Parameter estimation in nonlinear AR–GARCH models

Mika Meitz
Koç University

Pentti Saikkonen
University of Helsinki

University of Vienna

November 23, 2009
Outline of the Presentation

- Brief Introduction to (Refresher of) GARCH models
  - Basic ARCH and GARCH & Extensions and Different Flavors of GARCH
  - Is there interest in these topics?

- What we do in this paper
  - Consistency and Asymptotic Normality of the QMLE
  - Previous Literature

- Results
  - The Nonlinear AR–GARCH model we consider
  - Consistency
    - Road Map of a Basic Proof & Details of some Complications
  - Asymptotic Normality
    - Road Map of a Basic Proof & Outline of some Complications
  - Some Remarks of Assumptions Made
  - Examples
    - Linear AR–GARCH & Nonlinear AR–GARCH

- Conclusions
  - Summary of what we did & Relation to existing literature
Outline of the Presentation

- Brief Introduction to (Refresher of) GARCH models
  - Basic ARCH and GARCH & Extensions and Different Flavors of GARCH
  - Is there interest in these topics?

- What we do in this paper
  - Consistency and Asymptotic Normality of the QMLE
  - Previous Literature

- Results
  - The Nonlinear AR–GARCH model we consider
  - Consistency
    - Road Map of a Basic Proof & Details of some Complications
  - Asymptotic Normality
    - Road Map of a Basic Proof & Outline of some Complications
  - Some Remarks of Assumptions Made
  - Examples
    - Linear AR–GARCH & Nonlinear AR–GARCH

- Conclusions
  - Summary of what we did & Relation to existing literature
posed a discrete time model. If the price of risk were constant over time, then rising conditional variances would translate linearly into rising expected returns. Thus the mean of the return equation would no longer be estimated as zero, it would depend upon the past squared returns exactly in the same way that the conditional variance depends on past squared returns. This very strong coefficient restriction can be tested and used to estimate the price of risk. It can also be used to measure the coefficient of relative risk aversion of the representative agent under the same assumptions.

Empirical evidence on this measurement has been mixed. While Engle et al. (1987) find a positive and significant effect, Chou, Engle and Kane (1992), and Glosten, Jagannathan and Runkle (1993), find a relationship that varies over time and may be negative because of omitted variables. French, Schwert and Stambaugh (1987) showed that a positive volatility surprise should and does have a negative effect on asset prices. There is not simply one risky asset in the economy and the price of risk is not likely to be constant, hence the instability is not surprising and does not disprove the existence of the risk-return trade-off, but it is a challenge to better modeling of this trade-off.

The causes of volatility are more directly modeled. Since the basic ARCH model and its many variants describe the conditional variance as a function of lagged squared returns, these are perhaps the proximate causes of volatility. It is best to interpret these as observables that help in forecasting volatility rather than as causes. If the true causes were included in the specification, then the lags would not be needed.

A small collection of papers has followed this route. Andersen and Bollerslev (1998b) examined the effects of announcements on exchange rate

[Original Figure: Robert F. Engle, Nobel Lecture 2003, Figure 1]
Intro : Autocorrelations of Returns & Squared Returns

Historical volatility are based on rolling standard deviations of returns. In (Figure 7) these are constructed for 5 day, one year, and five year windows. While each of these approaches may seem reasonable, the answers are clearly very different. The 5 day estimate is extremely variable while the other two are much smoother. The 5 year estimate smooths over peaks and troughs that the other two see. It is particularly slow to recover after the 87 crash and particularly slow to reveal the rise in volatility in 1998–2000. In just the same way, the annual estimate fails to show all the details revealed by the 5 day volatility. However, some of these details may be just noise. Without any true measure of volatility, it is difficult to pick from these candidates.

The ARCH model provides a solution to this dilemma. From estimating the unknown parameters based on the historical data, we have forecasts for each day in the sample period and for any period after the sample. The natural first model to estimate is the GARCH (1,1). This model gives weights to the unconditional variance, the previous forecast, and the news measured as the square of yesterday's return. The weights are estimated to be (.004, .941, .055) respectively. Clearly the bulk of the information comes from the previous day forecast. The new information changes this a little and the long run average variance has a very small effect. It appears that the long run variance effect is so tiny that it might not be important. This is incorrect. When forecasting many steps ahead, the long run variance eventually dominates as the importance of news and other recent information fades away. It is naturally small because of the use of daily data.

In this example, we will use an asymmetric volatility model that is some-
Basic ARCH and GARCH

- The basic GARCH(1,1) model

\[ y_t = \sigma_t \varepsilon_t \]
\[ \sigma^2_t = \omega + \alpha y_{t-1}^2 + \beta \sigma^2_{t-1} \]
\[ \varepsilon_t \sim \text{IID } N(0, 1) \]

- \( y_t \) are the observed returns
- \( \sigma_t^2 \) is their conditional variance (\( \sigma^2_t = E[y_t^2|\mathcal{F}_{t-1}] \))

- Provides a simple, easily estimable, and empirically reasonably fitting characterization of evolution of volatility
- One of the most commonly applied models for financial time series
- Has found applications with a wide range of financial returns
  - US stocks; stocks traded at other developed markets & emerging markets
  - Indices of equity returns; exchange rates; bond & commodity returns
Extensions of GARCH

- The two most basic ‘extensions’ are to
  - include a (usually simple) conditional mean
  - consider nonlinear models

- Why a conditional mean?
  - Typically a small degree of autocorrelation in the returns
  - Including a mean provides a better fit

- Why nonlinear models?
  - Linearity is often just a convenient, simplifying assumption
  - “Using a term like nonlinear science is like referring to the bulk of zoology as the study of non-elephant animals” – Stanislaw Ulam
  - Economic reasons in the context of GARCH
    - Leverage effects
    - Asymmetric loss functions
    - Regime shifts

There exists a long list of different nonlinear GARCH models.
Extensions of GARCH

- The two most basic ‘extensions’ are to
  - include a (usually simple) conditional mean
  - consider nonlinear models

- Why a conditional mean?
  - Typically a small degree of autocorrelation in the returns
  - Including a mean provides a better fit

- Why nonlinear models?
  - Linearity is often just a convenient, simplifying assumption
  - “Using a term like nonlinear science is like referring to the bulk of zoology as the study of non-elephant animals”
    - Stanislaw Ulam
  - Economic reasons in the context of GARCH
    - Leverage effects
    - Asymmetric loss functions
    - Regime shifts
  - There exists a looong list of different nonlinear GARCH models
Different Flavors of GARCH

- **Linear GARCH**
  \[ y_t = \sigma_t \varepsilon_t \]
  \[ \sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 \]
  \[ \varepsilon_t \sim \text{IID}(0, 1) \]

- **Threshold (or ‘GJR’) GARCH**
  \[ \sigma_t^2 = \omega + (\alpha_1 I(y_{t-1} \geq 0) + \alpha_2 I(y_{t-1} < 0)) y_{t-1}^2 + \beta \sigma_{t-1}^2 \]

- **Smooth Transition GARCH**
  \[ \sigma_t^2 = \omega + (\alpha_1 + \alpha_2 G(y_{t-1})) y_{t-1}^2 + \beta \sigma_{t-1}^2 \]

- **‘General Nonlinear’ GARCH**
  \[ \sigma_t^2 = g(y_{t-1}, \sigma_{t-1}^2) \]
Different Flavors of GARCH

- **Linear AR–GARCH**

  \[ y_t = \phi_0 + \phi_1 y_{t-1} + \sigma_t \epsilon_t \]
  \[ u_t = y_t - (\phi_0 + \phi_1 y_{t-1}) \]
  \[ \sigma_t^2 = \omega + \alpha u_{t-1}^2 + \beta \sigma_{t-1}^2 \]

  (an AR(1) to make things simple; could be ARMA; could be more lags)

- **Nonlinear AR–GARCH**

  \[ y_t = f(y_{t-1}) + \sigma_t \epsilon_t \]
  \[ u_t = y_t - f(y_{t-1}) \]
  \[ \sigma_t^2 = g(u_{t-1}, \sigma_{t-1}^2) \]

  ▶ For example, \( f(y_{t-1}) \) could be

  \[ f(y_{t-1}) = (\phi_0 + \phi_1 y_{t-1}) + (\psi_0 + \psi_1 y_{t-1}) F(y_{t-1}) \]

  ▶ **Threshold AR (TAR) model**: \( F \) is an indicator function (of, say, \( y_{t-1} \geq 0 \))
  ▶ **Smooth Transition AR (STAR) model**: \( F \) is a nonlinear function with range \([0, 1]\)
Is there interest in these topics?

- Is there interest in GARCH models?
  - Kim, Morse, and Zingales (2006, *J. Economic Perspectives*), “What Has Mattered To Economics Since 1970”
    - #24 Bollerslev (1986, *J. Econometrics*)
  - Nobel Prize in Economics 2003 to Robert F. Engle
    “for methods of analyzing economic time series with time-varying volatility (ARCH)”
    - (joint with Clive Granger, for cointegration)
Is there interest in these topics?

- **Is there interest in GARCH models?**
  - Kim, Morse, and Zingales (2006, *J. Economic Perspectives*), “What Has Mattered To Economics Since 1970”
    - #24 Bollerslev (1986, *J. Econometrics*)
  - Nobel Prize in Economics 2003 to Robert F. Engle
    “for methods of analyzing economic time series with time-varying volatility (ARCH)”
    - (joint with Clive Granger, for cointegration)

- **Is there interest in nonlinear AR–GARCH models?**
  - Engle (1982) and Bollerslev (1986)
    - All their empirical applications use AR–GARCH models (not pure GARCH)
    - Numerous nonlinear GARCH models
    - Two examples: One using a linear AR–GARCH model, the other a nonlinear AR–GARCH model (mean&variance both nonlinear)
  - Subsequently, numerous articles
  - ‘Nonlinear AR–GARCH’ is more the rule than the exception
Outline of the Presentation

- Brief Introduction to (Refresher of) GARCH models
  - Basic ARCH and GARCH & Extensions and Different Flavors of GARCH
  - Is there interest in these topics?
- What we do in this paper
  - Consistency and Asymptotic Normality of the QMLE
  - Previous Literature
- Results
  - The Nonlinear AR–GARCH model we consider
  - Consistency
    - Road Map of a Basic Proof & Details of some Complications
  - Asymptotic Normality
    - Road Map of a Basic Proof & Outline of some Complications
  - Some Remarks of Assumptions Made
  - Examples
    - Linear AR–GARCH & Nonlinear AR–GARCH
- Conclusions
  - Summary of what we did & Relation to existing literature
What we do in this paper

What do we want to do in this paper? (more details will follow!)

■ The model we consider: A General Nonlinear AR–GARCH model

\[ y_t = f(y_{t-1}; \mu_0) + \sigma_t \varepsilon_t \]

\[ u_{0t} = y_t - f(y_{t-1}; \mu_0) \]

\[ \sigma_t^2 = g(u_{0,t-1}, \sigma_{t-1}^2; \theta_0) \]

▶ The true parameter \( \theta_0 \) belongs to some parameter space \( \Theta \)
▶ ‘Feasible’ Gaussian log-likelihood (multiplied by \( -\frac{2}{T} \), modulo a constant)

\[ L_T(\theta) = T^{-1} \sum_{t=1}^{T} l_t(\theta), \quad l_t(\theta) = \log(h_t(\theta)) + \frac{u_t^2(\theta)}{h_t(\theta)} \]

▶ Quasi-Maximum Likelihood (QML) estimator for the parameter \( \theta_0 \) is

\[ \hat{\theta}_T = \arg \min_{\theta \in \Theta} L_T(\theta) \]

■ What we want to show is

▶ Strong consistency of the QML estimator (Thm 1)

\[ \hat{\theta}_T \to \theta_0 \quad \text{a.s.} \]

▶ Asymptotic normality of the QML estimator (Thm 2)

\[ T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, \mathcal{J}(\theta_0)^{-1}\mathcal{I}(\theta_0)\mathcal{J}(\theta_0)^{-1}) \]
What we do in this paper

Previous Literature (a selection)

- Asymptotic estimation theory for these models
  - Linear ‘pure’ GARCH
    - Hall and Yao (2003, *Econometrica*)
  - Linear AR–GARCH
    - Ling and McAleer (2003, *Econometric Theory*)
    - Francq and Zakoïan (2004, *Bernoulli*)
    - Ling (2007, *Journal of Econometrics*)
    + Lange, Rahbek, and Jensen (WP) (ARCH)
  - Nonlinear ‘pure’ GARCH
  - Nonlinear AR–GARCH
    - —
Outline of the Presentation

- Brief Introduction to (Refresher of) GARCH models
  - Basic ARCH and GARCH & Extensions and Different Flavors of GARCH
  - Is there interest in these topics?
- What we do in this paper
  - Consistency and Asymptotic Normality of the QMLE
  - Previous Literature
- Results
  - The Nonlinear AR–GARCH model we consider
  - Consistency
    - Road Map of a Basic Proof & Details of some Complications
  - Asymptotic Normality
    - Road Map of a Basic Proof & Outline of some Complications
  - Some Remarks of Assumptions Made
  - Examples
    - Linear AR–GARCH & Nonlinear AR–GARCH
- Conclusions
  - Summary of what we did & Relation to existing literature
The Nonlinear AR–GARCH model we consider

- We assume the data is generated by

\[ y_t = f(y_{t-1}, \ldots, y_{t-p}; \mu_0) + \sigma_t \varepsilon_t, \quad t = 1, 2, \ldots \]

\[ f(\ldots) = \sum_{j=1}^{p} a_j(y_{t-1}, \ldots, y_{t-p}; \mu_0) y_{t-j} + b(y_{t-1}, \ldots, y_{t-p}; \mu_0) \]

(\(a_j, b\) some nonlinear functions)

\[ \sigma_t^2 = g(u_{0,t-1}, \sigma_{t-1}^2; \theta_0) \]

\[ u_{0,t} = y_t - f(y_{t-1}, \ldots, y_{t-p}; \mu_0) \]

\[ \varepsilon_t \text{ IID rv's; } E[\varepsilon_t] = 0, \quad E[\varepsilon_t^2] = 1, \quad \varepsilon_t \text{ indep. of } \{y_s, s < t\} \]

- For brevity, in the following I will often just write (with \(p = 1\)):

\[ y_t = f(y_{t-1}; \mu_0) + \sigma_t \varepsilon_t \]

\[ u_{0,t} = y_t - f(y_{t-1}; \mu_0) \]

\[ \sigma_t^2 = g(u_{0,t-1}^2, \sigma_{t-1}^2; \theta_0) \]
“Road Map of a Basic Consistency Proof”
(a.k.a. what we would have done if things were simple)

- **The Data Generating Process**
  - The Nonlinear AR–GARCH model
    \[
    y_t = f(y_{t-1}; \mu_0) + \sigma_t \epsilon_t \\
    u_{0,t} = y_t - f(y_{t-1}; \mu_0) \\
    \sigma^2_t = g(u_{0,t-1}, \sigma^2_{t-1}; \theta_0)
    \]
  - Likelihood Function
    \[
    L_T(\theta) = T^{-1} \sum_{t=1}^{T} l_t(\theta)
    \]
    \( (y_t, \sigma^2_t) \) is stationary and ergodic

- **Consistency of the QMLE**
  - ULLN for \( L_T(\theta) \):
    \[
    \sup_{\theta \in \Theta} \left| L_T(\theta) - E[l_t(\theta)] \right| \longrightarrow 0 \ a.s.
    \]
    \( l_t(\theta) \) stationary and ergodic
  - \( E[\sup_{\theta \in \Theta} |l_t(\theta)|] < \infty \)
  - Unique minimum of \( E[l_t(\theta)] \) at \( \theta = \theta_0 \)
    \( \) Identification condition for \( l_t(\theta_0) \)
Issues that require some attention in our case

1 Stationarity and ergodicity of the DGP \((y_t, \sigma^2_t)\)
   - At a minimum, we typically would like to know that a stationary and ergodic solution at least exists
   - Establishing this can be rather complicated in GARCH-type models

2 Defining two likelihood functions
   - “Feasible” likelihood to be used in practice
   - “Theoretically convenient” likelihood to be used in the proofs
   - Existence of the latter, and its connection to the former, need some attention

3 Proving consistency without Uniform Laws of Large Numbers
   - The quantities appearing in the likelihood do not have finite expectations
   - Traditional application of ULLN’s not directly possible

4 Global identification
   - Ensuring the existence of unique optimum complicated in general nonlinear models

→ Next, I’ll discuss each of these in a bit more detail
Stationary & Ergodicity of the DGP

- Stationarity & Ergodicity (S & E) of various quantities facilitates the development of asymptotic estimation theory
  - Ensure the validity of a Strong LLN: \( T^{-1} \sum_{t=1}^{T} X_t \rightarrow E[X_t] \) a.s.
  - Often, establishing S & E of the DGP the first step in asymptotic estimation theory
  - Lack of results for S & E explains the lack of results on estimation theory

- For nonlinear AR–GARCH models, such results we obtained in our earlier paper “Stability of nonlinear AR–GARCH models” (Meitz and Saikkonen 2008, Journal of Time Series Analysis)
  - Results on Stationarity, Ergodicity, Mixing, Existence of moments
  - First results for nonlinear AR–GARCH models
  - Opened up the way for the development of asymptotic estimation theory

- Hence, in this paper, we assume the following:
  **Assumption DGP:** The process \((y_t, \sigma_t^2)\) is stationary and ergodic with \(E[|y_t|^{2r}] < \infty\) and \(E[\sigma_t^{2r}] < \infty\) for some \(r > 0\).
  (in the examples, low-level conditions for this to hold are provided)
Defining two likelihood functions

- Nonlinear AR–GARCH

\[ y_t = f(y_{t-1}; \mu_0) + \sigma_t \varepsilon_t \]
\[ u_{0t} = y_t - f(y_{t-1}; \mu_0) \]
\[ \sigma_t^2 = g(u_{0,t-1}, \sigma_{t-1}^2; \theta_0) \]

- Here \( y_t \) is observable, but \( \sigma_t^2 \) is not observable
  - Moreover, even if we knew \( \theta_0 \), the distribution of \( \sigma_t \) is unknown and values for it cannot be computed
  - Estimation cannot be based on the ‘natural’ log-likelihood
    \[ l_t(\theta) = \log(\sigma_t^2(\theta)) + \frac{u_{2}^2(\theta)}{\sigma_t^2(\theta)} \]
  - In practical applications, an approximation to this model is the model always estimated
“Feasible” likelihood

- The ‘feasible’ likelihood used in practice is based on the approximation

\[
y_t = f(y_{t-1}; \mu) + h_t^{1/2}(\theta)\varepsilon_t
\]

\[
u_t(\theta) = y_t - f(y_{t-1}; \mu)
\]

\[
h_t(\theta) = \begin{cases} 
\varsigma_0 \in \mathbb{R}_+, & t = 0 \\
g(u_{t-1}(\theta), h_{t-1}(\theta); \theta), & t \geq 1 
\end{cases}
\]

> Note: Setting \(\varsigma_0\) to a draw from the distribution of \(\sigma_t^2\) not feasible

- ‘Feasible’ Gaussian log-likelihood
  (multiplied by \(-\frac{2}{T}\); modulo a constant)

\[
L_T(\theta) = T^{-1}\sum_{t=1}^{T} l_t(\theta), \quad l_t(\theta) = \log(h_t(\theta)) + \frac{u_t^2(\theta)}{h_t(\theta)}
\]

> Complication: The \(l_t(\theta)\) are not Stationary and Ergodic since \(h_t(\theta)\) are not

> It would be easier to work with S & E quantities in the proofs
“Theoretically convenient” likelihood

Solution? Find a S & E process \( h^*_t(\theta) \) that ‘approximates’ \( h_t(\theta) \)

- Approach similar to BHK2003, FZ2004, and SM2006

How is this process \( h^*_t(\theta) \) found?

- In the **linear case** (BHK, FZ), more straightforward as Explicit expressions for \( h_t(\theta) \) exist
- In the nonlinear case, \( h_t(\theta) \) is merely an infinite sequence resulting from iterated application of random functions – more complicated
- Proposition 1 – some details on next slide – Solutions to Banach space-valued iterated random functions

Now we can define

\[
L^*_T(\theta) = T^{-1} \sum_{t=1}^{T} l^*_t(\theta), \quad l^*_t(\theta) = \log(h^*_t(\theta)) + \frac{u^2_t(\theta)}{h^*_t(\theta)}
\]

- \( l^*_t(\theta) \) is S & E and approximates \( l_t(\theta) \)
- Also: Minimizers of \( L_T(\theta) \) and \( L^*_T(\theta) \) will be asymptotically equivalent
Details of Proposition 1

Assumptions:

C1 Compact parameter space $\Theta$

C2 Function $g(\cdot, \cdot; \cdot)$ continuous w.r.t. all its arguments and

(i) $g(u, x; \theta) \leq \varrho x + \kappa u^2 + \varpi$ with $0 < \varrho < 1$

(ii) $|g(u, x_1; \theta) - g(u, x_2; \theta)| \leq \kappa |x_1 - x_2|$ with $0 < \kappa < 1$

C3 Functions $a(\cdot; \cdot)$ and $b(\cdot; \cdot)$ bounded and measurable

Proposition 1: Suppose Assumptions DGP and C1–C3 hold. Then, for all $\theta \in \Theta$ there exists a stationary and ergodic solution $h^*_t(\theta)$ to the equation

$$h_t(\theta) = g(u_{t-1}, h_{t-1}(\theta); \theta), \quad t = 1, 2, \ldots.$$  (*)

This solution is continuous in $\theta$, measurable w.r.t. the $\sigma$–algebra generated by $(y_{t-1}, y_{t-2}, \ldots)$, and it is unique when (*) is extended to all $t \in \mathbb{Z}$. Furthermore, the solution $h^*_t(\theta)$ has the properties $h^*_t(\theta_0) = \sigma^2_t$ and $E[\sup_{\theta \in \Theta} h^*_t(\theta)] < \infty$. If $h_t(\theta), \theta \in \Theta$, are any other solutions to (*), then for some $\gamma > 1$, $\gamma^t \sup_{\theta \in \Theta} |h^*_t(\theta) - h_t(\theta)| \to 0$ in $L_r$–norm as $t \to \infty$. 
Proving consistency without ULLN’s

- A basic consistency proof would rely on a ULLN for $L_T^*(\theta)$:
  $$\sup_{\theta \in \Theta} | L_T^*(\theta) - E[l_t^*(\theta)] | \to 0 \text{ a.s.}$$
- This could require that $l_t^*(\theta)$ is S & E and $E[\sup_{\theta \in \Theta} | l_t^*(\theta) |] < \infty$
  (other sets of assumptions also possible)
- Under the assumptions we make, $E[l_t^*(\theta)] \in \mathbb{R} \cup \{+\infty\}$
  - ULLN can not be applied
  - Note: ULLN would apply if we made stronger assumptions
    - existence of 2nd moments — but this is much more than we need
- We have to use ‘more basic arguments’ (à la Pfanzagl (1969, Metrika))
  - In technical (and not very informative) terms, we prove the following
    - $\sup_{\theta \in \Theta} | L_T^*(\theta) - L_T(\theta) | \to 0 \text{ a.s.}$
    - $| L_T^*(\theta_0) - E[l_t^*(\theta_0)] | \to 0 \text{ a.s.}$
    - $\lim_{T \to \infty} \inf_{\theta \in B(\theta_0, \delta)^c} L_T^*(\theta) \geq \inf_{\theta \in B(\theta_0, \delta)^c} E[l_t^*(\theta)] > E[l_t^*(\theta_0)] \text{ a.s.}$
- This gives us the consistency result
Consistency result – Theorem 1

Assumptions:

C4 The functions $a(z; \cdot)$ and $b(z; \cdot)$ are continuous for every $z \in \mathbb{R}^p$

C5 The function $g$ is bounded away from zero: $g(\cdot, \cdot; \cdot) \geq g > 0$

C6 Global identification conditions

(i) $f(y_{t-1}, \ldots, y_{t-p}; \mu) = f(y_{t-1}, \ldots, y_{t-p}; \mu_0)$ a.s. iff $\mu = \mu_0$

(ii) $h^*_t(\mu_0, \lambda) = \sigma_t^2$ a.s. iff $\lambda = \lambda_0$

Theorem 1: Suppose Assumptions DGP and C1–C6 hold. Then the QML estimator $\hat{\Theta}_T$ is strongly consistent, that is, $\hat{\Theta}_T \rightarrow \theta_0$ a.s.

Note: I did not yet discuss the ‘Difficulty 4’: Global Identification

- Verifying C6 in Examples turns out to be rather complicated – more on this in connection with the Examples
Summary of Difficulties in the Consistency Proof

- \((y_t, \sigma^2_t)\) is S & E
  - Meitz & Saikkonen (2008, *JTSA*)

- Calculation of \(\sigma^2_t\) impossible
  - Replace with feasible \(h_t(\theta)\)

- \(h_t(\theta)\) or \(l_t(\theta)\) not S & E
  - Find a S & E process \(h^*_t(\theta)\) that ‘approximates’ \(h_t(\theta)\)

- ULLN can not be used since necessary moments not existing
  - Avoid using a ULLN by going to a ‘basic level consistency proof’

\[ \hat{\theta}_T \longrightarrow \theta_0 \quad \text{a.s.} \]
“Road Map of a Basic Asymptotic Normality Proof”

Typically done using a Taylor Series Expansion of the Score

\[ T^{1/2} \frac{\partial L_T(\hat{\theta}_T)}{\partial \theta} = T^{1/2} \frac{\partial L_T(\theta_0)}{\partial \theta} + \frac{\partial^2 L_T(\theta)}{\partial \theta \partial \theta'} T^{1/2}(\hat{\theta}_T - \theta_0) \]

Asymptotic Normality of the QMLE

- CLT for Score:
  \[ T^{1/2} \frac{\partial L_T(\theta_0)}{\partial \theta} \xrightarrow{d} N(0, I(\theta_0)), \quad I(\theta_0) = E \left[ \frac{\partial l_t(\theta_0)}{\partial \theta} \frac{\partial l_t(\theta_0)}{\partial \theta'} \right] \]
  - \( l_t(\theta) \) once continuously differentiable
  - \( \frac{\partial l_t(\theta_0)}{\partial \theta} \) stationary and ergodic MDS and \( E \left[ \frac{\partial^2 l_t(\theta_0)}{\partial \theta \partial \theta'} \right] < \infty \)

- ULLN for Hessian:
  \[ \sup_{\theta \in N(\theta_0)} \left| \frac{\partial^2 L_T(\theta)}{\partial \theta \partial \theta'} - J(\theta) \right| \longrightarrow 0 \ \text{a.s.}, \quad J(\theta) = E \left[ \frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'} \right] \]
  - \( l_t(\theta) \) twice continuously differentiable
  - \( \frac{\partial^2 l_t(\theta_0)}{\partial \theta \partial \theta'} \) stationary and ergodic and \( E \left[ \sup_{\theta \in N(\theta_0)} \left| \frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'} \right| \right] < \infty \)

- Matrices \( I(\theta_0) \) and \( J(\theta_0) \) Positive Definite
  - Condition of linear independence for the components of \( \frac{\partial l_t(\theta_0)}{\partial \theta} \)
Challenges in proving asymptotic normality – let’s only make brief remarks on issues requiring attention

1 Showing that $l_t^*(\theta)$ is twice continuously differentiable and that $\frac{\partial l_t^*(\theta_0)}{\partial \theta}$ & $\frac{\partial^2 l_t^*(\theta)}{\partial \theta \partial \theta'}$ are S & E
   ▶ Show that the S & E process $h_t^*(\theta)$ is twice continuously differentiable
   ▶ Show that these derivatives are S & E
   ▶ Show that the derivatives $l_t^*(\theta)$ approximate those of $l_t(\theta)$
   ▶ Propositions 2 and 3 in the paper
     ▶ Technicalities again similar to Bougerol (1993) and S&M (2006)

2 Showing that the “technically convenient” estimator is as. normally distributed
   ▶ $\tilde{\theta}_T = \arg \min_{\theta \in \Theta} L_T^*(\theta)$
   ▶ $T^{1/2}(\tilde{\theta}_T - \theta_0) \xrightarrow{d} N(0, J(\theta_0)^{-1} \mathcal{I}(\theta_0) J(\theta_0)^{-1})$
     ▶ CLT for Score
     ▶ ULLN for Hessian
     ▶ $\mathcal{I}$ & $\mathcal{J}$ positive definite

3 Finally, show that $\hat{\theta}_T$ and $\tilde{\theta}_T$ have asymptotically the same distribution
Asymptotic normality result – Theorem 2

N1 The true parameter value $\theta_0$ is an interior point of $\Theta$.

N2 Functions $a$, $b$, and $g$ twice continuously partially differentiable on a suitable set $\Theta_0$

N3 (i,ii) Partial derivatives of $a$, $b$, and $g$ ‘suitably bounded’
   (iii) Lipschitz-conditions on first and second derivatives of $g$

N4 Assumption DGP holds with $r = 2$; $E[\varepsilon_t^4] < \infty$;
   \[ \sup_{\theta \in \Theta_0} \left| \frac{h^*_t(t(\theta))}{h^*_t(\theta)} \right|_4 < \infty \quad \text{and} \quad \sup_{\theta \in \Theta_0} \left| \frac{h^*_{\theta\theta,t}(\theta)}{h^*_t(\theta)} \right|_2 < \infty. \]

N5 (i) The distribution of $\varepsilon_t$ is not concentrated at two points.
   (ii) $x'_{\mu} \frac{\partial f_t(\mu_0)}{\partial \mu} = 0$ a.s. iff $x_{\mu} = 0$ ($x_{\mu} \in \mathbb{R}^m$).
   (iii) $x'_{\lambda} \frac{\partial g(u_{0,t},\sigma^2_t;\theta_0)}{\partial \lambda} = 0$ a.s. iff $x_{\lambda} = 0$ ($x_{\lambda} \in \mathbb{R}^l$).

**Theorem 2:** Suppose Assumptions DGP, C1–C6, and N1–N5 hold. Then

\[ T^{1/2}(\hat{\theta}_T - \theta_0) \overset{d}{\to} N(0, \mathcal{I}(\theta_0)^{-1} \mathcal{J}(\theta_0) \mathcal{J}(\theta_0)^{-1}) \]

where matrices $\mathcal{I}(\theta_0)$ and $\mathcal{J}(\theta_0)$ are positive definite.
Results

What do we show? (repeated)

- The model we consider: A General Nonlinear AR–GARCH model

\[ y_t = f(y_{t-1}; \mu_0) + \sigma_t \varepsilon_t \]

\[ u_{0t} = y_t - f(y_{t-1}; \mu_0) \]

\[ \sigma_t^2 = g(u_{0,t-1}, \sigma_{t-1}^2; \theta_0) \]

- The true parameter \( \theta_0 \) belongs to some parameter space \( \Theta \)
- Feasible Gaussian log-likelihood (multiplied by \(-\frac{2}{T}\), modulo a constant)

\[ L_T(\theta) = T^{-1} \sum_{t=1}^{T} l_t(\theta), \quad l_t(\theta) = \log(h_t(\theta)) + \frac{u_t^2(\theta)}{h_t(\theta)} \]

- Quasi-maximum likelihood estimator for the parameter \( \theta_0 \) is

\[ \hat{\theta}_T = \arg \min_{\theta \in \Theta} L_T(\theta) \]

- What we show is
  - Strong consistency of the QML estimator (Thm 1)

\[ \hat{\theta}_T \longrightarrow \theta_0 \quad \text{a.s.} \]

  - Asymptotic normality of the QML estimator (Thm 2)

\[ T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, \mathcal{J}(\theta_0)^{-1}\mathcal{I}(\theta_0)\mathcal{J}(\theta_0)^{-1}) \]
Outline of the Presentation

- **Brief Introduction to (Refresher of) GARCH models**
  - Basic ARCH and GARCH & Extensions and Different Flavors of GARCH
  - Is there interest in these topics?

- **What we do in this paper**
  - Consistency and Asymptotic Normality of the QMLE
  - Previous Literature

- **Results**
  - The Nonlinear AR–GARCH model we consider
  - Consistency
    - Road Map of a Basic Proof & Details of some Complications
  - Asymptotic Normality
    - Road Map of a Basic Proof & Outline of some Complications
  - Some Remarks of Assumptions Made
  - Examples
    - Linear AR–GARCH & Nonlinear AR–GARCH

- **Conclusions**
  - Summary of what we did & Relation to existing literature
Some remarks of assumptions made

- **Data Generating Process**
  - Conditional Mean & Conditional Variance
    \[
    f(\ldots) = \sum_{j=1}^{p} a_j(y_{t-1}, \ldots, y_{t-p}; \mu) y_{t-j} + b(y_{t-1}, \ldots, y_{t-p}; \mu)
    \]
    \[
    g(u_{t-1}, \sigma^2_{t-1}; \theta) \text{ quite general nonlinear functions}
    \]
  - \((y_t, \sigma^2_t)\) Stat. & Erg. with \(E[|y_t|^{2r}] < \infty, E[\sigma^2_t] < \infty\) for some \(r > 0\).

- **Consistency**
  - Compact parameter space \(\Theta\)
  - \(f\) and \(g\) continuous w.r.t. \(\theta\)
  - Lipschitz condition on \(g\): \(|g(u, x_1; \theta) - g(u, x_2; \theta)| \leq \kappa|x_1 - x_2|\) (\(\kappa < 1\))
  - Identification conditions
    - \(f(y_{t-1}, \ldots, y_{t-p}; \mu) = f(y_{t-1}, \ldots, y_{t-p}; \mu_0)\) a.s. iff \(\mu = \mu_0\)
    - \(h^*_t(\mu_0, \lambda) = h^*_t(\mu_0, \lambda_0)\) a.s. iff \(\lambda = \lambda_0\)

- **Asymptotic Normality**
  - \(f\) and \(g\) twice continuously differentiable w.r.t. \(\theta\)
  - \(E[y_t^4] < \infty\) and \(E[\sigma^4_t] < \infty\)
  - Identification conditions
    - components of \(\frac{\partial f_t(\mu_0)}{\partial \mu}\) a.s. linearly independent; same for \(\frac{dh^*_t(\theta_0)}{d\lambda}\)

- **In addition many technical conditions**

- All assumptions rather innocuous & Similar to ones used in earlier
Example 1: Linear AR($p$)–GARCH(1,1)

- The model specification

\[ f (y_{t-1}, \ldots, y_{t-p}; \mu_0) = \phi_{0,0} + \phi_{0,1}y_{t-1} + \cdots + \phi_{0,p}y_{t-p} \]
\[ \sigma_t^2 = g (u_{0,t-1}, \sigma_{t-1}^2; \theta_0) = \omega_0 + \alpha_0 u_{0,t-1}^2 + \beta_0 \sigma_{t-1}^2 \]

- \[ u_{0,t} = y_t - (\phi_{0,0} + \sum_{j=1}^{p} \phi_{0,j}y_{t-j}) \]
- \[ \varepsilon_t \sim \text{IID}(0, 1) \text{ with } \varepsilon_t^2 \text{ having a non-degenerate distribution} \]

- Assumptions needed for Consistency
  - ‘strict stationarity of GARCH-part’ : \[ E[\ln(\beta_0 + \alpha_0 \varepsilon_t^2)] < 0 \]
  - ‘non-explosiveness of AR-part (root cond)’ : \[ 1 - \sum_{j=1}^{p} \phi_{0,j}z^j \neq 0, \ |z| \leq 1 \]

- Assumptions needed for Asymptotic Normality
  - ‘fourth moments of the GARCH-part’ : \[ E[(\beta_0 + \alpha_0 \varepsilon_t^2)^2] < 1 \]

- Importantly, same conditions as in previous literature (F&Z2004)
  - Our general nonlinear framework does not come at the cost of stronger assumptions
Example 2: AR($p$)–AGARCH(1,1)

- As in Ex. 1, but the AGARCH of Ding, Granger, and Engle (1993)

\[
f(y_{t-1}, \ldots, y_{t-p}; \mu_0) = \phi_{0,0} + \phi_{0,1}y_{t-1} + \cdots + \phi_{0,p}y_{t-p}
\]

\[
\sigma_t^2 = g(u_{0,t-1}, \sigma_{t-1}^2; \theta_0) = \omega_0 + \alpha_0(|u_{0,t-1}| - \gamma_0 u_{0,t-1})^2 + \beta_0 \sigma_{t-1}^2
\]

- Contains also so-called GJR–GARCH and TGARCH models (Glosten, Jaganathan, Runkle (1993); Zakoïan (1994))

- For this model, we only obtain a consistency result

- Assumption N2 ($g$ being cont. diff. w.r.t. all its parameters) violated due to the term $|u_{0,t-1}|$
  - Similar problem occurs in numerous other nonlinear GARCH models
  - Not a problem in pure GARCH models
Example 3: Nonlinear AR($p$)–GARCH(1,1)

- The model specification (for simplicity, $p = 1$)

$$f(y_{t-1}; \mu_0) = (\phi_{0,0} + \phi_{0,1}y_{t-1}) + (\psi_{0,0} + \psi_{0,1}y_{t-1}) F(y_{t-1}; \varphi_{0,1}, \varphi_{0,2})$$

$$\sigma_t^2 = g(u_{0,t-1}, \sigma_{t-1}^2; \theta_0) = \omega_0 + (\alpha_{0,1} + \alpha_{0,2} G(u_{0,t-1}; \gamma_{0,1}, \gamma_{0,2})) u_{0,t-1}^2 + \beta_0 \sigma_{t-1}^2$$

- $F$ and $G$: Nonlinear functions taking values in $[0, 1]$
  - cdf's of logistic distribution; or normal distr.; or anything rather general
- Functional-Coefficient AR model (Chen&Tsay, 1993, JASA), Smooth Transition AR (STAR) model (Teräsvirta, 1994, JASA)

- Assumptions needed for Consistency
  - 'strict stationarity of GARCH-part': mild log-moment condition
  - 'non-explosiveness of AR-part': root condition
  - 'identification of the nonlinear part': $\psi_{0,0} \neq 0$ or $\psi_{0,1} \neq 0$
  - 'S & E of DGP': $\varepsilon_t$ has a density positive on $\mathbb{R}$

- Assumptions needed for Asymptotic Normality
  - 'fourth moments of the GARCH-part': $E[(something)^2] < 1$

- Importantly, previously no results available for anything nonlinear
Assumption C6: Global identification conditions

(i) \( f(y_{t-1}, \ldots, y_{t-p}; \mu) = f(y_{t-1}, \ldots, y_{t-p}; \mu_0) \) a.s. if \( \mu = \mu_0 \)

(ii) \( h^*_t(\mu_0, \lambda) = \sigma^2_t \) a.s. if \( \lambda = \lambda_0 \)

- Showing that this condition implies identification is rather straightforward
- Showing that this condition actually holds is rather difficult in nonlinear cases
  - Very limited previous literature appears to be available

- For example, to verify C6(i)
  - Restrict \( f \) so that for every \( \mu \neq \mu_0 \) there exists a measurable set \( A \subset \mathbb{R}^p \) such that \( f(z; \mu) \neq f(z; \mu_0) \) for all \( z \in A \)
    - Long list of technical conditions
  - Show (using Markov chain techniques) that \( P\{(y_{t-1}, \ldots, y_{t-p}) \in A\} > 0 \)
    - Stationary distribution of \( y_t \) exists but is unknown

- In summary, ensuring global identification rather complicated
Outline of the Presentation

- **Brief Introduction to (Refresher of) GARCH models**
  - Basic ARCH and GARCH & Extensions and Different Flavors of GARCH
  - Is there interest in these topics?
- **What we do in this paper**
  - Consistency and Asymptotic Normality of the QMLE
  - Previous Literature
- **Results**
  - The Nonlinear AR–GARCH model we consider
  - Consistency
    - Road Map of a Basic Proof & Details of some Complications
  - Asymptotic Normality
    - Road Map of a Basic Proof & Outline of some Complications
  - Some Remarks of Assumptions Made
  - Examples
    - Linear AR–GARCH & Nonlinear AR–GARCH
- **Conclusions**
  - Summary of what we did & Relation to existing literature
Summary of what we did

- Reviewed (nonlinear AR–)GARCH models
  - Frequently used in applied work with financial time series

- Considered asymptotic estimation theory in nonlinear AR–GARCH models
  - Discussed some challenges in proving Consistency and Asymptotic Normality
  - Looked at some concrete examples

- Importance of our work?
  - Our paper is the first one to develop asymptotic estimation theory for nonlinear AR–GARCH models
Conclusions

Relation to existing literature (repeated)

- Asymptotic estimation theory for these models
  - Linear ‘pure’ GARCH
    - Hall and Yao (2003, *Econometrica*)
  - Linear AR–GARCH
    - Ling and McAleer (2003, *Econometric Theory*)
    - Francq and Zakoïan (2004, *Bernoulli*)
    - Ling (2007, *Journal of Econometrics*)
    - Lange, Rahbek, and Jensen (WP) (ARCH)
  - Nonlinear ‘pure’ GARCH
  - Nonlinear AR–GARCH
    - Meitz and Saikkonen (WP, 31 May 2008; R & R at *ET*)
THANK YOU!