat{V}_T(\hat{\phi}_j) = \nabla_{\hat{\phi}_j} \hat{N}_T(\hat{\phi}_j)$. Then, the Euclidean norm of $\hat{V}_T(\hat{\phi}_j)$ is bounded in probability, such that $\|\hat{V}_T(\hat{\phi}_j)\| = O_p(1)$ for all $\hat{\phi}_j \in \mathbb{B}$.

Assumption B2 is important in two different aspects: firstly, it implies that $u_t$ in (24) is a m.d.s.; secondly, the finite fourth moment is required as a condition to obtain a finite $\sigma_u^2$.

Finally, Assumption B3 establishes the Lipschitz condition of $\hat{N}_T(\hat{\phi}_j)$. This implies that $\hat{N}_T(\hat{\phi}_j)$ is stochastically equicontinuous. To show consistency of the NL-ILS estimator, the crucial point consists on proving that the population mapping is an ACM. Lemma 1 delivers this result.

**Lemma 1** Suppose Assumptions B1, B2 and B3 hold. Then, there exist an open ball centered at $\phi$ with closure $\mathbb{B}$, such that the mapping $N(\phi_j)$ is an ACM on $(\mathbb{B}, d)$, with $\phi_j \in \mathbb{B}$ for all $j > 0$.

Figure 1 displays the maximum eigenvalue associated with different combinations of parameters satisfying Assumption B1. From Lemma 1, the population mapping has a fixed point such that $N(\phi) = \phi$ holds and the following inequality is valid for all iterations:

$$|\phi_{j+1} - \phi_j| = |N(\phi_j) - N(\phi_{j-1})| \leq \kappa |\phi_j - \phi_{j-1}|$$  \hspace{1cm} (16)

Remark Lemma 1: Lemma 1 can also be extended to the ARMA(1,1) case. Note that the eigenvalues of the population mapping gradient evaluated on the true vector of parameters, $V(\phi)$, are given by:

$$\varepsilon = \left[ \frac{a + b}{1 + b}, \frac{a(a + b)}{1 + b}, \frac{a(a + b)}{1 + b} \right]'$$  \hspace{1cm} (17)

Under the ARMA(1,1) model, Assumption B1 is relaxed such that $|a| < 1$ and $|b| < 1$ hold. For all $b > 0$, the eigenvalues associated with (17) are smaller than one in absolute value. This result is particularly relevant for ARMA(1,1) models generated with a positive moving average parameter (close to unity) and a negative autoregressive parameter.
(potentially close to one in absolute value). Under such combination of parameter values, Dias and Kapetanios (2012) show that the IOLS estimator is not valid, because its population mapping is not an ACM. This implies that the NL-ILS estimator can, alternatively, be used when convergence is not achieved with the IOLS estimator.

To prove the consistency of the NL-ILS estimator, it is necessary to show that the population mapping and the population gradient converge uniformly to their sample counterparts when evaluated at the same points. These are given by Lemmas 3 and 4, respectively. Lemma 3 is obtained using the fact that $\bar{q} \to \infty$ at a logarithmic rate of $T$ and using the weak law of large numbers. Lemma 5 in the appendix shows that the sample mapping is also an ACM, implying that it has also a fixed point, denoted by $\hat{\phi}$, such that $\hat{N}_T(\hat{\phi}) = \hat{\phi}$.

With regard to the asymptotic distribution of the NL-ILS estimator, Lemma 6 gives the $\sqrt{T}$ convergence of $\hat{\phi}_j$ to $\hat{\phi}$. This is achieved by allowing the number of iterations goes to infinite as $T \to \infty$, such that $\ln(T) = o(1)$. As in Dias and Kapetanios (2012), I use the fact that, when evaluated at the true vector of parameters and $T \to \infty$, the lagged disturbances are no longer latent variables. This implies that asymptotic results from the NL-LS estimator can be used in the final bit of the proof.

Define the following quantities: $A = [I - V(\phi)]^{-1}$; $V(\phi)$ is the gradient of the population mapping evaluated on the true vector of parameters $\phi$; $C_0 = \text{plim}_{T} \sum_{t=1}^{T} \left[ \frac{\partial h_t(\phi)}{\partial \phi} \frac{\partial h_t(\phi)}{\partial \phi'} \right]$; and $h_t(\phi) = \psi_0 - \sum_{i=1}^{T} \psi u_{t-i}$.

$$A^{-1} = \begin{pmatrix} \frac{1-a}{b+1} & \frac{(a^2+a-1)b+1}{b+1} & -\frac{(a-1)(a+1)^2w}{b+1} \\ 0 & \frac{(a^3-2a+1)b}{b+1} & -\frac{(a^2-1)^2}{b+1} \\ 0 & \frac{(ab+1)^2}{b+1} & -(ab+2)a^2+b+2 \end{pmatrix}$$ (18)

$$C_0 = \begin{pmatrix} \frac{1}{(1-a)^3} & -\frac{\omega}{(1-a)^3} & 0 \\ -\frac{\omega}{(1-a)^3} & \frac{\omega^2}{(1-a)^3} + \sum_{i=0}^{\bar{q}} d_i^2 \sigma_u^2 & \sum_{i=0}^{\bar{q}} d_i a^i \sigma_u^2 \\ 0 & \sum_{i=0}^{\bar{q}} d_i a^i \sigma_u^2 & \sum_{i=0}^{\bar{q}} a^{2i} \sigma_u^2 \end{pmatrix}$$ (19)

The consistency and asymptotic distribution of the NL-ILS is therefore given by:
Theorem 1 Suppose Assumptions B1, B2 and B3 hold. Then

i. \[ \left| \hat{\phi} - \phi \right| = o_p(1) \text{ as } j \to \infty \text{ with } T \to \infty. \]

\[ \sqrt{T} \left( \phi - \hat{\phi} \right) \to^d \mathcal{N} \left( 0, \sigma^2_u \Sigma^{-1} \right) \text{ as } T \to \infty \text{ and } \frac{\ln(T)}{j} = o(1). \]

The proof of Theorem 1 is given in the Appendix. The asymptotic covariance matrix can be computed replacing the true vector of parameters with the consistent NL-ILS estimates of \( \phi \). From item (ii) in Theorem 1, it is clear that the asymptotic variance of the NL-ILS is in fact an augmented version of the asymptotic variance obtained from the NL-LS estimator. The closed form for the asymptotic variance of the NL-ILS estimator considering the parameters of GARCH(1,1) in its original form can be easily obtained.

Corollary 1 Suppose Assumptions B1 and B3 hold. If

i. \( \epsilon_t \) is a fourth-order stationary white noise process, such that \( u_t \) in (12) is a linear innovation with \( u_t \sim (0, \sigma^2_u) \) and \( \text{Cov}(u_t, \epsilon_{t-l}^2) \forall l > 0 \); Then, \( \left| \hat{\phi} - \phi \right| = o_p(1) \text{ as } j \to \infty \text{ with } T \to \infty. \)

2.2 GARCH(1,1)-in-mean

To explore the relation between risk and return, Engle, Lilien, and Robins (1987) proposed the ARCH-in-mean model. Following them, a process \( \{y_t\}_{t=1}^T \) is said to be a GARCH(1,1)-in-mean model if the structure below holds:

\[ y_t = \lambda \sigma_t + \epsilon_t \]

\[ \epsilon_t = \eta \sigma_t \]

\[ \sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \]
where $\eta_t$ be an iid $\sim (0, 1)$ sequence and $\sigma_t^2$ is the latent conditional variance. The parameter $\lambda$ is usually known as the risk premium parameter. The GARCH(1,1)-in-mean is second-order stationary provided that $\omega > 0$, $\alpha > 0$, $\beta > 0$ and $\alpha + \beta < 1$ hold. Similarly to the GARCH(1,1) model, (38) allows an ARMA(1,1) representation as:

$$
\epsilon_t^2 = \omega + a \epsilon_{t-1}^2 + u_t + bu_{t-1}
$$

(24)

where $a = (\alpha + \beta)$ and $b = -\beta$. I denote $\phi = (\omega, a, b)'$ and $\theta = (\lambda, \phi')'$. The generic model in (1), (3) and (4) nests the GARCH(1,1)-in-mean specification by setting $f(Y_t, X_t, \sigma_t, \theta_1) = \lambda \sigma_t^2$ and $\gamma_i = 0$ for all $i = 1, 2, \ldots, r$. Extension of Theorem 1 to the GARCH(1,1)-in-mean model does not carry any significant difference. The main point consists on showing that the gradient associated with the population mapping is an ACM. This paper provides numerical evidences indicating that the gradient of the population mapping is indeed an ACM.

**Definition 3** GARCH(1,1)-in-mean Mapping:

Define the population mapping for the GARCH(1,1)-in-mean model as $N(\theta_j)$ and its sample counterpart as $\hat{N}_T(\hat{\theta}_j)$, such that at any $j$ iteration, $N(\theta_j)$ maps from $\theta_j$ to $\theta_{j+1}$ and $\hat{N}_T(\hat{\theta}_j)$ maps from $\hat{\theta}_j$ to $\hat{\theta}_{j+1}$.

$$
\theta_{j+1} = N(\theta_j) = \min_{\phi_{j+1}} \mathbb{E} \left\{ \frac{1}{T} \sum_{t=1}^{T} \left[ y_t - \lambda_{j+1} \sigma_t, \psi_{j+1,0} - \sum_{i=0}^{\infty} \psi_{j+1,i} u_{j,t-1-i} \right]^2 \right\} \tag{25}
$$

$$
\hat{\theta}_{j+1} = \hat{N}_T(\hat{\theta}_j) = \min_{\phi_{j+1}} \frac{1}{T} \sum_{t=1}^{T} \left[ y_t - \hat{\lambda}_{j+1} \hat{\sigma}_t, \hat{\psi}_{j+1,0} - \sum_{i=0}^{\hat{q}} \hat{\psi}_{j+1,i} \hat{u}_{j,t-1-i} \right]^2 \tag{26}
$$

It is possible to split the sample mapping in two distinct procedures: the first mapping delivers estimates of $\lambda$, whereas the second one delivers the parameters from the ARMA(1,1) representation in (24). This result is formalized in Proposition 1.

**Proposition 1** Assume the model stated in (21), (22) and (23). Define the vector of free parameters in (24) on the $j + 1$ iteration as $\hat{\phi}_{j+1} = (\hat{\omega}_{j+1}, \hat{\phi}_{j+1,1}, \hat{\phi}_{j+1,2})'$. The sample mapping in (26) can be computed in two distinct procedures, such that:

1. $\hat{\lambda}_{j+1} = \left[ \sum_{t=1}^{T} \hat{\sigma}_{j,t}^2 \right]^{-1} \sum_{t=1}^{T} \hat{\sigma}_{j,t} y_t$

2. $\hat{\phi}_{j+1} = \min_{\phi_{j+1}} \frac{1}{T} \sum_{t=1}^{T} \left[ y_t - \hat{\lambda}_{j+1} \hat{\sigma}_t, \hat{\psi}_{j+1,0} - \sum_{i=0}^{\hat{q}} \hat{\psi}_{j+1,i} \hat{u}_{j,t-1-i} \right]^2$
Remark Proposition 1: it provides the necessary identification conditions for the use of NL-ILS estimator and alleviates the computational burden of computing parameter(s) in the mean equation. The proof of Proposition 1 is obtained from the first order condition computed from the sample mapping in Definition 3. Figure 2 displays the maximum eigenvalue associated with the numerical gradient of the sample mapping computed using results in Proposition 1. As discussed in Lemma 1, if the maximum eigenvalue is smaller than one in absolute value, this guarantees the ACM property. All the eigenvalues in Figure 2 are less than one in absolute value, indicating that the sample mapping is an ACM. Furthermore, preliminary calculations show that the contraction property of $N(\theta)$ does not depend on $\lambda$, being only governed by the parameters from the ARMA(1,1) representation of (23). If the ACM holds, asymptotic theory for the GARCH(1,1)-in-mean model can be extended following the steps in Theorem 1.

3 NL-ILS estimation procedure

I first describe the NL-ILS algorithm for the simple GARCH(1,1) model. As a natural extension of this procedure, I show that NL-ILS estimator can be also applied to the weak-GARCH(1,1) models. This variant of the GARCH(1,1) model was originally proposed by Drost and Nijman (1993) and it is robust to temporal aggregation. Secondly, I extend the algorithm for the GARCH(1,1)-in-mean model in the spirit of Engle, Lilien, and Robins (1987). This is a particularly interesting case since, under this specification, the mean equation has now a latent regressor. Finally, I show that the NL-ILS algorithm can also be implemented to estimate parameters from the RealGARCH(1,1)-in-mean model in the spirit of Hansen, Huang, and Shek (2011). This model turns out to be particularly important, because it is parameterized in such a way that there is a measurement equation linking the latent conditional variance to the realized measure. Hence, the RealGARCH model can be seen as an augmented GARCH model.

3.1 GARCH(1,1) and weak GARCH(1,1) models

I consider the GARCH(1,1) model as in (10) and (11). Using the sample mapping defined in (15), the NL-ILS algorithm is computed through the following steps:

Step 1: Given any initial estimate\(^6\) of $\phi$, denoted by $\hat{\phi}_0$ with $\hat{\phi}_0 \in \mathbb{B}$, compute, recur-

\(^6\)The starting value $\hat{\phi}_0$ can assume any value, provided that $\hat{\phi}_0 \in \mathbb{B}$, where $\mathbb{B}$ is the set of parameters satisfying the restrictions that guarantee the GARCH(1,1) model in (10) is second-order stationary. In both Monte Carlo study and empirical analysis, I obtain $\hat{\phi}_0$ by estimating (12) using residuals obtained from an AR($p$) model as regressors.
sively, estimates of $u_t$, denoted by $\hat{u}_{0,t}$, with:

$$\hat{u}_{0,t} = \epsilon_t^2 - \hat{\psi}_{0,0} - \sum_{i=1}^{q} \hat{\psi}_{0,i}\hat{u}_{0,t-i}$$  \hspace{1cm} (27)$$

where $\hat{\psi}_{0,0}$ and $\hat{\psi}_{0,i}$ denote the parameters from the $MA(\infty)$ representation computed using the starting values $\hat{\phi} = (\hat{\omega}_0, \hat{a}_0, \hat{b}_0)'$ and $\bar{q}$ is the truncation parameter defined exogenously. The Monte Carlo simulations showed that the size of $\bar{q}$ does not play a decisive role on both performance and convergence. As a standard rule, I fixed $\bar{q} = 3\sqrt{T}$. 

**Step 2:** Plug $\hat{u}_{0,t}$ into the sample mapping and minimize the sum of squared residuals with respect to $\hat{\phi}_1$ to obtain the first estimate of $\phi$, denoted by $\hat{\phi}_1$.

$$\hat{\phi}_1 = \hat{N}_T \left( \hat{\phi}_0 \right) = \min_{\hat{\phi}_1} \frac{1}{T} \sum_{t=1}^{T} \left[ \epsilon_t^2 - \hat{\psi}_{1,0} - \sum_{i=0}^{\bar{q}} \hat{\psi}_{1,i}\hat{u}_{1,t-1-i} \right]^2$$  \hspace{1cm} (28)$$

**Step 3:** Compute recursively a new set of residuals, denoted by $\hat{u}_{1,t}$, using $\hat{\phi}_1$ through (29):

$$\hat{u}_{1,t} = \epsilon_t^2 - \hat{\psi}_{1,1} - \sum_{i=0}^{\bar{q}} \hat{\psi}_{1,i}\hat{u}_{1,t-1-i}$$  \hspace{1cm} (29)$$

Repeat steps 2 and 3 $j$ times until $\hat{\phi}_j$ converges. I assume NL-ILS algorithm converges if $\left\| \hat{\phi}_j - \hat{\phi}_{j-1} \right\| < c$, where $c$ is exogenously defined. In both, Monte Carlo simulations and empirical application, I set $c = 1e-5$. Therefore, the $j^{th}$ iteration of the NL-ILS algorithm is given by the minimization below:

$$\hat{\phi}_j = \hat{N}_T \left( \hat{\phi}_{j-1} \right) = \min_{\hat{\phi}_j} \frac{1}{T} \sum_{t=1}^{T} \left[ \epsilon_t^2 - \hat{\psi}_{j,0} - \sum_{i=0}^{\bar{q}} \hat{\psi}_{j,i}\hat{u}_{j-1,t-1-i} \right]^2$$  \hspace{1cm} (30)$$

I denote the NL-ILS estimates obtained through the steps above by $\hat{\phi}$. The key factor that guarantees the NL-ILS algorithm converges is the contraction property yielded by the ACM condition on the population counterpart of (30). It is important to point out that the speed of convergence depends on the contraction parameter $\kappa$ as discussed in Section 2. Considering a specification such that $\alpha = 0.025$ and $\beta = 0.95$, the maximum eigenvalue associated with $V(\phi)$ is equal to 0.5. Adopting $c = 1e-5$ and provided that $\left| \hat{\phi}_0 - \phi \right| = 0.1$, convergence in this scenario would occur after ten iterations. This is 

Note that, under the true vector of parameters, the disturbances are iid process, implying, from Theorem 3.1 (Orthogonal Regression) in Greene (2008) - pg. 23, that estimates of $\psi$ are unbiased for any truncation parameter $\bar{q}$. At any iteration $j$, I only require the residuals to be uncorrelated and the resulting MA($\bar{q}$) model to be invertible.
in line with the results obtained in the Monte Carlo study, where convergence for the GARCH(1,1) takes, on average, 8 iterations.

The class of GARCH(1,1) model suffers from an important drawback: it is not closed under temporal aggregation. To overcome this issue, Drost and Nijman (1993) introduced the weak-GARCH(p,q) model. From their definition, a weak-GARCH(1,1) model, at some frequency \( m \), is given by:

\[
\begin{align*}
y_{(m)t} &= \epsilon_{(m)t} = \eta_{(m)t}\sigma_{(m)t} \\
\sigma^2_{(m)t} &= \omega_{(m)} + \alpha_{(m)}\epsilon^2_{(m)t-1} + \beta_{(m)}\sigma^2_{(m)t-1}
\end{align*}
\]  

(31)  

(32)

\[
E(\epsilon_{(m)t}) = 0
\]  

(33)

\[
P[\epsilon^2_{(m)t} | \epsilon_{(m)t-1}, \epsilon_{(m)t-2}, \ldots] = \sigma^2_{(m)t}
\]  

(34)

where \( P[\epsilon^2_{(m)t} | \epsilon_{(m)t-m}, \epsilon_{(m)t-2m}, \ldots] \) denotes the best linear predictor of \( \epsilon^2_{(m)t} \) in terms of the lagged values of \( \epsilon_{(m)t} \). An alternative definition of weak-GARCH (in terms of ARMA representation) is given by Francq and Zakoian (2010). They state that a process \( \epsilon_{(m)t} \) is generated by a weak-GARCH if \( \epsilon_{(m)t} \) is a white noise and \( \epsilon^2_{(m)t} \) admits an ARMA representation, such that \( u_{(m)t} \) in the ARMA(1,1) representation is a linear innovation with \( \text{Cov}(u_{(m)t}, \epsilon^2_{(m)t-l}) \) for all \( l > 0 \). By being closed under temporal aggregation, the weak-GARCH(1,1) model relaxes the assumption on sampling the data at the true data generation process frequency. This is particularly relevant when dealing with financial returns which are discrete representations from continuous processes. Drost and Werker (1996) establish the temporal aggregation, from the continuous time processes to the weak-GARCH models, providing closed solutions for the diffusion parameters as functions of the parameters of the weak-GARCH(1,1) model. Since this paper focuses on discrete time models, I restrict my analysis to the temporal aggregation provided by Drost and Nijman (1993). They define a bridge from the parameters of the strong GARCH(1,1) to the parameters of the weak-GARCH(1,1) sampled at some lower frequency \( m \) as the solution of the following system of equations:

\[
\omega_{(m)} = \omega \left[ \frac{1 - (\beta + \alpha)^m}{1 - (\beta + \alpha)} \right]
\]  

(35)

\[
\alpha_{(m)} = (\beta + \alpha)^m - \beta_{(m)}
\]  

(36)

\[
\beta_{(m)} = \frac{\beta (\beta + \alpha)^{m-1}}{1 + \alpha^2 \left[ \frac{1 - (\beta + \alpha)^{2m-2}}{1 - (\beta + \alpha)} \right] + \beta^2 (\beta + \alpha)^{2m-2}}
\]  

(37)

where \( \omega_{(m)} \), \( \alpha_{(m)} \) and \( \beta_{(m)} \) are parameters at some frequency \( m \) from the weak-GARCH(1,1) model and \( \phi_{(m)} = (\omega_{(m)}, \alpha_{(m)}, \beta_{(m)})' \).
Estimation of weak-GARCH models showed to be more difficult than its strong counterpart. In fact, the QMLE asymptotic theory for the weak-GARCH models remain to be established. Monte Carlo exercises, however, show that QMLE consistently estimate the parameters in (32). Francq and Zakoian (2000) establish the asymptotic theory for a two-stage least squares (LS) estimator.

Extending the NL-ILS algorithm to the weak-GARCH(1,1) is straightforward. Similar to the GARCH(1,1) model, (32) allows an ARMA(1,1) representation, implying that steps 1, 2 and 3 above can be performed to obtain estimates of \( \phi(m) \). The NL-ILS estimate of \( \phi(m) \) is denoted by \( \hat{\phi}(m) \). Corollary 1 delivers the consistency of the NL-ILS for the weak-GARCH(1,1) model. Convergence of the NL-ILS algorithm in the weak-GARCH(1,1) model is the same as in the GARCH(1,1) case, because they both share the same contraction parameter.

### 3.2 GARCH(1,1)-in-mean

I illustrate the NL-ILS estimator for the GARCH(1,1)-in-mean model. From (21), (22), (23) and (24), rewrite the model as:

\[
\begin{align*}
\sigma_t^2 &= \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \\
\epsilon_t &= y_t - \lambda \sigma_t \\
u_t &= \epsilon_t^2 - \omega - a \epsilon_{t-1}^2 - bu_{t-1}
\end{align*}
\]  

(38)  
(39)  
(40)

The NL-ILS algorithm is computed through the following steps:

**Step 1:** Choose an initial estimate of \( \theta_0 \), such that \( \theta_0 \in \mathbb{B} \), where \( \mathbb{B} \) is the set of parameters satisfying the second-order stationarity conditions. Using (39) and (38), compute recursively estimates of the conditional variance, denoted as \( \hat{\sigma}_{0,t}^2 \), and estimates of \( u_t \), denoted by \( \hat{u}_{0,t} \).

**Step 2:** From Proposition 1, the sample mapping in (26) can be split in two distinctive maps, such that \( \hat{\theta}_i \) is given by:

\[
\hat{\lambda}_1 = \left[ \hat{\sigma}_0' \hat{\sigma}_0 \right]^{-1} \hat{\sigma}_0' Y
\]

(41)

\[
\hat{\phi}_1 = \min_{\hat{\phi}_1} \frac{1}{T} \sum_{t=1}^{T} \left[ y_t - \hat{\lambda}_1 \hat{\sigma}_{0,t} \right]^2 - \sum_{i=0}^{q} \hat{\psi}_{1,i} \hat{u}_{0,t-1-i}\]

(42)

---

Note that under the temporal aggregation discussion in the previous subsection, the GARCH(1,1)-in-mean model depicted in (21), (22) and (23) is classified within the strong class of models. The literature, as far as I am aware, does not provide time aggregation results for the GARCH(1,1)-in-mean models. In this entire paper, therefore, I will refer to the strong GARCH(1,1)-in-mean as GARCH(1,1)-in-mean.

In practice, I firstly fix \( \lambda_0 = \left[ \frac{1}{T} \sum_{t=1}^{T} y_t \right] \left[ \text{Var}(y_t) \right]^{-1} \) and obtain \( \hat{\lambda}_0 \). As a second step, similarly to the GARCH(1,1) case, I estimate an AR(p) model having \( \hat{\sigma}_{0,t}^2 \) as dependent variable. This allows me to get initial estimates of \( u_t \) and hence compute \( \hat{\phi}_0 \).
where \( \tilde{\sigma}_0 \) and \( Y \) are \((T \times 1)\) vectors stacking all observations of \( \tilde{\sigma}_{0,t} \) and \( y_t \), respectively. Compute \( \hat{\lambda}_i \) through (41). Plug \( \hat{\lambda}_i \) into (42), and minimize with respect to \( \hat{\phi}_i \). Equations (41) and (42) deliver \( \hat{\theta}_i \).

**Step 3:** Using \( \hat{\theta}_i \), compute recursively \( \hat{\sigma}_{1,t}^2, \hat{\epsilon}_{1,t} \) and \( \hat{u}_{1,t} \) through (38), (39) and (40).

Repeat Steps 2 and 3 \( j \) times until \( \hat{\theta}_j \) converges. As in the GARCH(1,1) case, I assume \( \hat{\theta}_j \) converges if \( \| \hat{\theta}_j - \hat{\theta}_{j-1} \| < c \). On the \( j \)th iteration, the sample mapping resumes to:

\[
\hat{\lambda}_j = [\hat{\sigma}_{j-1}^2]^{-1} \hat{\sigma}_{j-1}^2 Y
\]

\[
\hat{\phi}_j = \min_{\phi_j} \frac{1}{T} \sum_{t=1}^{T} \left[ \left( y_t - \hat{\lambda}_j \hat{\sigma}_{j-1,t}^2 \right)^2 - \sum_{i=0}^{q} \hat{\psi}_{j,i} \hat{u}_{j-1,t-1-i} \right]^2
\]

Denote the resulting NL-ILS estimates as \( \hat{\theta} \). As in the GARCH(1,1) case, the key factor governing convergence of the NL-ILS estimator is the ACM property of \( N_T(\theta_j) \). It is important to stress that the procedure described above covers models with mean equation that accounts for more complex mean specifications. Among them, (39) can contain a constant, observed exogenous regressors or any function from the conditional variance. For all these scenarios, Proposition 1 holds, implying that sample mapping, on any \( j \)th iteration, can take the form of equations (43) and (44). With respect to computational issues, the NL-ILS estimator does not present significant numerical problems, achieving convergence for the great majority of replications. To this point, I found that by imposing constraints on the autoregressive and moving average parameters in (40), the rate of success of the NL-ILS algorithm improves. This showed to be a valid strategy when dealing with small samples and conditional variance specifications containing \( \beta \) very close to 1.

### 3.3 RealGARCH(1,1) and RealGARCH(1,1)-in-mean

The availability of high frequency data has triggered a new class of volatility estimators: the realized variance. These realized measures showed to be quite powerful on modeling the unobserved conditional variance. As pointed out by Andersen, Bollerslev, Diebold, and Labys (2003), realized volatility measures are able to respond faster to abrupt changes in the underline volatility than the standard GARCH framework, which may deliver massive improvements on volatility forecast. This important feature carried by the realized measures led to the establishment of models that combine the GARCH-type approach with the realized variance. Among these models, I point out the GARCH-X, HEAVY

\(^{10}\)As in the GARCH(1,1) case, I fix \( c = 1e - 5 \).
and RealGARCH proposed by Engle (2002), Shepard and Sheppard (2010) and Hansen, Huang, and Shek (2011), respectively. In this paper, I focus on extending the NL-ILS estimator to the case of RealGARCH(1,1) and RealGARCH(1,1)-in-mean model, since these models present the nice feature of having a measurement equation relating the realized variance to the latent conditional variance. The measurement equation accommodates the measurement error that arises from the difference between realized variance and the latent conditional variance, as pointed out by Asai, McAleer, and Medeiros (2011). This paper does not address the different realized variance estimators, (an extensive survey can be found on McAleer and Medeiros (2008)), I simply assume that the realized measures are obtained through consistent estimators and therefore are treated as observed variables. Following Hansen, Huang, and Shek (2011), a log-linear RealGARCH(1,1)-in-mean model is given by:

\[ y_t = \lambda \sigma_t + \epsilon_t \] (45)
\[ \epsilon_t = \eta_t \sigma_t \] (46)
\[ \ln \sigma_t^2 = \omega + \beta \ln \sigma_{t-1}^2 + \gamma \ln \nu_{t-1} \] (47)
\[ \ln \nu_t = \xi + \varphi \ln \sigma_t^2 + \tau(\eta_t) + z_t \] (48)

where \( \nu_t \) accounts for the realized variance, \( \tau(\eta_t) \) is a leverage function capturing asymmetries on the response of the realized measure to positive or negative shocks in \( \eta_t \). \( \eta_t \) and \( z_t \) are a iid processes with zero mean and variances equal to 1 and \( \sigma_z^2 \) respectively. As in Hansen, Huang, and Shek (2011)), I define the leverage function as:

\[ \tau(\eta_t) = \tau_1 \eta_t + \tau_2 (\eta_t^2 - 1) \] (49)

Note that (48) is a measurement equation relating \( \sigma_t^2 \) to \( \nu_t \). Its importance is twofold: firstly, it allows multi-step-ahead forecast, since the dynamics of \( \nu_t \) is fully specified; secondly, it helps identifying the parameters in (45), (47) and (48) when NL-ILS estimator is adopted. Denote \( \theta = (\lambda, \omega, \beta, \gamma, \xi, \varphi, \tau_1, \tau_2)' \). Estimation of the log-linear RealGARCH(1,1)-in-mean is originally undertaken through QMLE procedure. The QMLE estimates are denoted by \( \hat{\theta}_Q \). Hansen, Huang, and Shek (2011) discuss the asymptotic properties of \( \hat{\theta}_Q \) for the standard log-linear RealGARCH(1,1) case.

Hansen, Huang, and Shek (2011) derives a VARMA(1,1) representation of the two processes: \( \ln \epsilon_t^2 \) and \( \ln \nu_t \):

\[
\begin{pmatrix}
\ln \nu_t \\
\ln \epsilon_t^2
\end{pmatrix} =
\begin{pmatrix}
\mu_\nu & 0 \\
0 & \mu_\epsilon
\end{pmatrix}
\begin{pmatrix}
\ln \nu_{t-1} \\
\ln \epsilon_{t-1}^2
\end{pmatrix} +
\begin{pmatrix}
w_t \\
u_t
\end{pmatrix}
+
\begin{pmatrix}
-\beta & 0 \\
\gamma & -\rho
\end{pmatrix}
\begin{pmatrix}
w_{t-1} \\
u_{t-1}
\end{pmatrix} \] (50)
where \( \mu_t = \omega + \gamma \xi + (1 - \beta - \varphi \gamma) \bar{\eta}_t^2 \), \( \mu_\nu = \varphi \omega + (1 - \beta) \xi \), \( \rho = \beta + \varphi \gamma \), \( \bar{\eta}_t^2 = \mathbb{E} (\ln \eta_t^2) \), \( w_t = \tau (\eta_t) + z_t \) and \( u_t = \ln \eta_t^2 - \bar{\eta}_t^2 \). Note that \( \ln \nu_t \) in (50) does not depend on the \( \ln \epsilon_t^2 \) nor on \( u_t \), implying that the parameters on the first equation of the VARMA(1,1) representation can be estimated separately from the parameters governing \( \ln \nu_t \). Furthermore, necessary condition an ARMA(1,1) model to be second-order stationarity implies \( |\rho| < 1 \). Using this result, (50) can be expressed as a VMA(\( \infty \)) process:

\[
\begin{pmatrix}
\ln \nu_t \\
\ln \epsilon_t^2
\end{pmatrix} = \begin{pmatrix}
\frac{\mu_\nu}{1 - \rho} \\
\frac{\mu_\nu}{1 - \rho}
\end{pmatrix} + \begin{pmatrix}
w_t \\
u_t
\end{pmatrix} + \begin{pmatrix}
\sum_{i=0}^{\infty} \rho^i (\rho - \beta) L^i \\
\sum_{i=0}^{\infty} \gamma \rho^{i+1} L^i
\end{pmatrix} \begin{pmatrix}
w_{t-1} \\
u_{t-1}
\end{pmatrix} \tag{51}
\]

Note that the generic model in (1), (3) and (4) nests the RealGARCH(1,1)-in-mean by setting the mean equation as \( \lambda \sigma_t \); \( \alpha \), \( \beta \) and \( \gamma \) to one; and (51) satisfies (4). Considering the baseline log-linear RealGARCH(1,1)-in-mean model and the compact representation of the \( (\ln \epsilon_t^2, \ln \nu_t)' \) in (51), the NL-ILS is computed through the following steps:

**Step 1:** Choose an initial estimate of \( \theta \), such that \( \theta_0 \in \mathbb{B} \), where \( \mathbb{B} \) is the set of parameters satisfying the second-order stationarity conditions. Compute initial estimates of \( \mu_\nu \) and \( \rho \), denoted as \( \hat{\mu}_\nu,0 \) and \( \hat{\rho}_0 \) respectively. Denote \( \hat{\phi}_0 = (\hat{\mu}_\nu,0, \hat{\rho}_0, \hat{\beta}_0)' \) as the vector of parameters describing the \( \ln \nu_t \) process. Compute recursively an initial set of disturbances \( \hat{\omega}_{0,t} \) using:

\[
\hat{\omega}_{0,t} = \ln \nu_t - \hat{\mu}_\nu,0 - \hat{\rho}_0 \ln \nu_{t-1} + \hat{\beta}_0 \hat{\omega}_{0,t-1} \tag{52}
\]

**Step 2:** Recursively in (47), compute \( \hat{\sigma}_{0,t}^2 \) assuming \( \hat{\theta}_0 \).

**Step 3:** By truncating, at some lag order \( \tilde{q} \), the first row in (51), write the first sample mapping similarly to the GARCH(1,1) case as:

\[
\hat{\phi}_1 = \hat{N}_T \left( \hat{\phi}_0 \right) = \min_{\hat{\phi}_1} \frac{1}{T} \sum_{t=1}^{T} \left[ \ln \nu_t^2 - \hat{\psi}_{1,o} - \sum_{i=0}^{q} \hat{\psi}_{1,i} \hat{\omega}_{0,t-i} \right]^2 \tag{53}
\]

where \( \hat{\psi}_{1,o} = \hat{\mu}_{\nu,1} / (1 - \hat{\rho}_1) \) and \( \hat{\psi}_{1,i} = \hat{\rho}_1 \left( \hat{\rho}_1 - \hat{\beta}_1 \right) \). Minimize (53) with respect to \( \hat{\phi}_1 \) and obtain \( \hat{\phi}_1 \). Using (52), compute recursively \( \hat{\omega}_{1,t} \).

**Step 4:** From the second equation in (51), define the second sample mapping:

\[
\hat{\zeta}_1 = \hat{M}_T \left( \zeta_0 \right) = \min_{\hat{\zeta}_1} \frac{1}{T} \sum_{t=1}^{T} \left[ \ln \left[ \left( y_t - \hat{\lambda}_1 \hat{\sigma}_{0,t} \right)^2 \right] - \frac{\hat{\mu}_{\epsilon,0}}{1 - \hat{\rho}_1} - \sum_{i=0}^{q} \hat{\rho}_1 \hat{\gamma}_1 \hat{\omega}_{1,t-1-i} \right]^2 \tag{54}
\]

where \( \hat{\zeta}_1 = (\hat{\lambda}_1, \hat{\mu}_{\epsilon,0}, \hat{\gamma}_0)' \). Similarly to Step 2 in the GARCH(1,1)-in-mean case,
Proposition 1 allows to split the mapping in (54) such that:

\[ \hat{\lambda}_i = \left[ \hat{\sigma}_o' \hat{\sigma}_o \right]^{-1} \hat{\sigma}_o' Y \]  

(55)

\[ \hat{\zeta}_i^* = \min_{\hat{\zeta}_i} \frac{1}{T} \sum_{t=1}^{T} \left[ \ln \left( \left( y_t - \hat{\lambda}_i \hat{\sigma}_{0,t} \right)^2 \right) - \frac{\hat{\mu}_{c,0}}{1 - \hat{\rho}_1} - \sum_{i=0}^{\hat{q}} \hat{\gamma}_i \hat{\omega}_{i,t-1} \right]^2 \]  

(56)

where \( \hat{\sigma}_o \) and \( Y \) are \((T \times 1)\) vectors stacking all observations of \( \hat{\sigma}_{o,t} \) and \( y_t \). Compute \( \hat{\lambda}_i \) through (55). Plug \( \hat{\lambda}_i \) into (56), and minimize with respect to \( \hat{\zeta}_i^* = (\hat{\mu}_{c,1}, \hat{\gamma}_1)' \).

**Step 5:** Based on \( \hat{\phi}_1 \), \( \hat{\zeta}_1 \), and \( \bar{\eta}^2 \) solve the following system of equations to find \( \hat{\omega}_1 \), \( \hat{\xi}_1 \), and \( \hat{\phi}_1 \).

\[ \hat{\phi}_1 = \frac{\hat{\rho}_1 - \hat{\beta}_1}{\hat{\gamma}_1} \]  

(57)

\[ \hat{\xi}_1 = \left[ \hat{\mu}_{\nu_1} - \hat{\phi}_1 \hat{\mu}_{c_1} + \hat{\phi}_1 \left( 1 - \hat{\beta}_1 - \hat{\phi}_1 \hat{\gamma}_1 \right) \right] \left( 1 - \hat{\beta}_1 - \hat{\phi}_1 \hat{\gamma}_1 \right) \]  

(58)

\[ \hat{\omega}_1 = \hat{\mu}_{\nu_1} - \hat{\gamma}_1 \hat{\xi}_1 - \left( 1 - \hat{\beta}_1 - \hat{\phi}_1 \hat{\gamma}_1 \right) \bar{\eta}^2 \]  

(59)

**Step 6:** Recursively in (47), compute \( \hat{\sigma}_{1,t}^2 \) using \( \left( \hat{\omega}_1, \hat{\beta}_1, \hat{\gamma}_1 \right)' \). Retrieve estimates of \( \eta_t \) through \( \bar{\eta}_{1,t} = \frac{\left( y_t - \hat{\lambda}_1 \hat{\sigma}_{1,t} \right)}{\hat{\sigma}_{1,t}} \) and obtain \( \hat{\gamma}_1 \) by estimating (48) using \( \hat{\sigma}_{1,t} \) as a regressor.

Repeat Steps 3, 4, 5 and 6 until \( \hat{\theta}_j \) converges. As in the previous models, I assume convergence occurs if \( \left\| \hat{\theta}_j - \hat{\theta}_{j-1} \right\| < c \). Note that the NL-ILS algorithm requires \( \bar{\eta}^2 \) to be defined exogenously, which implies that some distributional assumption has to be made on \( \eta_t \). The Monte Carlo simulations showed that the NL-ILS algorithm takes more iterations to converge, which indicates that the contraction parameter associated with the RealGARCH(1,1)-in-mean model is higher than the one found in the GARCH(1,1) model. The steps above also hold for computing the RealGARCH(1,1) model. In this case, (55) in step 4 drops out.

### 4 Monte Carlo Study

This section addresses the performance of the NL-ILS estimator discussed in the previous section, when estimating the GARCH(1,1), weak-GARCH(1,1), GARCH(1,1)-in-mean and RealGARCH(1,1)-in-mean models. My main focuses are on assessing consistency, efficiency and forecast performance of the NL-ILS estimator compared with the benchmark estimator: the MLE. For the GARCH(1,1)-in-mean case, I discuss an additional study: I appreciate the behavior of the \( \lambda \) estimates (risk premium parameter) when the conditional variance is misspecified. This set of experiments plays the role of robustness
analysis, since it is known that MLE estimates of $\lambda$ can be biased when the conditional variance is misspecified. This follows from the fact that the information matrix is no longer block diagonal in the GARCH(1,1)-in-mean specification. In all exercises, I fix the number of replications to 1500 unless otherwise stated. I also discard the initial 500 observations to reduce dependence on initial conditions. All models are estimated using the CML optimization library in GAUSS. Results for the GARCH(1,1) and weak-GARCH(1,1) models are reported in terms of the median and the relative root mean squared error (RRMSE). The relative measures are computed using the MLE benchmark (in the denominator), implying that NL-ILS outperforms the MLE estimator when the relative measures are lower than one.

The first set of simulations analyzes the performance of the NL-ILS for the GARCH(1,1) model. The data generating process follows the baseline model displayed in (10) and (11). The stochastic term $\eta_t$ is set to be normally distributed, with zero mean and variance equal to one. Table 1 displays results for two different specifications and five different sample sizes ($T = 100, T = 200, T = 300, T = 400$ and $T = 500$). I focus on high persistent GARCH(1,1) specifications with $(\alpha + \beta)$ close to one, because these are the most usual cases reported when modeling financial returns. As an overall picture, I find that NL-ILS estimates outperform the MLE benchmark when $T$ is small. This was somehow expected, since MLE estimator is known to suffer from numerical problems either when $T$ is small or $(\alpha + \beta)$ approaches to one. The outstanding performance in small samples is particularly welcome when dealing with variables that may have structural breaks and also for forecasting purposes (see the work of Giraitis, Kapetanios, and Yates (2010)). When $(\alpha + \beta) = 0.995$, the NL-ILS-ILS estimator outperforms the MLE in all sample sizes, achieving its best performance when $T = 100$, with gains of 61%, 44% and 61% for the $\omega$, $\alpha$ and $\beta$ parameters, respectively. Considering the specification where $(\alpha + \beta) = 0.97$, I find that the MLE estimator improves its performance, yielding more accurate estimates than the NL-ILS estimator for all sample sizes, but $T = 100$.

Table 2 and 3 report results for the weak-GARCH(1,1) model. The weak-GARCH(1,1) processes are generated in the spirit of Drost and Nijman (1993). I firstly generate a GARCH(1,1) process using (10) and (11). The vector $\theta = (\omega, \alpha, \beta)'$ used in this specification contains the high frequency parameters. The stochastic term is assumed to be normally distributed, such that $\eta_t \sim (0, 1)$. Given the high frequency GARCH(1,1) process, I re-sample $y_t$ at different frequencies $m$, yielding $y_{(m)t}$. When re-sampling, I assume $y_t$ is a stock variable, rendering $y_{(m)t}$, $t = m, 2m, ...., T$. The low frequency parameters are computed through (35), (36) and (37) and denoted as $\theta_m$. Table 2 displays

\[\text{CML (Constrained Maximum Likelihood Estimation) library in GAUSS designed to solve maximum likelihood functions subject to linear and nonlinear constraints. In all Monte Carlo simulations, I set global variables in CML to their default values, because this specification is flexible enough to accommodate endogenous changes in both algorithms and grid search procedures.}\]

\[\text{Additional results considering different specifications are available upon request.}\]
results obtained when the high frequency α and β parameters are set equal to 0.05 and 0.94 respectively. In this scenario, as m increases (the resulting weak-GARCH(1,1) is sampled at a lower frequency), the NL-ILS estimator improves its performance when compared to the MLE benchmark. I argue that the reasons for that are twofold: firstly, as observed in the GARCH(1,1) case, NL-ILS has a better performance than MLE estimator for small samples. This plays an important role in this setup, since as m increases T(m) decreases, following the fact that T (the high frequency sample size) is constant. The second reason arises from the robustness of the NL-ILS estimator to disturbances that present nonlinear dependence. Comparing the relative measures obtained in the weak-GARCH(1,1) experiment, with the ones obtained with the GARCH(1,1), I find that performance gains from the NL-ILS estimator with respect to the MLE estimator are higher for the weak-GARCH(1,1) model. It is also relevant to point out that NL-ILS improves its performance with respect to MLE estimator when β is approaches one. Comparing the results when m = 3 in tables 2 and 3, I find that NL-ILS improves the RRMedSE in 26% and 17% for α(m) and β(m) respectively. This is a particular relevant result, since it mimics financial series that usually display β very close to one. NL-ILS procedure delivers less biased estimates than the MLE benchmark, when the bias is assessed through the median. The reason for this discrepancy between the mean and the median arises from the presence of outliers. From the data generation process specification, α is very close to zero and β is close to one, rendering autoregressive and moving average parameters very close to each other. This feature may lead to problems in the optimization of the sample mapping, yielding local minimums instead of global ones. This numerical problem usually generates outliers, impacting the mean and the variance of the NL-ILS estimator.

The third set of simulations focuses on analyzing the performance of the NL-ILS estimator for the GARCH(1,1)-in-mean specification. Table 4 and 5 report results considering different specifications and sample sizes. I generate the data using (21), (22) and (23), where ηt is assumed to be normally distributed with zero mean and variance equal to one. To reduce the impact of the outliers when assessing the comparison between the NL-ILS and the MLE estimators, I display results in terms of relative root median squared error (RRMedSE) and the relative root median squared forecast error (RRMedSFE). The former one is adopted when analyzing the parameters of the GARCH(1,1)-in-mean, whereas the RRMedSFE is used when evaluating the out-of-sample forecast performance for both risk premium and conditional variance. I denote the out-of-sample risk premium forecast, at some horizon h, as: \( \hat{\pi}_{t+h} = \hat{\lambda}_t \hat{\sigma}_{t+h} \). As in the GARCH(1,1) and weak-GARCH(1,1) cases, the relative measures are computed having the MLE as the benchmark. I also report results considering the MLE algorithm computed using the NL-ILS estimates as starting values. I denote this results as MLE*.

Table 4 displays results for two different GARCH(1,1)-in-mean specifications. These
specifications present \((\alpha + \beta)\) close to one (0.995 and 0.97, respectively) and \(\alpha = 0.025\). With respect to the RMedSE associated with the parameters, I find that NL-ILS outperforms the MLE benchmark for sample sizes up to \(T = 300\) (except for the \(\lambda\)) when \((\alpha + \beta) = 0.995\). This conclusion is in line with my previous findings for the GARCH(1,1) and weak-GARCH(1,1). When \(T\) gets large, the MLE estimator performance improves reasonably fast, outperforming the NL-ILS estimator. Again, this pattern is expected, since the MLE is extremely difficult to be beaten in medium samples and the model is correctly specified. As discussed in Section 2, NL-ILS is consistent, presenting a bias\(^{13}\) of only 0.008 when \(T=750\), whereas the bias associated with the MLE estimator is 0.010. The main determinant of the poorer performance of the NL-ILS estimator for large \(T\) lies on the presence of many more outliers than the ones found when the MLE algorithm is implemented. Concluding this point, I find that MLE algorithm is able to reduce the variance associated with the estimates much faster than the NL-ILS algorithm as \(T\) gets large. The poorer performance of the MLE* algorithm is also explained by the outliers. Hence, when \(T\) is small and the starting values in the MLE algorithm are very bad, the final outcome is likely to be also very poor. As expected, as \(T\) gets large MLE* algorithm converges to the standard MLE estimator.

The forecast performance of models estimated using the NL-ILS estimator is also worthily to be highlighted. In particular, I find very good results on forecasting the conditional variance up to \(T = 300\). I report gains of up to 51\% in terms of the RRMedFE. The surprisingly good performance of MLE* algorithm arises from the bias on estimating \(\omega\) and \(\alpha\). In practise, when \(\beta\) is very high (as is the case in this specification), models that present a bias combination such that, \(\hat{\beta}\) is downward biased and \(\hat{\omega}\) and \(\hat{\alpha}\) are upward biased, tend to perform well on forecasting, due to level effect. Regarding the forecast of the risk premium, I find that NL-ILS is not able to outperform the MLE benchmark for any value of \(T\). The reason for that is the poorer performance of the NL-ILS algorithm on estimating the risk premium parameter \(\lambda\).

The second model in table 4 presents a similar pattern, across different sample sizes, than the first specification discussed above. The main difference arises from the higher RRMEDSE associated with the parameters (except for \(\alpha\)). As noted before in the strong- and weak-GARCH(1,1) cases, MLE improves its performance, compared to NL-ILS, as \(\beta\) decreases.

Table 5 displays two additional GARCH(1,1)-in-mean specifications. Their main difference lies on the higher value impounded to \(\alpha\): \(\alpha = 0.08\). Apart from the case where \(T = 100\), I find that MLE delivers more accurate estimates and forecasts than the NL-ILS. This result is in line with my previous findings, indicating that the NL-ILS algorithm has an outstanding performance either when \(T\) is small or \(\beta\) is very close to one. Considering the forecast performance analysis, it is important to point out that, when outperformed

\(^{13}\)I compute the bias using the median within all replications.
by the MLE estimator, the NL-ILS results are, on average, only 5% to 10% worse than
the results obtained with the MLE estimator.

Table 6 displays results considering the performance of the NL-ILS estimator when applied
to the log-linear RealGARCH(1,1)-in-mean. The parameters in the RealGARCH(1,1)-
in-mean are set as the ones Hansen, Huang, and Shek (2011) found in their empirical
application. Overall, the results are favorable to the MLE estimator for both parameter
estimation and out-of-the-sample forecast of the conditional variance. The RRMSE of the
parameter estimates are very high indicating a poor performance of the NL-ILS estimator.
I argue that this poor performance comes from the higher variance associated with
the NL-ILS estimates (presence of outliers), since the bias assessed through the median
within all replications is neglectful when $T$ is large. In particular, it is important to point
out that the MLE estimator is able to extract a considerable benefit out of the inclusion
of the measurement equation. Turning the analysis to the forecast performance, the NL-
ILS is able to outperform the MLE benchmark for the realized variance for sample sizes
up to $T = 300$. This good performance on forecasting the realized variance contributed
for a decent performance on forecasting the conditional variance.

4.1 Robustness

Mixed evidences, in both sign and significance of the $\lambda$ parameter, have been found in the
literature when estimating the risk premium using the full parametric GARCH-in-mean
model and its variants. While French, Schwert, and Stambaugh (1987) found a positive
value for $\lambda$, Glosten, Jagannathan, and Runkle (1993) found an opposite sign and Baillie
and DeGennaro (1990) found very little evidence for a statistically significant $\lambda$. Con-
sidering the semiparametric approach, Linton and Perron (2003), Conrad and Mammen
(2008) and Christensen, Dahl, and Iglesias (2012) found strong evidences of nonlinearity
governing the risk premium function. Focusing on the full parametric GARCH-in-mean
models, mixed results on the $\lambda$ estimates can be motivated by lack of consistency of the
QMLE estimator. As discussed in Bollerslev, Chou, and Kroner (1992), QMLE estimates
of GARCH-in-mean parameters may be inconsistent when the conditional variance is
misspecified. This drawback arises because the information matrix is not block diagonal
between the parameters in the conditional mean and the conditional variance. The task
of correctly specifying the conditional variance is extremely difficult given the large menu
of alternative models available in the literature (see Francq and Zakoian (2010), Bollerslev
(2008) for a surveys on GARCH-type models). To study the performance of the NL-ILS
estimator when the conditional variance is misspecified, I carry out four different experi-
ments: in the first one, the conditional variance is specified as being an APARCH model
(see Ding, Granger, and Engle (1993)); in the second exercise, the conditional variance is
set to follow an EGARCH model in the spirit of Nelson (1991); the third exercise consists
on modeling the conditional variance as a JGR-GARCH as in Glosten, Jagannathan, and Runkle (1993); the final simulation is carried out using a GARCH(2,2) specification. Note that both EGARCH and JGR-GARCH models capture asymmetric responses of the conditional variance to positive and negative shocks, whereas the APARCH specification manages to capture three important stylized facts: long memory, dependence on some power transformation of the conditional standard deviations and asymmetric responses to positive and negative shocks. I assess performance on estimating $\lambda$ through the RRMSE and bias. Forecast performance is assessed using the RRMSFE. As in the previous experiments, the MLE estimator is the benchmark for all the relative measures. I also report the results for the MLE estimates which are computed using the NL-ILS estimates as starting values (MLE*).

Table 7 displays results for the APARCH and EGARCH models. Considering the APARCH results, I find that the NL-ILS estimator outperforms the MLE benchmark for all the different sample sizes. The difference in performance achieves 22% when $T = 1750$. When looking at the bias, the conclusion is even more favorable to the NL-ILS estimator. I find that MLE estimates are downward biased in 15%, when $T = 1750$, whereas the bias related to the NL-ILS is neglectful. In spite of the good performance on estimating $\lambda$, the NL-ILS algorithm fails on achieving outstanding results on forecasting both the risk premium and the conditional variance. In both cases, the MLE estimator delivers more accurate forecasts.

When the conditional variance is misspecified using the EGARCH specification, the results regarding the estimation of $\lambda$ are, again, extremely favorable to the NL-ILS estimator. Table 7 reports gains of 47% in terms of the RRMSE with respect to the MLE benchmark. The outstanding difference in performance is consistent through all the sample sizes, showing the robustness of the NL-ILS estimator. Analyzing the bias computed from both estimators, I find a similar picture as in the APARCH case: NL-ILS delivers neglectful bias, indicating consistency, whereas MLE is upward biased in 15%. A different picture arises when considering the forecast performance of the risk premium function. The NL-ILS estimator is now able to outperform the MLE benchmark in up to 28%, considering the median within all forecast horizons. I claim that this difference in performance comes mostly from the best estimation of $\lambda$, since NL-ILS does a worse job on forecasting the conditional variance.

Table 8 displays results for models using GJR-GARCH and GARCH(2,2) specifications. Considering the latter one, the results are very similar to the standard GARCH(1,1)-in-mean experiments carried out previously in this section. Overall, MLE provides more accurate results for both the estimation of $\lambda$ and the variance and risk premium forecasts. It is important to point out, however, that bias associated with the NL-ILS estimates of $\lambda$ is still neglectful. The results associated with the GJR-GARCH model follows the same pattern as the ones obtained with the APARCH and EGARCH specifications. I find that
the MLE estimates of $\lambda$ are biased, leading to a poorer performance of this estimator when compared to the NL-ILS procedure.

Overall, the conclusion obtained from this set of experiments is that NL-ILS is more robust than the MLE benchmark when the conditional variance is misspecified. Moreover, MLE delivers biased estimates of $\lambda$ when the conditional variance is misspecified in such a way that it possesses either asymmetric responses to positive and negative shocks or dependence at different moments than the second one. Finally, I claim that, under misspecification of the conditional variance, inference using the MLE framework may no longer be a valid alternative.

5 Empirical application

I examine the significance of the risk premium parameter using the GARCH(1,1)-in-mean framework by adopting the NL-ILS estimator discussed in the previous sections. As discussed in the Monte Carlo section, the performance of both NL-ILS and QMLE estimator may vary when dealing with weak processes, such as the weak-GARCH(1,1). As the true data generation process governing the excess returns are believed not to be discrete (such as daily, weekly or monthly), the impact of time aggregation on the consistency of the $\lambda$ estimates needs to be addressed. To this purpose, I construct nine different data sets on excess returns, comprehending three different indices (CRSP value-weighted index, S&P500 and S&P100) at three different frequencies: daily, weekly and monthly. The CRSP value-weighted index is considered as the most complete (in market sense) index, being therefore the best proxy for the market, as pointed out by Linton and Perron (2003). Hence, by using “least complete” indices, I check whether the significance of the risk premium parameter depends on the coverage of the index. Excess returns for the three indices are computed deducting the risk free rate (one-month Treasure bill rate) from their log returns\textsuperscript{14}. Table 9 reports the descriptive statistics. The daily CRSP and S&P500 indices spans from 28/06/1963 to 29/09/2011, accounting for 12,148 observations. The S&P100 index spans from a smaller period, (04/08/1982 - 29/09/2011), yielding 7,364 observations. CRSP and S&P500 indices have 2,426 and 740 observations for weekly and monthly frequencies, respectively. S&P100 index contains 1,469 and 330 observations on the weekly and monthly frequencies, respectively. Standard errors for the NL-ILS estimator are computed using block bootstrap with one thousand replications, whereas for the QMLE estimator the Bollerslev-Wooldridge robust standard errors are implemented.

I start discussing the results reported in Table 10, where I estimate a GARCH(1,1)-in-mean model using both, NL-ILS and QMLE estimators. At daily frequency, the $\lambda$

\textsuperscript{14}CRSP value-weighted index and one-month Treasure bill rate were downloaded from WRDS - Wharton Research Data Services, whereas S&P500 and S&P100 indices were obtained from Yahoo! finance.
estimates obtained using the NL-ILS estimator are significant at 5%\textsuperscript{15} level only for the CRSP index, whereas QMLE estimator delivers significant estimates for all series. Considering the parameters in the conditional variance equation, both procedures deliver highly significant estimates of the parameters $\alpha$ and $\beta$. Both methodologies also yield high degrees of persistence ($\alpha + \beta$). With respect to this point, it is relevant to mention that, in all indices, the persistence obtained using QMLE estimator is always higher (average of 0.99502) than the ones obtained using the NL-ILS estimator. To check this issue, I also performed the MLE estimation using the GED distribution. Results did not show any significant quantitative change. There is an important difference in magnitude from the $\lambda$ estimates obtained with the NL-ILS algorithm and the ones obtained with the MLE methodology. In fact, the difference between them turned out to be statistically significant\textsuperscript{16}, indicating the possibility of MLE being upward biased following discussion in Section 4.

Moving to the weekly and monthly frequencies, their patterns remain very similar to the one previously discussed. NL-ILS delivers $\lambda$ estimates which are significant at 5% level only for the CRSP index, whereas QMLE estimates are significant for all indices. The $\alpha$ and $\beta$ parameters from the conditional variance equation remain highly significant, yielding a high degree of persistence in the conditional variance. The results in table 10 turn out to be consistent with the previous findings in the literature. Linton and Perron (2003) found a value of $\hat{\lambda}$ when estimating a EGARCH-in-mean very close to the NL-ILS estimates. The same applies for Christensen, Dahl, and Iglesias (2012), who found significant QMLE estimates of $\lambda$ for the daily S&P500 index.

As a second step of my investigation, I incorporate realized measures of volatility in this analysis by estimating the RealGARCH(1,1)-in-mean for the S&P500 index\textsuperscript{17}. Table 11. Results in table corroborate my previous findings: risk premium parameter is not significant to any of the sample frequencies.

Analyzing the empirical results at the light of the results obtained in the Monte Carlo section, I claim that this difference in magnitude and significance may be caused by bias on the QMLE estimates, following misspecification of the conditional variance. Assuming my claim is correct, I outline two conclusions: firstly, the risk premium parameter is only significant for the most complete index (CRSP), whereas for the “less complete” indices the risk-return tradeoff does not hold. This finding is consistent with the theoretical results in Merton (1973), that requires the existence of a market portfolio. Hence, the fact that the $\lambda$ estimates obtained from the S&P500 and S&P100 are not significant may

\textsuperscript{15}T-statistics for the NL-ILS estimate of $\lambda$ is on the boundary of 5% level significance. However, looking at the empirical distribution computed from the bootstrapped estimates of $\lambda$, the NL-ILS turned out to be significant at 5% level.

\textsuperscript{16}T-statistics are 4.77, 3.44, 2.50 for CRSP, S&P500 and S&P100, respectively.

\textsuperscript{17}Realized measures of the conditional variance were obtained from the Oxford-Man Institute of Quantitative Finance (realized Library). Unfortunately, among the three different indices I adopt in this paper, there is only availability of data for the S&P500 index.
imply that these two indices are not good proxies for the market. Secondly, I conclude that the NL-ILS estimator is the most suitable for dealing with the task of estimating the risk premium parameter, since, as observed in the section 4, it is robust to misspecification of the conditional variance.

5.1 Empirical application: Robustness

As a robustness check, I estimate the risk premium parameter using three alternative models: APARCH(1,1,1)-in-mean, EGARCH(1,1,1)-in-mean and GJR-GARCH(1,1,1)-in-mean models. All the three models are estimated using the QMLE procedure. Table 12 reports results for all indices at all frequencies. By using models that allow for asymmetric response of the conditional variance to positive and negative shocks, it turns out that the risk premium parameter $\lambda$ is only significant for the CRSP index. This finding strengthens the conclusion that the GARCH(1,1)-in-mean estimated with the MLE estimator is not robust enough to misspecification of the conditional variance equation, leading to misleading results.

6 Conclusion

This paper introduces a novel estimator: the nonlinear iterative least squares (NL-ILS). To illustrate the NL-ILS estimator, I provide algorithms covering the GARCH(1,1), weak-GARCH(1,1), GARCH(1,1)-in-mean and RealGARCH(1,1)-in-mean models. I show that the NL-ILS estimator is particularly useful when innovations in the mean equation have some degree of dependence or the variance equation is misspecified. These both features may lead to inconsistency when the QMLE procedure is implemented. I establish the consistency and asymptotic distribution for the NL-ILS estimator covering the GARCH(1,1) model and extend the consistency result for the weak-GARCH(1,1) model. The assumptions I require for the asymptotic theory are compatible with the QMLE estimator. Through an extensive Monte Carlo study, I show that the NL-ILS estimator outperforms the MLE benchmark in a variety of scenarios including the following: the sample size is small; the $\beta$ parameter in the conditional variance has values very close to one, as widely found in empirical studies; or the true data generation process (DGP) is the weak-GARCH(1,1), indicating that the NL-ILS estimator is more robust to the presence of dependence on the innovations. Moreover, I show that the NL-ILS estimator is more robust to misspecified conditional variance, delivering neglectful biases on estimating the risk premium parameter in a GARCH(1,1)-in-mean model. In contrast with the NL-ILS algorithm, the MLE estimator presents biases of up to 15%, leading to the differences in performances of up to 22%, in terms of the relative mean squared error, when estimating the risk premium parameter. The NL-ILS estimator also delivers
more accurate out-of-the-sample forecasts for the risk premium function when the DGP is either the EGARCH(1,1,1)-in-mean or the GJR-GARCH(1,1,1)-in-mean models.

An empirical application addressing the significance of the risk premium parameter through a full parametric GARCH-in-mean and RealGARCH(1,1)-in-mean models is provided. I undertake my analysis through two different dimensions: temporal aggregation and market representation. The latter dimension is appraised by using the CRSP, S&P500 and S&P100 indices, which possess distinct market coverages, whereas the former dimension is assessed by aggregating the series at daily, weekly and monthly basis. When adopting the robust NL-ILS estimator and the QMLE benchmark to assess significance of the risk premium parameter, the results turned out to be very different: the NL-ILS estimator delivered risk premium estimates which are significant only for the CRSP index at all its frequencies; the QMLE estimator, however, provides estimates which are significant to all three data sets, in all frequencies. Moreover, the difference in magnitude between the NL-ILS and QMLE estimates are also significant in some data sets, indicating a potential bias. By using the Monte Carlo results, I argue that the QMLE estimator provides biased estimates following a misspecified conditional variance. As a robustness check for the empirical results, I estimate RealGARCH(1,1)-in-mean, EGARCH(1,1,1)-in-mean, APARCH(1,1,1)-in-mean and GJR-GARCH(1,1,1)-in-mean models and their results corroborate my findings using the GARCH(1,1)-in-mean estimated with NL-ILS algorithm: the risk premium parameter is only significant for the CRSP index at all frequencies. Ultimately, this paper suggests the use of the NL-ILS estimator on modeling the conditional volatility in the presence of dependent errors and misspecification. I highlight the robustness properties of the NL-ILS estimator assessing the risk premium in different indices and sampling frequencies.
References


Dias, G. F., and G. Kapetanios (2012): “Forecasting Medium and Large Datasets with Vector Autoregressive Moving Average (VARMA) models,” working paper, Queen Mary University of London.


7 Appendix

Proof of Lemma 1: It mirrors the proof in Lemma 5 in Dominitz and Sherman (2005) and Dias and Kapetanios (2012). Define the population mapping evaluated at some vectors of parameters $\xi$ and $\zeta$, such that $\xi, \zeta \in \mathbb{B}$. By Taylor expansion, rewrite $|N(\xi) - N(\zeta)|$ defining a bound that contains the gradient of the population mapping evaluated on $\phi$.

$$|N(\xi) - N(\zeta)| = |V(\xi)\ [\xi - \zeta]| \leq |V(\phi)\ [\xi - \zeta]| + 
+ |[V(\xi) - V(\phi)]\ [\xi - \zeta]| + o_p(|\xi - \zeta|)$$\hfill (60)

Using Dominitz and Sherman (2005) result (Lemma 5), it suffices to show that the maximum eigenvalue of $V(\phi)$ is less than one in absolute value to prove Lemma 1. By applying the NR procedure, the population and sample mapping in (14) and (15) can be linearized as:

$$\phi_{j+1} = N(\phi_j) = \phi_j - [H(\phi_j)]^{-1} G(\phi_j)$$ \hfill (61)

$$\hat{\phi}_{j+1} = \hat{N}_T(\hat{\phi}_j) = \hat{\phi}_j - [\hat{H}_T(\hat{\phi}_j)]^{-1} \hat{G}_T(\hat{\phi}_j)$$ \hfill (62)

where $\hat{G}_T(\hat{\phi}_j)$ and $\hat{H}_T(\hat{\phi}_j)$ are the gradient and Hessian of $Q_T(\hat{\phi}_{j+1})$ evaluated on $\hat{\phi}_j$, and $G(\phi_j)$ and $H(\phi_j)$ are their population counterparts. Using (61), the gradient of the population mapping on the $(j+1)^{th}$ iteration, defined as $V(\phi_j) = \nabla_{\phi_j} N(\phi_j)$, is given by:

$$V(\phi_j)_{\phi} = [\nabla_{\phi_j} N(\phi_j)]_{\phi} = I_3 - \left\{ \left[ I_1 \otimes [H(\phi_j)]^{-1} \right]_{\phi} \frac{\partial \text{vec}(G(\phi_j))}{\partial \phi} \right\}_{\phi} + 
+ \left[ G(\phi_j)' \otimes I_3 \right]_{\phi} \left\{ \frac{\partial \text{vec}([H(\phi_j)]^{-1})}{\partial \phi} \right\}_{\phi}$$ \hfill (63)

When evaluated at the true vector of parameters, the second term on the right-hand side of (63) is zero, following $[G(\phi_j)' \otimes I_3]_{\phi} = 0$. Hence, (63) reduces to:

$$V(\phi_j)_{\phi} = I - [H(\phi_j)]^{-1} \left[ \nabla_{\phi_j} G(\phi_j) \right]_{\phi}$$ \hfill (64)
The expressions for \( [H(\phi_j)]^{-1} \) and \([\nabla_{\phi_j}G(\phi_j)]\) are given by:

\[
[H(\phi_j)]^{-1} = \begin{pmatrix}
\frac{(-1+a)}{2} & \frac{-1+a-((1+a)^2)w}{(a+b)^2\sigma_w^2} & \frac{(-1+a)^2(1+a)^2w}{2(a+b)^2\sigma_w^2} & \frac{(-1+a)(1+a)^2(1+ab)w}{2(a+b)^2\sigma_w^2} \\
\frac{-1+a^2}{2(a+b)^2\sigma_w^2} & \frac{-1+a^2}{2(a+b)^2\sigma_w^2} & \frac{-1+a^2}{2(a+b)^2\sigma_w^2} & \frac{-1+a^2}{2(a+b)^2\sigma_w^2} \\
\frac{(-1+a)(1+a)^2(1+ab)w}{2(a+b)^2\sigma_w^2} & \frac{-1+a^2}{2(a+b)^2\sigma_w^2} & \frac{-1+a^2}{2(a+b)^2\sigma_w^2} & \frac{-1+a^2}{2(a+b)^2\sigma_w^2} \\
\frac{(-1+a^2)}{2(a+b)^2\sigma_w^2} & \frac{-1+a^2}{2(a+b)^2\sigma_w^2} & \frac{-1+a^2}{2(a+b)^2\sigma_w^2} & \frac{-1+a^2}{2(a+b)^2\sigma_w^2}
\end{pmatrix}
\]

(65)

\[
[\nabla_{\phi_j}G(\phi_j)] =
\begin{pmatrix}
\frac{2}{(-1+a)(1+b)} & \frac{2w}{(-1+a)^2(1+b)} & \frac{2w}{(-1+a)^2(1+b)} & \frac{2w}{(-1+a)^2(1+b)} \\
2w & -\frac{2(1+a)^3w^2-2(1-a^2+b+a(1+a-4a^2+a^4)b+a(4-5a+a^2)b^2+(-1+a)^2b)\sigma_w^2}{(-1+a)^2(1+b)} & 0 & \frac{-2(1+a+b)(1+ab)\sigma_w^2}{(-1+a)(1+a)^2(1+b)} \\
0 & 0 & \frac{2((-1+a)(1+a)^2+(-1+a+a^2)b+b^2)\sigma_w^2}{(-1+a)(1+a)^2(1+b)} & \frac{2(1+a+b)\sigma_w^2}{(1+a)(1+b)}
\end{pmatrix}
\]

(66)

Using results in (66) and (65) and collecting terms in (64), \( V(\phi) \) reduces to:

\[
V(\phi) = \begin{pmatrix}
\frac{a+b}{1+b} & \frac{(1+(-1+a+a^2)b)w}{1+b} & \frac{(-1+a)(1+a)^2w}{1+b} \\
0 & \frac{(1+2ab-a^2b)}{1+b} & \frac{(-1+a^2)^2}{1+b} \\
0 & \frac{-1+ab^2}{1+b} & \frac{(-1+a^2)(2+ab)}{1+b}
\end{pmatrix}
\]

(67)

Define the Eigenvalues associated with \( V(\phi) \) as \( \varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3)' \). By solving (67), \( \varepsilon \) is given by:

\[
\varepsilon = \begin{pmatrix}
a + b \\
1 + b \\
1 + b
\end{pmatrix}
\]

(68)

Remark: In Lemma 1, it is important to point out that the eigenvalues associated with (67) do not depend on \( w \) nor on \( \sigma_w^2 \). This allows to focus only with the parameters \( a, b \) which are bounded by Assumption B1. To this purpose, I evaluate the properties of (68) performing a numerical grid search through different combinations of parameters \( a \) and \( b \).
that satisfy Assumption B1. Figure 1 displays the maximum eigenvalue computed using (68). From Figure 1, the maximum eigenvalue associated with $V(\phi)$ is smaller than one, in absolute value, for all combinations in the grid search. This is enough to prove Lemma 1, yielding that $N(\xi)$ is an ACM for all $\xi \in B$.

**Lemma 2** Denote $\Psi_i = \varrho_i(\phi)$ and $\bar{\Psi}_i = \varrho_i(\bar{\phi})$, with $\phi, \bar{\phi} \in B$. Suppose Assumptions B1, B2 and B3 hold. Then,

$$
\mathbb{E}\left\{\left[ (\psi_{j+1,0} - \sum_{i=0}^{\infty} \psi_{j+1,0} u_{j,t-1-i}) - (\bar{\psi}_{j+1,0} - \sum_{i=0}^{\infty} \bar{\psi}_{j+1,0} u_{j,t-1-i}) \right]^2 \right\} \neq 0 \quad \text{for all} \quad \bar{\phi} \neq \phi
$$

**Proof of Lemma 2**: To prove Lemma 2, rewrite the target expression as:

$$
\mathbb{E}\left\{\left[ (\psi_0 - \bar{\psi}_0) - \left( \sum_{i=0}^{\infty} \bar{\psi}_i u_{t-1-i} - \sum_{i=0}^{\infty} \psi_i u_{t-1-i} \right) \right]^2 \right\} \neq 0
$$

$$
\mathbb{E}\left\{\left( \psi_0 - \bar{\psi}_0 \right)^2 - 2 \left( \psi_0 - \bar{\psi}_0 \right) \sum_{i=0}^{\infty} (\bar{\psi}_i - \psi_i) u_{t-1-i} + \left( \sum_{i=0}^{\infty} \bar{\psi}_i u_{t-1-i} - \sum_{i=0}^{\infty} \psi_i u_{t-1-i} \right)^2 \right\} \neq 0
$$

$$
\left( \psi_0 - \bar{\psi}_0 \right)^2 + \mathbb{E}\left\{ \sum_{i=0}^{\infty} \left( \bar{\psi}_i - \psi_i \right)^2 u_{t-1-i}^2 \right\} \neq 0
$$

$$
\left( \psi_0 - \bar{\psi}_0 \right)^2 + \mathbb{E}\left\{ \sum_{i=0}^{\infty} \left( \bar{\psi}_i - \psi_i \right)^2 u_{t-1-i}^2 + 2 \sum_{i=0}^{\infty} \left( \bar{\psi}_i - \psi_i \right) u_{t-1-i} \left[ \sum_{i=0}^{\infty} \left( \bar{\psi}_i - \psi_i \right) u_{t-1-i} \right] \right\} \neq 0
$$

$$
\left( \psi_0 - \bar{\psi}_0 \right)^2 + \sum_{i=0}^{\infty} \left( \bar{\psi}_i - \psi_i \right)^2 \sigma_u^2 \neq 0 \quad (69)
$$

Showing that (69) holds is equivalent to prove:

$$
|\psi_0 - \bar{\psi}_0| + \sigma_u^2 \sum_{i=0}^{\infty} |\bar{\psi}_i - \psi_i| > 0 \quad (70)
$$

I shall prove (70) by contradiction. To this purpose, I show that $\bar{\phi} = \phi$ is the only vector that sets (70) to zero. Define $\phi^*$ as vector located in the segment line between $\phi$ and $\bar{\phi}$. Using the first order Taylor expansion, the first term on the left-hand side of (70) reduces to:

$$
|\psi_0 - \bar{\psi}_0| = \left| \frac{\partial (\frac{\psi}{1-a})}{\partial \phi'} \right|_{\phi^*} |\phi - \bar{\phi}| = \left| \left[ \frac{1}{1-a^*}, \frac{\omega^*}{(1-a^*)^2}, 0 \right] \right| |\phi - \bar{\phi}| \quad (71)
$$
Assumption B1 guarantees that the first two elements of \( \frac{\partial \psi}{\partial \phi} \bigg|_{\phi^*} \) are strictly positive. Given that, (71) is equal to zero only if \( \tilde{\phi} = [\omega, a, \tilde{b}]' \), for any \( \tilde{b} \) satisfying Assumption B1. Hence it makes necessary to show that the second term on the left-hand side of (70) is greater than zero when evaluated at \( \tilde{\phi} = [\omega, a, \tilde{b}]' \). To this purpose, I apply the first order Taylor expansion such that:

\[
\sigma^2 \sum_{i=0}^{\infty} |\tilde{\psi}_i - \psi_i| = \sigma^2 \sum_{i=0}^{\infty} \left| \frac{\partial \psi_i}{\partial \phi} \bigg|_{\phi^*} \phi - \tilde{\phi} \right| = \sigma^2 \sum_{i=0}^{\infty} \left| [0, ia^* (a^* + b^*), a^* i] \phi - \tilde{\phi} \right| \quad (72)
\]

The third element of \( \frac{\partial \psi}{\partial \phi} \bigg|_{\phi^*} \) is strictly greater than zero for all \( i \geq 1 \) and \( b^* \) satisfying Assumption B1, implying that when evaluated on \( \tilde{\phi} = [\omega, a, \tilde{b}]' \) (72) is strictly greater than zero. Hence, the only vector that sets (70) to zero is \( \tilde{\phi} = \phi \). This concludes the proof of Lemma 2.

**Lemma 3** Suppose Assumptions B1, B2 and B3 hold. Then,

\[
\sup_{\xi \in \mathbb{R}} \left| \hat{N}_T (\xi) - N (\xi) \right| = o_p (1) \text{ as } T \to \infty
\]

**Proof of Lemma 3:** By evaluating both (61) and (62) on \( \phi_j \), the absolute difference between the population mapping and its sample counterpart is given by:

\[
\left| \hat{N}_T (\phi_j) - N (\phi_j) \right| = \left[ \phi_j - [H (\phi_j)]^{-1} G (\phi_j) \right] - \left[ \phi_j - \left[ \hat{H}_T (\phi_j) \right]^{-1} \hat{G}_T (\phi_j) \right] \quad (73)
\]

Subtracting and adding \( \left[ \hat{H}_T (\phi_j) \right]^{-1} G (\phi_j) \) in (73):

\[
\left| \hat{N}_T (\phi_j) - N (\phi_j) \right| \leq \left| \left[ \hat{H}_T (\phi_j) \right]^{-1} - [H (\phi_j)]^{-1} \right| G (\phi_j) \right| - \left| \left[ \hat{H}_T (\phi_j) \right]^{-1} \left[ \hat{G}_T (\phi_j) - G (\phi_j) \right] \right| \quad (74)
\]

To prove point-wise convergence of the population and sample mappings evaluated at the same vector of parameters, it suffices to show that both terms on the right-hand side of (74) have order \( o_p (1) \) as \( T \to \infty \). This implies showing that sample gradient and Hessian converge to their population counterparts, when evaluated on the true vector of
parameters $\phi$. The sample and population gradient are then given by:

$$\hat{G}_T(\phi) = \begin{pmatrix} \frac{1}{T} \sum_{t=1}^{T} \left\{ \frac{-2}{1-a} \dot{u}_t \right\} \\ \frac{1}{T} \sum_{t=1}^{T} \left\{ 2 \dot{u}_t \left( \frac{-\omega}{1-a} - \sum_{i=0}^{q} d_i u_{t-i} \right) \right\} \\ \frac{1}{T} \sum_{t=1}^{T} \left\{ -2 \left( \sum_{i=0}^{q} a^i u_{t-i} \right) \dot{u}_t \right\} \end{pmatrix}$$

(75)

and

$$G(\phi) = \begin{pmatrix} \mathbb{E} \left\{ \frac{-2}{1-a} \left( \dot{u}_t - \sum_{i=q+1}^{\infty} a^i (a+b) u_{t-1-i} \right) \right\} \\ \mathbb{E} \left\{ 2 \left( \dot{u}_t - \sum_{i=q+1}^{\infty} a^i (a+b) u_{t-1-i} \right) \left( -\frac{\omega}{1-a} - \sum_{i=0}^{q} d_i u_{t-1-i} - \sum_{i=q+1}^{\infty} d_i u_{t-1-i} \right) \right\} \\ \mathbb{E} \left\{ -2 \left( \sum_{i=0}^{q} a^i u_{t-1-i} - \sum_{i=q+1}^{\infty} a^i u_{t-1-i} \right) \dot{u}_t \right\} \end{pmatrix}$$

(76)

where $d_i = a^i + ia^{i-1} (a+b)$ and $\dot{u}_t = \epsilon_t^2 - \frac{\omega}{1-a} - \sum_{i=0}^{q} a^i (a+b) u_{t-1-i}$. Provided that $q \rightarrow \infty$ as $T \rightarrow \infty$, and $\sum_{i=0}^{\infty} |a^i| < \infty$ following $|a| < 1$, the additional terms in the population mapping converges in probability to zero as $T \rightarrow \infty$, such that $\sum_{i=q+1}^{\infty} a^i \overset{p}{\to} 0$ and $\sum_{i=q+1}^{\infty} d_i \overset{p}{\to} 0$. Furthermore, all elements in (75) are averages of m.d.s. processes. This allows the use of the weak law of large numbers, such that (78) holds. Similar steps are conducted to show that the sample Hessian converges in probability to its population counterpart, yielding (79).

$$\hat{G}_T(\phi_j) \overset{p}{\to} G(\phi_j)$$

(78)

$$\hat{H}_T(\phi_j) \overset{p}{\to} H(\phi_j)$$

(79)

To obtain uniform convergence in probability, the sample mapping needs to be stochastically equicontinuous. Assumption B3 provides the Lipschitz condition for $\tilde{N}_T(\phi_j)$ for all $\phi_j \in \mathbb{B}$. Following Lemma 2.9 in Newey and McFadden (1994), the Lipschitz condition implies that the sample mapping is stochastically equicontinuous, which allows the use of theorem 21.9 (pg. 337) in Davidson (1994), yielding uniform convergence between sample and population mappings.

**Lemma 4** Suppose Assumptions B1, B2 and B3 hold and fix $\xi$ and $\varsigma$ in $\mathbb{B}$. Then,

$$\sup_{\xi,\varsigma \in \mathbb{B}} \left| \hat{\Lambda}_T(\xi,\varsigma) - \Lambda(\xi,\varsigma) \right| = o_p(1) \text{ as } T \rightarrow \infty$$

37
Proof of Lemma 4: Proof of Lemma 4 mirrors the steps of Lemma 4 in Dias and Kapetanios (2012). Using their result, rewrite sup_{ξ,ς ∈ B} |Λ_T (ξ,ς) - Λ (ξ,ς)| = o_p (1) as:

\[
\sup_{ξ,ς ∈ B} \left| Λ (ξ,ς) - Λ_T (ξ,ς) \right| \leq \frac{1}{|ξ - ς|} \left[ \sup_{ξ,ς ∈ B} \left| N (ξ) - N_T (ξ) \right| + \sup_{ξ,ς ∈ B} \left| N (ς) - N_T (ς) \right| \right]
\]

(80)

Lemma 3 implies that both terms inside the brackets have order o_p (1). Assumption B1 states that [ξ - ς] is bounded, implying that the right-hand side of (80) converges in probability to zero, as T → ∞.

Lemma 5 Suppose Assumptions B1, B2 and B3 hold and fix ξ and ς in B. If

i) sup_{ξ ∈ B} |N_T (ξ) - N (ξ)| = o_p (1) as T → ∞

ii) sup_{ξ,ς ∈ B} |Λ_T (ξ,ς) - Λ (ξ,ς)| = o_p (1) as T → ∞

then, \(N_T (ξ)\) is an ACM on \((B, d)\), with \(ξ ∈ B\) and it has fixed point denoted by \(\hat{φ}\), such that \(|\hat{φ}_j - \hat{φ}| = o_p (1)\), as \(j → ∞\) with \(T → ∞\).

Proof of Lemma 5: see Lemma 5 in Dias and Kapetanios (2012).

Lemma 6 Suppose Assumptions B1, B2 and B3 hold. If \(N_T (γ)\) is an ACM on \((B, d)\) Then, \(\sqrt{T} |\hat{φ}_j - \hat{φ}| = o_p (1)\) as \(T → ∞\) and \(j → ∞\)

Proof of Lemma 6: Lemma 6 in Dias and Kapetanios (2012) gives:

\[
\sqrt{T} |\hat{φ}_j - \hat{φ}| \leq \sqrt{T}k^j |\hat{φ}_0 - \hat{φ}|
\]

(81)

The right-hand side converges in probability to zero if \(\frac{ln(T)}{j} = o (1)\). In fact, \(j \gg -\frac{1}{2} \frac{ln(T)}{ln(ε)}\) needs to hold, implying that speed of convergence depends on the contraction parameter of the population mapping.

Proof of Theorem 1: I divide this proof in two sections. In the first part, I prove the consistency of the NL-ILS estimator (item (i) in Theorem 1), whereas the second part focuses on the asymptotic distribution (part (ii) in Theorem 1). From Dominitz and Sherman (2005), if \(N (ξ)\) is an ACM on \((B, d)\), then \(N (ξ)\) is also a contraction map. Lemmas 1 and 5 state that the population and sample mapping are ACM. These allow the use of standard fixed-point theorem as stated in Burden and Faires (1993) and Judd (1998) to show consistency of the NL-ILS estimator. Identification on the population mapping gives \(N (φ) = φ\). To show that \(|\hat{φ} - φ| = o_p (1)\), rewrite this term:

\[
|\hat{φ} - φ| \leq |φ_j - φ| + |\hat{φ} - φ_j|
\]

(82)
The first term on the right-hand side can be expressed only as function of the population mapping. Rewriting it in this way and substituting recursively using the ACM bound, \(|\phi_j - \phi|\) resumes to

\[
|\phi_j - \phi| = |N(\phi_j) - N(\phi)| \leq \kappa |\phi_j - \phi|
\]

\[
|\phi_j - \phi| \leq \kappa |N(\phi_j) - N(\phi)| \leq \kappa^2 |\phi_j - \phi|
\]

\[
|\phi_j - \phi| \leq \kappa^j |N(\phi_j) - N(\phi)|
\]  \hspace{1cm} (83)

Provided that \( j \rightarrow \infty \) as \( T \rightarrow \infty \), the right-hand side of (83) converges in probability to zero. Hence, to show consistency of the NL-ILS estimator it remains to show that the second term on the right-hand side of (82) converges in probability to zero. Rewrite this term as:

\[
|\hat{\phi} - \phi_j| \leq |\hat{\phi} - \hat{\phi}_j| + |\hat{\phi}_j - \phi_j| \]  \hspace{1cm} (84)

The first term on the right-hand side of (84) has order \( o_p\left(T^{1/2}\right) \) following Lemma 6. The second term on the right-hand side of (84) is bounded as

\[
|\hat{\phi}_j - \phi_j| \leq |\hat{\phi}_j - \hat{\phi}_{j-1} - N(\hat{\phi}_{j-1}) - N(\phi_{j-1})| \]  \hspace{1cm} (85)

The first term on the right-hand side of (85) has order \( o_p(1) \) following Lemma 3. The remaining term of (85) can be rewritten using the ACM bound, such that:

\[
|N(\hat{\phi}_{j-1}) - N(\phi_{j-1})| \leq \kappa \left|\hat{\phi}_{j-1} - \phi_{j-1}\right| \]  \hspace{1cm} (86)

Applying recursively the same strategy as in (85) and (86), equation (84) reduces to

\[
|\hat{\phi} - \phi_j| \leq \kappa^j |\hat{\phi}_0 - \hat{\phi}_0| \]  \hspace{1cm} (87)

Note that \( |\hat{\phi}_0 - \phi_0| \) is bounded, provided that \( \hat{\phi}_0, \phi_0 \in \mathbb{B} \). As \( j \rightarrow \infty \) with \( T \rightarrow \infty \), the right-hand side of (87) has order \( o_p(1) \), implying \( |\hat{\phi} - \phi| = o_p(1) \).

I now prove the asymptotic distribution of the NL-ILS estimator. This proof mirrors the steps of Theorem 4 in Dominitz and Sherman (2005). To establish the asymptotic distribution of \( \sqrt{T} [\hat{\phi}_j - \phi] \), firstly rewrite it as:

\[
\sqrt{T} [\hat{\phi}_j - \phi] = \sqrt{T} [\hat{\phi}_j - \hat{\phi}] + \sqrt{T} [\hat{\phi} - \phi] \]  \hspace{1cm} (88)
The first term on the right-hand side of equation (88) has order $o_p(1)$ following Lemma 6 and provided that $\frac{h_0(T)}{T} = o(1)$. The second term of (88) resumes to

$$\sqrt{T} \left[ \hat{\phi} - \phi \right] = \sqrt{T} \left[ \hat{N}_T (\hat{\phi}) - N (\phi) \right] = \sqrt{T} \left[ \left[ \hat{N}_T (\hat{\phi}) - \hat{N}_T (\phi) \right] + \left[ \hat{N}_T (\phi) - \phi \right] \right]$$  \hspace{1cm} (89)

Define $\hat{\Lambda}_T (\hat{\phi}, \phi) = \int_0^1 \hat{\tilde{V}}_T \left( \hat{\phi} + \xi (\hat{\phi} - \phi) \right) d\xi$, such that the first term on the right-hand side of (89) is given by

$$\left[ \hat{N}_T (\hat{\phi}) - \hat{N}_T (\phi) \right] = \hat{\Lambda}_T (\hat{\phi}, \phi) \left[ \hat{\phi} - \phi \right]$$  \hspace{1cm} (90)

Plugging (90) into (89), the latter reduces to:

$$\sqrt{T} \left[ \hat{\phi} - \phi \right] = \sqrt{T} \left[ \hat{\Lambda}_T (\hat{\phi}, \phi) \left[ \hat{\phi} - \phi \right] \right] + \sqrt{T} \left[ \hat{N}_T (\phi) - \phi \right]$$

$$\sqrt{T} \left[ \hat{\phi} - \phi \right] = \sqrt{T} \left[ \left[ I_3 - \hat{\Lambda}_T (\hat{\phi}, \phi) \right]^{-1} \left[ \hat{N}_T (\phi) - \phi \right] \right]$$  \hspace{1cm} (91)

As in Dominitz and Sherman (2005), I initially show that $\hat{\Lambda}_T (\hat{\phi}, \phi) \xrightarrow{p} V (\phi)$. To this purpose, write $\hat{\Lambda}_T (\hat{\phi}, \phi)$ as

$$\hat{\Lambda}_T (\hat{\phi}, \phi) = V (\phi) + \left[ \Lambda \left( \hat{\phi}, \phi \right) - V (\phi) \right] + \left[ \hat{\Lambda}_T (\hat{\phi}, \phi) - \Lambda \left( \hat{\phi}, \phi \right) \right]$$  \hspace{1cm} (92)

Item i in Theorem 1 states that $\hat{\phi}$ converges in probability to $\phi$ as $j \xrightarrow{} \infty$ with $T \xrightarrow{} \infty$. This implies that $\Lambda \left( \hat{\phi}, \phi \right) \xrightarrow{p} V (\phi)$, yielding that the second term on the right-hand of (92) converges in probability to zero. Lemma 4 implies that the third term on the right-hand side of (92) has order $o_p(1)$. Hence, (91) reduces to

$$\sqrt{T} \left[ \hat{\phi} - \phi \right] = \sqrt{T} \left[ I_3 - V (\phi) \right]^{-1} \left[ \hat{N}_T (\phi) - \phi \right]$$  \hspace{1cm} (93)

It remains to study the asymptotic distribution of $\sqrt{T} \left[ \hat{N}_T (\phi) - \phi \right]$. Note that, when $T \xrightarrow{} \infty$ and $N_T (\cdot)$ is evaluated on the true vector of parameter, the sample mapping reduces, asymptotically, to the case where the latent variable becomes observed regressors. Given that, the asymptotic distribution of $\sqrt{T} \left[ \hat{N}_T (\phi) - \phi \right]$ reduces to the asymptotic distribution of the NL-LS estimator. As in Greene (2008), the asymptotic variance of the NL-LS estimator is given by $\sigma_C^2 C^{-1}$, where $C_0 = \text{plim} \frac{1}{T} \sum_{t=1}^T \left[ \frac{\partial h_t (\theta)}{\partial \theta} \frac{\partial h_t (\theta)}{\partial \theta} \right]$ and $h_t (\theta)$ is the nonlinear function in $Q_T (y_t, x_t; \theta) = (y_t - h_t (x_t, \theta))^2$. Considering the sample mapping of the NL-ILS estimator, the function $h_t (x_t, \phi)$ is given by the MA($q$), such
that \( h_t(x_t, \phi) = \psi_0 + \sum_{i=0}^{q} \psi_i x_{t-i} \). Given that, \( C_0 \) resumes to

\[
C_0 = \begin{pmatrix}
\frac{1}{(1-a)^2} & -\frac{\omega}{(1-a)^3} & 0 \\
-\frac{\omega}{(1-a)^3} & \frac{\omega^2}{(1-a)^4} + \sum_{i=0}^{\bar{q}} d_i^2 \sigma_u^2 & \sum_{i=0}^{\bar{q}} d_i a_i \sigma_u^2 \\
0 & \sum_{i=0}^{\bar{q}} d_i a_i \sigma_u^2 & \sum_{i=0}^{\bar{q}} a_i^2 \sigma_u^2
\end{pmatrix}
\] (94)

where \( d_i = ia^{i-1} (a + b) + a^i \). Applying the central limit theorem for martingale difference sequences, the asymptotic distribution of \( \sqrt{T} \left[ \hat{N}_T(\phi) - \phi \right] \) is given by

\[
\sqrt{T} \left[ \hat{N}_T(\phi) - \phi \right] \xrightarrow{d} \mathcal{N}(0, \sigma_u^2 C_0^{-1})
\] (95)

Equation 67 gives the analytical solution of \( V(\phi) \). Define \( A = [I - V(\phi)]^{-1} \), then the asymptotic distribution of the NL-ILS is given by

\[
\sqrt{T} \left[ \hat{\phi} - \phi \right] \xrightarrow{d} \mathcal{N}(0, \sigma_u^2 A C_0^{-1} A')
\] (96)

Proof of Corollary 1: This proof follows the item (i) in Theorem 1. Note that Lemma 1 holds because \( u_t \) has zero mean, finite variance and autocovariance equal to zero for all lags greater than zero. It is relevant to discuss the validity of Lemmas 4 and 3. Both of them are based on the weak law of large numbers. Note that, from item (i) in corollary 1, the \( u_t \) is a linear projection with \( \text{Cov}(u_{t-i}, u_{t-j}) = 0 \) for all \( i \neq j \). This is sufficient to allow the use of the weak law of large numbers as stated in Hamilton (1994) - pg. 186, implying that Lemma 3 holds. If Lemma 3 holds, then Lemma 4 also holds, extending the validity of item (i) in Theorem 1 to this Corollary.

Proof of Proposition 1: The regressors in (26) do not depend on \( \hat{\lambda}_{j+1} \). This implies that the first derivative of the sample mapping with respect to \( \hat{\lambda}_{j+1} \) is

\[
4 \frac{1}{T} \sum_{t=1}^{T} \left\{ \left[ y_t - \hat{\lambda}_{j+1} \hat{\sigma}_{j,t} \right] - \hat{\psi}_{j+1,0} - \sum_{i=0}^{\bar{q}} \hat{\psi}_{j+1,i} \hat{u}_{j,t-i} \right\}^2 \left[ y_t - \hat{\lambda}_{j+1} \hat{\sigma}_{j,t} \right] \hat{\sigma}_{j,t} = 0
\]

\[
\frac{1}{T} \sum_{t=1}^{T} \left\{ y_t - \hat{\lambda}_{j+1} \hat{\sigma}_{j,t} \right\} \hat{\sigma}_{j,t} = 0
\] (97)
By manipulating (97), $\hat{\lambda}_{j+1}$ resumes to:

$$\hat{\lambda}_{j+1} = \left[ \sum_{t=1}^{T} \hat{\sigma}_{j,t}^{2} \right]^{-1} \sum_{t=1}^{T} \hat{\sigma}_{j,t} y_{t}$$ (98)

The remaining first order conditions do not have a closed solution, implying that $\hat{\phi}_{j+1}$ has to be recovers through optimization. This concludes the proof of Proposition 1.
Figure 1: GARCH(1,1): ACM property

Figure 1 plots the highest element of $|\varepsilon|$ in (68) using different combinations of $\alpha$ and $\beta$, such that Assumption B1 is satisfied. The grid is fixed in 0.001.

Figure 2: GARCH(1,1)-in-mean: ACM property

Figure 2 displays the maximum eigenvalue computed from the numerical gradient of the NL-ILS mapping.
Table 1: GARCH(1,1)

<table>
<thead>
<tr>
<th></th>
<th>T = 100</th>
<th></th>
<th>T = 200</th>
<th></th>
<th>T = 300</th>
<th></th>
<th>T = 500</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NL-ILS</td>
<td>MLE</td>
<td>NL-ILS</td>
<td>MLE</td>
<td>NL-ILS</td>
<td>MLE</td>
<td>NL-ILS</td>
<td>MLE</td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td>-------</td>
<td>--------------</td>
<td>-------</td>
<td>--------------</td>
<td>-------</td>
<td>--------------</td>
<td>-------</td>
</tr>
<tr>
<td>ω</td>
<td>0.010</td>
<td>0.002</td>
<td>0.058</td>
<td>0.391</td>
<td>0.002</td>
<td>0.048</td>
<td>0.003</td>
<td>0.599</td>
</tr>
<tr>
<td></td>
<td>0.002</td>
<td>0.048</td>
<td>0.611</td>
<td>0.461</td>
<td>0.039</td>
<td>0.599</td>
<td>0.003</td>
<td>0.733</td>
</tr>
<tr>
<td>α</td>
<td>0.026</td>
<td>0.033</td>
<td>0.026</td>
<td>0.684</td>
<td>0.023</td>
<td>0.760</td>
<td>0.021</td>
<td>0.910</td>
</tr>
<tr>
<td></td>
<td>0.038</td>
<td>0.684</td>
<td>0.737</td>
<td>0.760</td>
<td>0.026</td>
<td>0.910</td>
<td>0.030</td>
<td>0.977</td>
</tr>
<tr>
<td>β</td>
<td>0.970</td>
<td>0.961</td>
<td>0.908</td>
<td>0.535</td>
<td>0.974</td>
<td>0.950</td>
<td>0.976</td>
<td>0.958</td>
</tr>
<tr>
<td></td>
<td>0.952</td>
<td>0.950</td>
<td>0.976</td>
<td>0.958</td>
<td>0.976</td>
<td>0.958</td>
<td>0.976</td>
<td>0.958</td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td>-------</td>
<td>--------------</td>
<td>-------</td>
<td>--------------</td>
<td>-------</td>
<td>--------------</td>
<td>-------</td>
</tr>
</tbody>
</table>

NL-ILS and MLE account for results obtained using the NL-ILS and MLE algorithms. Results reported in terms of median and the relative root mean squared error (RRMSE) for the conditional variance parameters. RRMSE is computed using the RMSE from the QMLE estimates on the denominator and the RMSE of the NL-ILS estimates on the numerator. Truncation parameter is fixed to $q = 3/\sqrt{T}$. I perform 5000 replications. Replications that do not achieve convergence are discarded when computing the relative measures.
Table 2: weak-GARCH(1,1): Specification 1

<table>
<thead>
<tr>
<th>m=2, T = 500</th>
<th>m=3, T = 333</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Freq.</td>
<td>Low freq.</td>
</tr>
<tr>
<td>ω</td>
<td>0.001</td>
</tr>
<tr>
<td>α</td>
<td>0.050</td>
</tr>
<tr>
<td>β</td>
<td>0.940</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>m=4, T = 250</th>
<th>m=5, T = 200</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Freq.</td>
<td>Low freq.</td>
</tr>
<tr>
<td>ω</td>
<td>0.001</td>
</tr>
<tr>
<td>α</td>
<td>0.050</td>
</tr>
<tr>
<td>β</td>
<td>0.940</td>
</tr>
</tbody>
</table>

NL-ILS and MLE account for results obtained using the NL-ILS and MLE algorithms. Results reported in terms of the median and the relative root mean squared error (RRMSE) for the conditional variance parameters. RRMSE is computed using the RMSE from the QMLE estimates on the denominator and the RMSE of the NL-ILS estimates on the numerator. Truncation parameter is fixed to $\hat{q} = 3 \sqrt{T}$. I perform 1500 replications. Replications that do not achieve convergence are discarded when computing the relative measures. The variable $m$ denotes sampling frequency.

Table 3: weak-GARCH(1,1): Specification 2

<table>
<thead>
<tr>
<th>m=2, T = 500</th>
<th>m=3, T = 333</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Freq.</td>
<td>Low freq.</td>
</tr>
<tr>
<td>ω</td>
<td>0.001</td>
</tr>
<tr>
<td>α</td>
<td>0.020</td>
</tr>
<tr>
<td>β</td>
<td>0.970</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>m=4, T = 250</th>
<th>m=5, T = 200</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Freq.</td>
<td>Low freq.</td>
</tr>
<tr>
<td>ω</td>
<td>0.001</td>
</tr>
<tr>
<td>α</td>
<td>0.020</td>
</tr>
<tr>
<td>β</td>
<td>0.970</td>
</tr>
</tbody>
</table>

NL-ILS and MLE account for results obtained using the NL-ILS and MLE algorithms. Results reported in terms of the median and the relative root mean squared error (RRMSE) for the conditional variance parameters. RRMSE is computed using the RMSE from the QMLE estimates on the denominator and the RMSE of the NL-ILS estimates on the numerator. Truncation parameter is fixed to $\hat{q} = 3 \sqrt{T}$. I perform 1500 replications. Replications that do not achieve convergence are discarded when computing the relative measures. The variable $m$ denotes sampling frequency.
Table 4: GARCH(1,1)-in-mean

<table>
<thead>
<tr>
<th></th>
<th>$T = 100$</th>
<th>$T = 200$</th>
<th>$T = 300$</th>
<th>$T = 400$</th>
<th>$T = 500$</th>
<th>$T = 750$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda = 0.1$</td>
<td>1.25</td>
<td>0.96</td>
<td>1.11</td>
<td>1.00</td>
<td>1.11</td>
<td>0.99</td>
</tr>
<tr>
<td>$\omega = 0.01$</td>
<td>0.16</td>
<td>2.54</td>
<td>0.18</td>
<td>1.23</td>
<td>0.23</td>
<td>1.12</td>
</tr>
<tr>
<td>$\alpha = 0.025$</td>
<td>0.85</td>
<td>1.48</td>
<td>0.68</td>
<td>0.99</td>
<td>0.87</td>
<td>0.97</td>
</tr>
<tr>
<td>$\beta = 0.97$</td>
<td>0.60</td>
<td>1.83</td>
<td>0.63</td>
<td>1.28</td>
<td>0.70</td>
<td>1.13</td>
</tr>
<tr>
<td>min $\tilde{\pi}_{t+h}$</td>
<td>1.23</td>
<td>1.00</td>
<td>1.09</td>
<td>1.00</td>
<td>1.09</td>
<td>0.99</td>
</tr>
<tr>
<td>med $\tilde{\pi}_{t+h}$</td>
<td>1.25</td>
<td>1.01</td>
<td>1.10</td>
<td>1.00</td>
<td>1.10</td>
<td>0.99</td>
</tr>
<tr>
<td>max $\tilde{\pi}_{t+h}$</td>
<td>1.25</td>
<td>1.03</td>
<td>1.11</td>
<td>1.00</td>
<td>1.11</td>
<td>1.00</td>
</tr>
<tr>
<td>min $\hat{\sigma}^2_{t+h}$</td>
<td>0.48</td>
<td>0.70</td>
<td>0.72</td>
<td>0.95</td>
<td>0.87</td>
<td>0.97</td>
</tr>
<tr>
<td>med $\hat{\sigma}^2_{t+h}$</td>
<td>0.52</td>
<td>0.74</td>
<td>0.78</td>
<td>0.97</td>
<td>0.90</td>
<td>0.98</td>
</tr>
<tr>
<td>max $\hat{\sigma}^2_{t+h}$</td>
<td>0.57</td>
<td>0.76</td>
<td>0.81</td>
<td>0.98</td>
<td>0.93</td>
<td>0.99</td>
</tr>
</tbody>
</table>

NL-ILS and MLE account for results obtained using the NL-ILS and MLE algorithms. MLE* accounts for results obtained using the MLE estimator computed using NL-ILS estimates as the initial values. Results for the GARCH(1,1)-in-mean parameters and in-sample conditional variance are reported in terms of the relative root median squared error (RMedSE). Forecast accuracy is accessed through the RRMedSFE (relative root median squared forecast error). Relative measures are computed with respect to the MLE benchmark. Relative measures less than one imply NL-ILS estimator outperforms the MLE methodology. Truncation parameter is fixed to $\bar{q} = 3 \sqrt{T}$. I perform 1500 replications. Replications that do not achieve convergence are discarded for computing the relative measures.
<table>
<thead>
<tr>
<th></th>
<th>T = 100</th>
<th>T = 200</th>
<th>T = 300</th>
<th>T = 400</th>
<th>T = 500</th>
<th>T = 750</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NL-ILS</td>
<td>MLE*</td>
<td>NL-ILS</td>
<td>MLE*</td>
<td>NL-ILS</td>
<td>MLE*</td>
</tr>
<tr>
<td>( \lambda = 0.1 )</td>
<td>1.14</td>
<td>0.95</td>
<td>1.06</td>
<td>1.00</td>
<td>1.10</td>
<td>1.02</td>
</tr>
<tr>
<td>( \omega = 0.01 )</td>
<td>0.34</td>
<td>1.12</td>
<td>0.75</td>
<td>1.12</td>
<td>1.21</td>
<td>1.11</td>
</tr>
<tr>
<td>( \alpha = 0.08 )</td>
<td>0.88</td>
<td>0.93</td>
<td>1.22</td>
<td>1.03</td>
<td>1.39</td>
<td>1.03</td>
</tr>
<tr>
<td>( \beta = 0.90 )</td>
<td>0.80</td>
<td>1.01</td>
<td>1.37</td>
<td>1.07</td>
<td>1.77</td>
<td>1.11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>T = 100</th>
<th>T = 200</th>
<th>T = 300</th>
<th>T = 400</th>
<th>T = 500</th>
<th>T = 750</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NL-ILS</td>
<td>MLE*</td>
<td>NL-ILS</td>
<td>MLE*</td>
<td>NL-ILS</td>
<td>MLE*</td>
</tr>
<tr>
<td>( \lambda = 0.1 )</td>
<td>1.14</td>
<td>0.95</td>
<td>1.06</td>
<td>1.00</td>
<td>1.10</td>
<td>1.02</td>
</tr>
<tr>
<td>( \omega = 0.01 )</td>
<td>0.34</td>
<td>1.12</td>
<td>0.75</td>
<td>1.12</td>
<td>1.21</td>
<td>1.11</td>
</tr>
<tr>
<td>( \alpha = 0.08 )</td>
<td>0.88</td>
<td>0.93</td>
<td>1.22</td>
<td>1.03</td>
<td>1.39</td>
<td>1.03</td>
</tr>
<tr>
<td>( \beta = 0.90 )</td>
<td>0.80</td>
<td>1.01</td>
<td>1.37</td>
<td>1.07</td>
<td>1.77</td>
<td>1.11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>T = 100</th>
<th>T = 200</th>
<th>T = 300</th>
<th>T = 400</th>
<th>T = 500</th>
<th>T = 750</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NL-ILS</td>
<td>MLE*</td>
<td>NL-ILS</td>
<td>MLE*</td>
<td>NL-ILS</td>
<td>MLE*</td>
</tr>
<tr>
<td>( \lambda = 0.1 )</td>
<td>1.10</td>
<td>1.02</td>
<td>1.15</td>
<td>1.00</td>
<td>1.10</td>
<td>1.00</td>
</tr>
<tr>
<td>( \omega = 0.01 )</td>
<td>0.89</td>
<td>1.31</td>
<td>1.53</td>
<td>1.16</td>
<td>1.59</td>
<td>1.19</td>
</tr>
<tr>
<td>( \alpha = 0.08 )</td>
<td>0.94</td>
<td>1.00</td>
<td>1.09</td>
<td>1.06</td>
<td>1.12</td>
<td>1.03</td>
</tr>
<tr>
<td>( \beta = 0.875 )</td>
<td>0.98</td>
<td>1.12</td>
<td>1.51</td>
<td>1.16</td>
<td>1.67</td>
<td>1.20</td>
</tr>
</tbody>
</table>

Table 5: GARCH(1,1)-in-mean

NL-ILS and MLE account for results obtained using the NL-ILS and MLE algorithms. MLE* accounts for results obtained using the MLE estimator computed using NL-ILS estimates as the initial values. Results for the GARCH(1,1)-in-mean parameters and in-sample conditional variance are reported in terms of the relative root median squared error (RMedSE). Forecast accuracy is assessed through the RRMedSFE (relative root median squared forecast error). Relative measures are computed with respect to the MLE benchmark. Relative measures less than one imply NL-ILS estimator outperforms the MLE methodology. Truncation parameter is fixed to \( \bar{q} = \frac{3}{\sqrt{T}} \). I perform 1500 replications. Replications that do not achieve convergence are discarded for computing the relative measures.
Table 6: RealGARCH(1,1)-in-mean

<table>
<thead>
<tr>
<th></th>
<th>T = 100</th>
<th>T = 200</th>
<th>T = 300</th>
<th>T = 400</th>
<th>T = 500</th>
<th>T = 2000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NL-ILS</td>
<td>NL-ILS</td>
<td>NL-ILS</td>
<td>NL-ILS</td>
<td>NL-ILS</td>
<td>NL-ILS</td>
</tr>
<tr>
<td>( \lambda = 0.1 )</td>
<td>1.21</td>
<td>1.41</td>
<td>1.63</td>
<td>1.71</td>
<td>1.81</td>
<td>2.35</td>
</tr>
<tr>
<td></td>
<td>(0.089)</td>
<td>(0.105)</td>
<td>(0.104)</td>
<td>(0.103)</td>
<td>(0.104)</td>
<td>(0.102)</td>
</tr>
<tr>
<td>( \omega = 0.06 )</td>
<td>1.74</td>
<td>1.89</td>
<td>1.95</td>
<td>2.00</td>
<td>2.01</td>
<td>2.27</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.05)</td>
<td>(0.049)</td>
<td>(0.05)</td>
<td>(0.053)</td>
<td>(0.058)</td>
</tr>
<tr>
<td>( \beta = 0.45 )</td>
<td>1.65</td>
<td>2.45</td>
<td>2.83</td>
<td>3.30</td>
<td>3.70</td>
<td>5.68</td>
</tr>
<tr>
<td></td>
<td>(0.393)</td>
<td>(0.423)</td>
<td>(0.442)</td>
<td>(0.442)</td>
<td>(0.444)</td>
<td>(0.458)</td>
</tr>
<tr>
<td>( \gamma = 0.51 )</td>
<td>1.80</td>
<td>2.63</td>
<td>3.10</td>
<td>3.49</td>
<td>3.84</td>
<td>5.39</td>
</tr>
<tr>
<td></td>
<td>(0.519)</td>
<td>(0.506)</td>
<td>(0.496)</td>
<td>(0.501)</td>
<td>(0.505)</td>
<td>(0.501)</td>
</tr>
<tr>
<td>( \xi = -0.18 )</td>
<td>1.78</td>
<td>2.03</td>
<td>2.10</td>
<td>2.19</td>
<td>2.21</td>
<td>2.08</td>
</tr>
<tr>
<td></td>
<td>(−0.318)</td>
<td>(−0.227)</td>
<td>(−0.22)</td>
<td>(−0.207)</td>
<td>(−0.209)</td>
<td>(−0.189)</td>
</tr>
<tr>
<td>( \varphi = 1.04 )</td>
<td>2.01</td>
<td>2.40</td>
<td>2.54</td>
<td>2.60</td>
<td>2.65</td>
<td>2.51</td>
</tr>
<tr>
<td></td>
<td>(0.942)</td>
<td>(1.012)</td>
<td>(1.017)</td>
<td>(1.021)</td>
<td>(1.027)</td>
<td>(1.038)</td>
</tr>
<tr>
<td>( \tau_1 = -0.11 )</td>
<td>1.03</td>
<td>1.03</td>
<td>1.07</td>
<td>1.06</td>
<td>1.10</td>
<td>1.37</td>
</tr>
<tr>
<td></td>
<td>(−0.119)</td>
<td>(−0.118)</td>
<td>(−0.121)</td>
<td>(−0.122)</td>
<td>(−0.121)</td>
<td>(−0.122)</td>
</tr>
<tr>
<td>( \tau_2 = 0.07 )</td>
<td>1.03</td>
<td>1.11</td>
<td>1.16</td>
<td>1.14</td>
<td>1.21</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>(0.073)</td>
<td>(0.073)</td>
<td>(0.076)</td>
<td>(0.074)</td>
<td>(0.076)</td>
<td>(0.073)</td>
</tr>
</tbody>
</table>

|                | \( \hat{\sigma}^2_t \) | 1.00 | 0.94 | 0.96 | 0.95 | 0.96 | 1.00 |
|                | \( \min \hat{\pi}_{t+h} \) | 1.10 | 1.20 | 1.24 | 1.27 | 1.28 | 1.27 |
|                | \( \text{med } \hat{\pi}_{t+h} \) | 1.14 | 1.26 | 1.34 | 1.34 | 1.36 | 1.33 |
|                | \( \max \hat{\pi}_{t+h} \) | 1.19 | 1.35 | 1.45 | 1.45 | 1.50 | 1.45 |
|                | \( \min \hat{\sigma}^2_{t+h} \) | 0.98 | 1.01 | 1.00 | 1.01 | 1.01 | 0.99 |
|                | \( \text{med } \hat{\sigma}^2_{t+h} \) | 1.02 | 1.04 | 1.01 | 1.02 | 1.03 | 1.00 |
|                | \( \max \hat{\sigma}^2_{t+h} \) | 1.09 | 1.07 | 1.03 | 1.05 | 1.05 | 1.01 |
|                | \( \min \hat{\nu}_{t+h} \) | 0.92 | 0.97 | 0.96 | 0.99 | 0.98 | 0.98 |
|                | \( \text{med } \hat{\nu}_{t+h} \) | 0.97 | 0.99 | 0.97 | 1.00 | 1.00 | 0.99 |
|                | \( \max \hat{\nu}_{t+h} \) | 1.00 | 1.01 | 0.99 | 1.02 | 1.01 | 1.00 |

Table 6 reports the results obtained using the NL-ILS a estimator. Results for the RealGARCH(1,1)-in-mean parameters and in-sample conditional variance are reported in terms of the Relative root Mean Squared Error (RRMedSFE). Values inside the brackets refer to the median computed within all valid replications. Forecast accuracy is accessed through the RMSE. Relative measures are computed with respect to the MLE benchmark. Relative measures less than one imply NL-ILS estimator outperforms the MLE methodology. Truncation parameter is fixed to \( q = 3 \sqrt{T} \). I perform 1500 replications. Replications that do not achieve convergence are discarded for computing the relative measures.
<table>
<thead>
<tr>
<th></th>
<th>T = 500</th>
<th>T = 750</th>
<th>T = 1000</th>
<th>T = 1250</th>
<th>T = 1500</th>
<th>T = 1750</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>APARCH(1,1)-in-mean</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda = 0.20$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\min \hat{\pi}_{t+h}$</td>
<td>1.12</td>
<td>0.99</td>
<td>1.13</td>
<td>0.98</td>
<td>1.00</td>
<td>0.99</td>
</tr>
<tr>
<td>$\text{med } \hat{\pi}_{t+h}$</td>
<td>1.16</td>
<td>1.00</td>
<td>1.19</td>
<td>0.98</td>
<td>1.11</td>
<td>0.99</td>
</tr>
<tr>
<td>$\max \hat{\pi}_{t+h}$</td>
<td>1.24</td>
<td>1.00</td>
<td>1.31</td>
<td>0.99</td>
<td>1.28</td>
<td>1.02</td>
</tr>
<tr>
<td>$\min \hat{\sigma}^2_{t+h}$</td>
<td>1.29</td>
<td>0.96</td>
<td>1.32</td>
<td>0.93</td>
<td>1.35</td>
<td>0.98</td>
</tr>
<tr>
<td>$\text{med } \hat{\sigma}^2_{t+h}$</td>
<td>1.52</td>
<td>0.96</td>
<td>1.61</td>
<td>0.94</td>
<td>1.65</td>
<td>1.00</td>
</tr>
<tr>
<td>$\max \hat{\sigma}^2_{t+h}$</td>
<td>1.73</td>
<td>0.97</td>
<td>1.90</td>
<td>0.95</td>
<td>1.96</td>
<td>1.03</td>
</tr>
<tr>
<td><strong>EGARCH(1,1,1)-in-mean</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda = 0.20$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\min \hat{\pi}_{t+h}$</td>
<td>0.88</td>
<td>1.00</td>
<td>0.83</td>
<td>1.01</td>
<td>0.76</td>
<td>0.98</td>
</tr>
<tr>
<td>$\text{med } \hat{\pi}_{t+h}$</td>
<td>0.88</td>
<td>1.01</td>
<td>0.83</td>
<td>1.01</td>
<td>0.78</td>
<td>0.98</td>
</tr>
<tr>
<td>$\max \hat{\pi}_{t+h}$</td>
<td>0.89</td>
<td>1.01</td>
<td>0.84</td>
<td>1.01</td>
<td>0.80</td>
<td>0.99</td>
</tr>
<tr>
<td>$\min \hat{\sigma}^2_{t+h}$</td>
<td>1.18</td>
<td>1.00</td>
<td>1.11</td>
<td>1.00</td>
<td>1.11</td>
<td>0.99</td>
</tr>
<tr>
<td>$\text{med } \hat{\sigma}^2_{t+h}$</td>
<td>1.27</td>
<td>1.01</td>
<td>1.20</td>
<td>1.00</td>
<td>1.20</td>
<td>0.99</td>
</tr>
<tr>
<td>$\max \hat{\sigma}^2_{t+h}$</td>
<td>1.33</td>
<td>1.01</td>
<td>1.26</td>
<td>1.00</td>
<td>1.26</td>
<td>0.99</td>
</tr>
</tbody>
</table>

NL-ILS accounts for results obtained using the NL-ILS algorithm. MLE* accounts for results obtained using the MLE estimator computed using NL-ILS estimates as the initial values. Results for the $\lambda$ parameter are reported in terms of the Relative Mean Squared Error (RMSE). Values inside brackets refer to the bias obtained for the $\lambda$ estimates. Forecast accuracy is accessed through the RRMSFE (relative root mean squared forecast error). Relative measures are computed with respect to the MLE benchmark. Relative measures less than one imply NL-ILS estimator outperforms the MLE methodology. Truncation parameter is fixed to $\bar{q} = 3 \sqrt{T}$. I perform 1500 replications. Replications that do not achieve convergence are discarded for computing the relative measures.
Table 8: Robustness analysis: conditional variance misspecification

<table>
<thead>
<tr>
<th>GJR GARCH(1,1,1)-in-mean</th>
<th>( T = 500 )</th>
<th>( T = 750 )</th>
<th>( T = 1000 )</th>
<th>( T = 1250 )</th>
<th>( T = 1500 )</th>
<th>( T = 1750 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda = 0.20 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\pi}_{t+h} ) min</td>
<td>0.98</td>
<td>1.00</td>
<td>0.96</td>
<td>1.00</td>
<td>0.93</td>
<td>1.00</td>
</tr>
<tr>
<td>( \hat{\pi}_{t+h} ) med</td>
<td>1.02</td>
<td>1.00</td>
<td>1.01</td>
<td>1.00</td>
<td>0.91</td>
<td>1.00</td>
</tr>
<tr>
<td>( \hat{\pi}_{t+h} ) max</td>
<td>1.05</td>
<td>1.00</td>
<td>1.04</td>
<td>1.00</td>
<td>1.01</td>
<td>1.00</td>
</tr>
<tr>
<td>( \hat{\sigma}^2_{t+h} ) min</td>
<td>1.06</td>
<td>0.99</td>
<td>1.06</td>
<td>1.00</td>
<td>1.04</td>
<td>1.00</td>
</tr>
<tr>
<td>( \hat{\sigma}^2_{t+h} ) med</td>
<td>1.12</td>
<td>1.00</td>
<td>1.13</td>
<td>1.00</td>
<td>1.09</td>
<td>1.00</td>
</tr>
<tr>
<td>( \hat{\sigma}^2_{t+h} ) max</td>
<td>1.28</td>
<td>1.00</td>
<td>1.30</td>
<td>1.00</td>
<td>1.22</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GARCH(2,2)-in-mean</th>
<th>( T = 500 )</th>
<th>( T = 750 )</th>
<th>( T = 1000 )</th>
<th>( T = 1250 )</th>
<th>( T = 1500 )</th>
<th>( T = 1750 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda = 0.20 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\pi}_{t+h} ) min</td>
<td>1.13</td>
<td>1.00</td>
<td>1.13</td>
<td>1.00</td>
<td>1.15</td>
<td>1.01</td>
</tr>
<tr>
<td>( \hat{\pi}_{t+h} ) med</td>
<td>1.16</td>
<td>1.00</td>
<td>1.14</td>
<td>1.00</td>
<td>1.17</td>
<td>1.01</td>
</tr>
<tr>
<td>( \hat{\pi}_{t+h} ) max</td>
<td>1.20</td>
<td>1.00</td>
<td>1.19</td>
<td>1.00</td>
<td>1.25</td>
<td>1.01</td>
</tr>
<tr>
<td>( \hat{\sigma}^2_{t+h} ) min</td>
<td>1.17</td>
<td>1.00</td>
<td>1.13</td>
<td>0.99</td>
<td>1.15</td>
<td>0.99</td>
</tr>
<tr>
<td>( \hat{\sigma}^2_{t+h} ) med</td>
<td>1.27</td>
<td>1.01</td>
<td>1.25</td>
<td>1.00</td>
<td>1.27</td>
<td>1.00</td>
</tr>
<tr>
<td>( \hat{\sigma}^2_{t+h} ) max</td>
<td>1.48</td>
<td>1.01</td>
<td>1.52</td>
<td>1.01</td>
<td>1.68</td>
<td>1.00</td>
</tr>
</tbody>
</table>

NL-ILS accounts for results obtained using the NL-ILS algorithm. MLE* accounts for results obtained using the MLE estimator computed using NL-ILS estimates as the initial values. Results for the \( \lambda \) parameter are reported in terms of the relative root mean squared error (RMSE). Values inside brackets refer to the bias obtained for the \( \lambda \) estimates. Forecast accuracy is assessed through the RRMSFE (relative root mean squared forecast error). Relative measures are computed with respect to the MLE benchmark. Relative measures less than one imply NL-ILS estimator outperforms the MLE methodology. Truncation parameter is fixed to \( \bar{q} = 3 \sqrt{T} \). I perform 1500 replications. Replications that do not achieve convergence are discarded for computing the relative measures.
Table 9: Descriptive statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Std. Dev.</th>
<th>Kurtosis</th>
<th>N. Obs</th>
<th>Start Date</th>
<th>End Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRSP†</td>
<td>0.0002</td>
<td>0.0005</td>
<td>0.0099</td>
<td>19.7</td>
<td>12,148</td>
<td>28/06/1963</td>
<td>29/09/2011</td>
</tr>
<tr>
<td>S&amp;P500†</td>
<td>0.0000</td>
<td>0.0002</td>
<td>0.0104</td>
<td>31.1</td>
<td>12,148</td>
<td>28/06/1963</td>
<td>29/09/2011</td>
</tr>
<tr>
<td>S&amp;P100†</td>
<td>0.0001</td>
<td>0.0004</td>
<td>0.0121</td>
<td>30.5</td>
<td>7,364</td>
<td>04/08/1982</td>
<td>29/09/2011</td>
</tr>
<tr>
<td>CRSP‡</td>
<td>0.0010</td>
<td>0.0026</td>
<td>0.0227</td>
<td>9.0</td>
<td>2,426</td>
<td>05/07/1963</td>
<td>30/09/2011</td>
</tr>
<tr>
<td>S&amp;P500‡</td>
<td>0.0001</td>
<td>0.0010</td>
<td>0.0227</td>
<td>11.6</td>
<td>2,426</td>
<td>05/07/1963</td>
<td>30/09/2011</td>
</tr>
<tr>
<td>S&amp;P100‡</td>
<td>0.0007</td>
<td>0.0019</td>
<td>0.0243</td>
<td>8.3</td>
<td>1,469</td>
<td>04/08/1982</td>
<td>29/09/2011</td>
</tr>
<tr>
<td>CRSP§</td>
<td>0.0061</td>
<td>0.0096</td>
<td>0.0545</td>
<td>10.4</td>
<td>740</td>
<td>01/01/1950</td>
<td>01/08/2011</td>
</tr>
<tr>
<td>S&amp;P500§</td>
<td>0.0019</td>
<td>0.0054</td>
<td>0.0424</td>
<td>5.3</td>
<td>740</td>
<td>01/01/1950</td>
<td>01/08/2011</td>
</tr>
<tr>
<td>S&amp;P100§</td>
<td>0.0023</td>
<td>0.0061</td>
<td>0.0492</td>
<td>7.1</td>
<td>330</td>
<td>02/04/1984</td>
<td>02/10/2011</td>
</tr>
</tbody>
</table>

Superscripts †, ‡ and § denote daily, weekly and monthly frequencies, respectively. The null hypothesis in the Jarque-Bera test is reject in all indices and frequencies.
Table 10: Empirical application: risk premium estimation

<table>
<thead>
<tr>
<th></th>
<th>CRSP</th>
<th></th>
<th>S&amp;P500</th>
<th></th>
<th>S&amp;P100</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NL-ILS</td>
<td>QMLE</td>
<td>NL-ILS</td>
<td>QMLE</td>
<td>NL-ILS</td>
<td>QMLE</td>
</tr>
<tr>
<td><strong>λ</strong></td>
<td>0.02*</td>
<td>0.07***</td>
<td>0.01</td>
<td>0.04***</td>
<td>0.02*</td>
<td>0.05***</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.009)</td>
<td>(0.010)</td>
<td>(0.009)</td>
<td>(0.012)</td>
<td>(0.011)</td>
</tr>
<tr>
<td><strong>ω</strong></td>
<td>0.00</td>
<td>0.00***</td>
<td>0.00</td>
<td>0.00***</td>
<td>0.00</td>
<td>0.00***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td><strong>α</strong></td>
<td>0.12***</td>
<td>0.09***</td>
<td>0.10***</td>
<td>0.08***</td>
<td>0.11***</td>
<td>0.08***</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.002)</td>
<td>(0.022)</td>
<td>(0.002)</td>
<td>(0.015)</td>
<td>(0.002)</td>
</tr>
<tr>
<td><strong>β</strong></td>
<td>0.84***</td>
<td>0.91***</td>
<td>0.83***</td>
<td>0.92***</td>
<td>0.80***</td>
<td>0.91***</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.003)</td>
<td>(0.034)</td>
<td>(0.002)</td>
<td>(0.077)</td>
<td>(0.003)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>CRSP</th>
<th></th>
<th>S&amp;P500</th>
<th></th>
<th>S&amp;P100</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NL-ILS</td>
<td>QMLE</td>
<td>NL-ILS</td>
<td>QMLE</td>
<td>NL-ILS</td>
<td>QMLE</td>
</tr>
<tr>
<td><strong>λ</strong></td>
<td>0.06**</td>
<td>0.11***</td>
<td>0.01</td>
<td>0.06***</td>
<td>0.03</td>
<td>0.08***</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.019)</td>
<td>(0.022)</td>
<td>(0.020)</td>
<td>(0.029)</td>
<td>(0.025)</td>
</tr>
<tr>
<td><strong>ω</strong></td>
<td>0.00</td>
<td>0.00***</td>
<td>0.00</td>
<td>0.00***</td>
<td>0.00</td>
<td>0.00***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td><strong>α</strong></td>
<td>0.12***</td>
<td>0.14***</td>
<td>0.11**</td>
<td>0.13***</td>
<td>0.14***</td>
<td>0.14***</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.011)</td>
<td>(0.058)</td>
<td>(0.009)</td>
<td>(0.054)</td>
<td>(0.012)</td>
</tr>
<tr>
<td><strong>β</strong></td>
<td>0.77***</td>
<td>0.84***</td>
<td>0.81***</td>
<td>0.85***</td>
<td>0.74***</td>
<td>0.84***</td>
</tr>
<tr>
<td></td>
<td>(0.209)</td>
<td>(0.013)</td>
<td>(0.204)</td>
<td>(0.012)</td>
<td>(0.207)</td>
<td>(0.015)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>CRSP</th>
<th></th>
<th>S&amp;P500</th>
<th></th>
<th>S&amp;P100</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NL-ILS</td>
<td>QMLE</td>
<td>NL-ILS</td>
<td>QMLE</td>
<td>NL-ILS</td>
<td>QMLE</td>
</tr>
<tr>
<td><strong>λ</strong></td>
<td>0.12***</td>
<td>0.18***</td>
<td>0.05</td>
<td>0.07*</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td>(0.031)</td>
<td>(0.044)</td>
<td>(0.038)</td>
<td>(0.063)</td>
<td>(0.058)</td>
</tr>
<tr>
<td><strong>ω</strong></td>
<td>0.00</td>
<td>0.00***</td>
<td>0.00</td>
<td>0.00***</td>
<td>0.00</td>
<td>0.00*</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.001)</td>
<td>(0.000)</td>
</tr>
<tr>
<td><strong>α</strong></td>
<td>0.09</td>
<td>0.14***</td>
<td>0.11***</td>
<td>0.11***</td>
<td>0.04</td>
<td>0.14***</td>
</tr>
<tr>
<td></td>
<td>(0.063)</td>
<td>(0.019)</td>
<td>(0.042)</td>
<td>(0.025)</td>
<td>(0.078)</td>
<td>(0.042)</td>
</tr>
<tr>
<td><strong>β</strong></td>
<td>0.88***</td>
<td>0.84***</td>
<td>0.71***</td>
<td>0.85***</td>
<td>0.77***</td>
<td>0.82***</td>
</tr>
<tr>
<td></td>
<td>(0.138)</td>
<td>(0.018)</td>
<td>(0.186)</td>
<td>(0.028)</td>
<td>(0.218)</td>
<td>(0.055)</td>
</tr>
</tbody>
</table>

Standard errors are reported inside the brackets. NL-ILS standard errors are obtained using block bootstrap algorithm with 1000 replications. QMLE standard errors are computed using Bollerslev-Wooldridge robust estimator. The symbols *, **, and *** denote significance 10%, 5% and 1%, respectively.
Table 11: Empirical application: risk premium estimation
- RealGARCH(1,1)-in-mean

<table>
<thead>
<tr>
<th></th>
<th>CRSP</th>
<th>S&amp;P500</th>
<th>S&amp;P100*</th>
<th>CRSP</th>
<th>S&amp;P500</th>
<th>S&amp;P100*</th>
<th>CRSP</th>
<th>S&amp;P500</th>
<th>S&amp;P100*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily freq.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL-ILS</td>
<td>0.01</td>
<td>0.01</td>
<td></td>
<td>0.03</td>
<td>0.02</td>
<td></td>
<td>0.10</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>(0.019)</td>
<td>(0.020)</td>
<td>(0.052)</td>
<td>(0.045)</td>
<td>(0.115)</td>
<td>(0.110)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weekly freq.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL-ILS</td>
<td>1.44***</td>
<td>0.59***</td>
<td></td>
<td>0.50</td>
<td>0.23</td>
<td></td>
<td>0.66</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>(0.242)</td>
<td>(0.165)</td>
<td>(0.493)</td>
<td>(0.521)</td>
<td>(0.941)</td>
<td>(1.277)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monthly freq.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL-ILS</td>
<td>0.51***</td>
<td>0.56***</td>
<td></td>
<td>0.29***</td>
<td>0.28***</td>
<td></td>
<td>0.25**</td>
<td>0.28**</td>
<td></td>
</tr>
<tr>
<td>(0.045)</td>
<td>(0.032)</td>
<td>(0.074)</td>
<td>(0.053)</td>
<td>(0.109)</td>
<td>(0.130)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standard errors are reported inside the brackets. NL-ILS and MLE standard errors are obtained using block bootstrap algorithm with 1000 replications. The symbols *, **, and *** denote significance 10%, 5% and 1%, respectively.

Table 12: Robustness check: risk premium estimation

<table>
<thead>
<tr>
<th></th>
<th>CRSP</th>
<th>S&amp;P500</th>
<th>S&amp;P100*</th>
<th>CRSP</th>
<th>S&amp;P500</th>
<th>S&amp;P100*</th>
<th>CRSP</th>
<th>S&amp;P500</th>
<th>S&amp;P100*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily freq.</td>
<td>0.031***</td>
<td>0.004</td>
<td>0.013</td>
<td>0.033***</td>
<td>0.006</td>
<td>0.015</td>
<td>0.039**</td>
<td>0.012</td>
<td>0.022*</td>
</tr>
<tr>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.012)</td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.012)</td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.012)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>Weekly freq.</td>
<td>0.064***</td>
<td>0.004</td>
<td>0.014</td>
<td>0.062***</td>
<td>0.005</td>
<td>0.042</td>
<td>0.069**</td>
<td>0.017</td>
<td>0.042</td>
</tr>
<tr>
<td>(0.021)</td>
<td>(0.020)</td>
<td>(0.026)</td>
<td>(0.021)</td>
<td>(0.021)</td>
<td>(0.026)</td>
<td>(0.021)</td>
<td>(0.021)</td>
<td>(0.021)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>Monthly freq.</td>
<td>0.152***</td>
<td>0.065*</td>
<td>0.051</td>
<td>0.152***</td>
<td>0.014</td>
<td>0.015</td>
<td>0.156**</td>
<td>0.063</td>
<td>0.060</td>
</tr>
<tr>
<td>(0.032)</td>
<td>(0.030)</td>
<td>(0.064)</td>
<td>(0.032)</td>
<td>(0.033)</td>
<td>(0.012)</td>
<td>(0.032)</td>
<td>(0.039)</td>
<td>(0.062)</td>
<td></td>
</tr>
</tbody>
</table>

I report estimates of λ computed using QMLE methodology. Standard errors are computed using the Bollerslev-Wooldridge estimator. The symbols *, **, and *** denote significance 10%, 5% and 1%, respectively.