Volatility prediction under misspecification

Guangyuan Ji[1] and Genaro Sucarrat[2]

This version: 8th March 2024

(Please do not distribute)

Abstract

The volatility models used in practice are unlikely to equal the Data Generating Process (DGP). Accordingly, models that are valid under misspecification is of great importance. We establish exact, general and mild conditions under which a large class of volatility prediction specifications exists. Crucially, the specifications within the class generate volatility predictions that are weakly identified for volatility under misspecification. Next, we derive a consistent and asymptotically normal estimator that is valid under dependence of unknown form. The volatility prediction specifications we consider in more detail are modifications of the log-ARCH-X model. The specifications are highly interpretable and versatile, and accommodate zero returns (in contrast to the classic log-ARCH specification), short-term and long-term persistence, asymmetry, volatility proxies and additional covariates. Since the volatility specifications are in logs, inference is standard under nullity of the parameters, and positivity of the volatility predictions are guaranteed. In our simulation experiments the predictions are both unbiased and identified for the benchmark model, whereas in our empirical illustration the volatility predictions compare well with those of the benchmark volatility model.

JEL Classification: C01, C13, C14, C22, C53, C58

Keywords: Volatility, log-volatility, prediction, misspecification, ARCH, log-ARCH, log-ARCH-X

1 Introduction

Let y_t denote the variable of interest to predict. In empirical applications, an important and common task is that of estimating a model $f(\boldsymbol{\theta}, \boldsymbol{x}_{t-1})$ of y_t, where $\boldsymbol{x}_t = [x_{1t}, \ldots, x_{Kt}]'$ is a set of predictors, and $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_L]'$ is a finite-dimensional parameter. Arguably, the most common example of $f(\boldsymbol{\theta}, \boldsymbol{x}_{t-1})$ is the linear prediction

$$\boldsymbol{\theta}'\boldsymbol{x}_{t-1} = \theta_1 x_{1,t-1} + \cdots + \theta_K x_{K,t-1}. \tag{1}$$

Often, it is assumed that $f(\boldsymbol{\theta}, \boldsymbol{x}_{t-1})$ coincides with the conditional expectation, i.e. $E_{t-1}(y_t) = f(\boldsymbol{\theta}, \boldsymbol{x}_{t-1})$, where the subscript $t-1$ is shorthand for conditioning on past observables. This, however, is a very restrictive and unrealistic assumption, since it is unlikely to hold in practice. More realistically, the specification $f(\boldsymbol{\theta}, \boldsymbol{x}_{t-1})$ is a prediction of y_t, which most likely differs from $E_{t-1}(y_t)$. The linear prediction $\boldsymbol{\theta}'\boldsymbol{x}_{t-1}$ is commonly referred to as the OLS prediction or OLS projection, since the Ordinary Least Squares (OLS) estimator is consistent for $\boldsymbol{\theta}$ under suitable

[1]Department of Economics, BI Norwegian Business School, Nydalsveien 37, 0484 Oslo, Norway. Webpage: https://www.bi.edu/about-bi/employees/department-of-economics/guangyuan-ji/. Corresponding author.

[2]Department of Economics, BI Norwegian Business School, Nydalsveien 37, 0484 Oslo, Norway. Webpage: https://www.sucarrat.net/.

assumptions. When this is the case, the OLS prediction is a valid statistical representation of the true but unknown Data Generating Process (DGP), even though it is *misspecified* for $E_{t-1}(y_t)$.

Here, the focus is on *misspecified* predictions of volatility. Let r_t denote a financial return (possibly mean corrected), and let $r_t^2 = \sigma_t^2 \eta_t^2$ where $\sigma_t^2 = E_{t-1}(r_t^2)$. We will refer to σ_t^2 as the true volatility, and η_t^2 as the true squared innovation.[3] So the specification that governs σ_t^2 is the DGP of volatility. Note that, since $\sigma_t^2 = E_{t-1}(r_t^2)$, it follows that $E(\eta_t^2) = E(r_t^2/\sigma_t^2) = 1$ (identification). Let h_t denote a prediction of σ_t^2. A main contribution of this paper is to establish the exact conditions under which a large class of predictions h_t satisfy

$$r_t^2 = h_t z_t^2, \qquad \ln h_t = \boldsymbol{\theta}' \boldsymbol{x}_{t-1}, \qquad \boldsymbol{\theta} = [\theta_1, \dots, \theta_K]' \qquad \text{and} \qquad E(z_t^2) = 1 \tag{2}$$

for all t. We refer to $\ln h_t$ as the log-volatility prediction, and h_t can be misspecified in the sense that we can have $h_t \neq \sigma_t^2$. The property $E(z_t^2) = 1$ ensures h_t is a prediction of σ_t^2 in the sense of Sucarrat (2021), i.e. h_t is weakly identified for σ_t^2. Because if $E(z_t^2)$ were unequal to 1, then it would not be at the same scale-level as σ_t^2. A second main contribution of our paper is to establish the exact conditions under which an estimator of $\boldsymbol{\theta}$ is consistent and asymptotically normal.

The rest of the paper is organised as follows. The next section, Section 2, contains our theoretical results. In Section 3 we consider a specific class of log-volatility specifications that is both highly interpretable and highly versatile. The class is a heterogeneous version of the log-ARCH-X model, and constitutes a generalisation of the volatility model proposed by Sucarrat and Escribano (2012). Section 4 contains an empirical illustration of our results. Finally, Section 5 concludes.

2 Theoretical results

2.1 Structure, existence and consistency

Let $c > 0$ denote a real-valued fixed (i.e. non-stochastic) and known scalar that is chosen by the researcher. A straightforward example is $c = 1$. The motivation for c is to avoid taking the logarithm on zero values. Next, let

$$y_t = \begin{cases} \ln r_t^2 & \text{if } r_t \neq 0 \\ \ln c & \text{if } r_t = 0 \end{cases} \tag{3}$$

and let $\boldsymbol{x}_t$ denote a $(K \times 1)$ vector of covariates. Throughout, we rely on the following assumptions on r_t^2 and $\boldsymbol{x}_t$.

A 1 *$\{(r_t^2, \boldsymbol{x}_t')'\}$ is strictly stationary and ergodic, and the first entry of $\boldsymbol{x}_t$ is a constant equal to 1 for all t.*

A 2 *Moments: a) $0 < E(y_t^2) < \infty$ and b) $E(\boldsymbol{x}_t \boldsymbol{x}_t')$ has full rank.*

[3]We will refer to both σ_t^2 and σ_t as volatility, since the latter is obtained as a straightforward transformation of the former.

Notice that a direct consequence of A 1 is that also $\{y_t\}$ is strictly stationary and ergodic. Define the OLS estimator made up of y_t and $\boldsymbol{x}_{t-1}$:

$$\widehat{\boldsymbol{\theta}}^* \equiv \left[\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{t-1}y_t\right]. \tag{4}$$

A 1 and A 2 ensure this expression converges almost surely to a point we denote by $\boldsymbol{\theta}_0^* \equiv [\theta_{01}^*, \theta_{02}^*, \ldots, \theta_{0K}^*]'$. The following result, which summarises some well-known properties, is an intermediate but necessary step towards our main existence result further below (Proposition 2).

Proposition 1 *Suppose A 1 – A 2 hold. Then*

a) there exists a linear prediction $\boldsymbol{\theta}_0^{*\prime}\boldsymbol{x}_{t-1}$ of y_t with $\boldsymbol{\theta}_0^* \equiv E(\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}')^{-1}E(\boldsymbol{x}_{t-1}y_t)$ and $u_t \equiv y_t - \boldsymbol{\theta}_0^{*\prime}\boldsymbol{x}_{t-1}$;

b) the linear prediction $\boldsymbol{\theta}_0^{*\prime}\boldsymbol{x}_{t-1}$ is the best linear predictor of y_t in that it minimises the unconditional Mean Squared Error(MSE) $E(u_t^2)$;

c) the OLS estimator in (4) satisfies $\widehat{\boldsymbol{\theta}}^* \underset{p}{\to} \boldsymbol{\theta}_0^*$;

d) $E(u_t) = 0$ for all t.

Proof. In a linear regression context, the proofs of similar results are well-known, see e.g. Section 2.9 in Hayashi (2000). Nevertheless, for completeness we give the proof of Proposition 1 in Appendix A.1.

It is important to note that, in general, $\exp(\boldsymbol{\theta}_0^{*\prime}\boldsymbol{x}_{t-1})$ will not be identified as a volatility prediction, since in general $E\big(r_t^2/\exp(\boldsymbol{\theta}_0^{*\prime}\boldsymbol{x}_{t-1})\big)$ will not equal 1. However, as we now prove under mild and general assumptions, there exists a correction that transforms $\exp(\boldsymbol{\theta}_0^{*\prime}\boldsymbol{x}_{t-1})$ into an identified volatility prediction. Define the function $m_t(\boldsymbol{\theta}^*) \equiv r_t^2/\exp(\boldsymbol{\theta}^{*\prime}\boldsymbol{x}_{t-1})$, so that $m_t(\boldsymbol{\theta}^*) : \mathbb{R}^K \to \mathbb{R}^1$ with $\boldsymbol{\theta} \in \boldsymbol{\Theta}^*$. Next, assume

A 3 $0 < E(z_t^{*2}) < \infty$, *where* $z_t^{*2} = m_t(\boldsymbol{\theta}_0^*) = r_t^2/\exp(\boldsymbol{\theta}_0^{*\prime}\boldsymbol{x}_{t-1})$,

and define

$$z_t^2 \equiv \frac{z_t^{*2}}{E(z_t^{*2})} \qquad \text{and} \qquad \boldsymbol{\theta}_0 \equiv [\theta_{01}, \theta_{02}, \ldots, \theta_{0K}]', \tag{5}$$

where

$$\theta_{01} = \theta_{01}^* + \ln E(z_t^{*2}) \qquad \text{and} \qquad \theta_{0j} = \theta_{0j}^* \qquad \text{for} \qquad j = 2, \ldots, K.$$

The next result, Proposition 2, ensures the log-volatility prediction $\ln h_t = \boldsymbol{\theta}_0'\boldsymbol{x}_{t-1}$ exists, that $h_t = \exp(\boldsymbol{\theta}_0'\boldsymbol{x}_{t-1})$ is weakly identified as volatility in the sense that $E(r_t^2/h_t) = 1$, and that the correction $E(z_t^{*2})$ can be consistently estimated. The latter implies that also $\boldsymbol{\theta}_0$ can be consistently estimated. For this, we need the following additional assumption.

A 4 *There exists a compact set $\boldsymbol{\Theta}^*$ about $\boldsymbol{\theta}_0^*$ such that* $E\Big(\sup_{\boldsymbol{\theta}^* \in \boldsymbol{\Theta}^*} \frac{r_t^2}{\exp(\boldsymbol{\theta}^{*\prime}\boldsymbol{x}_{t-1})}\Big) < \infty$.

Proposition 2 *Suppose A 3 and A 4 hold in addition to the assumptions of Proposition 1. Then*

a) the log-volatility prediction $\ln h_t = \boldsymbol{\theta}_0' \boldsymbol{x}_{t-1}$ *exists for all* t *(existence);*

b) $E(r_t^2/h_t) = E(z_t^2) = 1$ *for all* t *(weak identification);*

c) $\widehat{E}(z_t^{*2}) \underset{p}{\to} E(z_t^{*2})$, *where* $\widehat{E}(z_t^{*2}) = \left(T^{-1}\sum_{t=1}^{T}\widehat{z}_t^{*2}\right)$ *and* $\widehat{z}_t^{*2} = r_t^2/\exp(\widehat{\boldsymbol{\theta}}^{*\prime}\boldsymbol{x}_{t-1})$ *(consistency of* $\widehat{E}(z_t^{*2})$*);*

d) $\widehat{\boldsymbol{\theta}} \underset{p}{\to} \boldsymbol{\theta}_0$ *with* $\widehat{\boldsymbol{\theta}} = [\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_K]'$, *where* $\widehat{\theta}_1 = \widehat{\theta}_1^* + \ln \widehat{E}(z_t^{*2})$ *and* $\widehat{\theta}_j = \widehat{\theta}_j$ *for* $j = 2, \ldots, K$ *(consistency of* $\widehat{\boldsymbol{\theta}}_0$*)*

Proof. See Section A.2.

2.2 Asymptotic normality

To derive the joint distribution of our estimator, we rely on the following additional assumption.

A 5 *The joint distribution of the* $(K+1)\times 1$ *vector* $\boldsymbol{w}_t$ *satisfies* $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{w}_t \underset{d}{\to} N(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{w}})$, *where* $\boldsymbol{w}_t \equiv \left[\boldsymbol{s}_t, z_t^{*2} - E(z_t^{*2}\right]'$ *with* $\boldsymbol{s}_t \equiv u_t \boldsymbol{x}_{t-1}$.

Note that the assumption implies $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{s}_t \underset{d}{\to} N(\boldsymbol{0}, \boldsymbol{S})$, where $\boldsymbol{S}$ is the first $K \times K$ elements in $\boldsymbol{\Sigma}_{\boldsymbol{w}}$. This leads to the following results.

Proposition 3 (asymptotic normality of $\widehat{\boldsymbol{\theta}}^*$) *Suppose A 5 holds in addition to the assumptions of Proposition 2. Then* $\sqrt{T}(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^*) \underset{d}{\to} N(\boldsymbol{0}, \boldsymbol{\Sigma}^*)$ *with* $\boldsymbol{\Sigma}^* = E(\boldsymbol{x}_t\boldsymbol{x}_t')^{-1}\,\boldsymbol{S}\,E(\boldsymbol{x}_t\boldsymbol{x}_t')^{-1}$.

Proof. See Section A.3.

The main use of this result is that it enables inference and variable selection ("machine learning") on the predictive power of $\boldsymbol{x}_{t-1}$. This is useful, since variable selection can be undertaken before the correction $E(z_t^{*2})$ is applied. Note that the form of the covariance matrix $\boldsymbol{S}$ depends on the properties of $\{\boldsymbol{s}_t\}$. There are three main cases: (i) $\{\boldsymbol{s}_t\}$ is homoscedastic and not autocorrelated, (ii) $\{\boldsymbol{s}_t\}$ is heteroscedastic but not autocorrelated, and (iii) $\{\boldsymbol{s}_t\}$ is both heteroscedastic and autocorrelated. In case (i), the ordinary OLS covariance matrix can be used for inference. In case (ii), a heteroscedascticity robust covariance matrix can be used. In case (iii), a Heteroscedasticity and Autocorrelation (HAC) robust covariance matrix, e.g. that of Newey and West (1987), can be used. In other words, numerical methods for all three cases are widely available in public software packages. The next proposition is an intermediate result, which is needed to establish the asymptotic normality result of $\widehat{\boldsymbol{\theta}}$ that succeeds it.

Proposition 4 (asymptotic normality of $\widehat{\boldsymbol{\phi}}$) *Let* $\widehat{\boldsymbol{\phi}} = [\widehat{\boldsymbol{\theta}}^*, \widehat{\tau}]'$, *where* $\widehat{\tau} \equiv \ln\left(T^{-1}\sum_{t=1}^{T}\widehat{z}_t^{*2}\right)$. *Suppose the assumptions of Proposition 3 hold. Then* $\sqrt{T}(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) \underset{d}{\to} N(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{\phi}_0})$, *where* $\boldsymbol{\Sigma}_{\boldsymbol{\phi}_0} = \boldsymbol{G}\boldsymbol{\Sigma}_w\boldsymbol{G}'$ *with* $\boldsymbol{G} = \begin{bmatrix} E(\boldsymbol{x}_t\boldsymbol{x}_t')^{-1} & \boldsymbol{0} \\ -\frac{E(z_t^{*2}\boldsymbol{x}_{t-1}')}{E(z_t^{*2})E(\boldsymbol{x}_t\boldsymbol{x}_t')} & E(z_t^{*2})^{-1} \end{bmatrix}$.

Proof. See Section A.4.

Corollary 1 (asymptotic normality of $\widehat{\boldsymbol{\theta}}$) $\sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \underset{d}{\to} N(\mathbf{0}, \boldsymbol{A}\boldsymbol{\Sigma}_{\phi_0}\boldsymbol{A})$, *where*

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & \dots & 0 & 1 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

is a $K \times (K+1)$ matrix.

Proof. See Section A.5.

3 The heterogenous log-ARCH-X prediction

3.1 Specification

The log-ARCH-X class of specifications provides a very general, flexible and successful approach that ensures the volatility prediction h_t are non-negative in empirical practice. Moreover, in comparison with ARCH-X specifications, inference is standard under nullity, since in this case the parameters are not on the border of the feasible parameter space under nullity. The heterogenous log-ARCH-X model that we consider is given by

$$\ln h_t = \omega + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{j \in q} \beta_j \ln \mathrm{EqWMA}_{j,t-1} + \sum_{k=1}^{r} \lambda_k y_{t-k} I_{\{r_{t-i}<0\}} + \sum_{l=1}^{s} \delta_l x_{l,t-1}, \tag{6}$$

so

$$\boldsymbol{\theta} = [\omega, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q, \lambda_1, \dots, \lambda_r, \delta_1, \dots, \delta_s]'.$$

The p is the log-ARCH order. EqWMA is short for Equally Weighted Moving Average, and

$$\mathrm{EqWMA}_{j,t-1} \equiv \frac{(r_{t-1}^2 + \cdots + r_{t-j}^2)}{j}, \qquad j \in \{1, 2, \dots, \}. \tag{7}$$

In other words, the $\mathrm{EqWMA}_{j,t-1}$'s can be viewed as a lagged volatility proxies made up of past values of r_t^2. They can be viewed as proxies of omitted log-GARCH terms, e.g. $\ln \sigma_{t-1}^2$. However, it should be noted that they can also be given an additional interpretation of interest. Specifically, if r_t is a daily financial return, and if the returns are recorded over weekdays only, then $\mathrm{EqWMA}_{5,t-1}$, $\mathrm{EqWMA}_{20,t-1}$ and $\mathrm{EqWMA}_{60,t-1}$ can be interpreted as the "weekly", "monthly" and "quarterly" volatilities, respectively. The log-proxies thus provide great flexibility in modelling the persistence of log-volatility. Also, note that $\ln \mathrm{EqWMA}_{j,t-1} = y_{t-1}$, i.e. the ARCH(1) term, when $j = 1$. The r is the number of logarithmic asymmetry terms (i.e., "leverage") analogous to those of Glosten et al. (1993). So $I_{r_{t-k}<0}$ is an indicator function equal to 1 if $r_{t-k} < 0$ and 0 otherwise. Finally, s is the number of covariates. Of course, additional volatility proxies, e.g. terms made up of Realised Volatility (RV) or the range, can be included here. If h_t were equal to σ_t^2, then many well-known specifications would be contained in (6). Examples include the constant variance model, the log-ARCH(p) model, the logarithmic version of the HAR model (see Corsi 2009, Corsi et al. 2012), and the log-RV model, where RV is short for Realised Volatility.

3.2 h_t as an approximation to the GARCH(1,1)

Here we study how well a misspecified volatility prediction h_t contained in (6) approximates the true volatility σ_t^2 when it is governed by a GARCH(1,1). The DGP is given by the standard GARCH(1,1) model with normal innovations:

$$\text{DGP:} \qquad r_t = \sigma_t \eta_t, \qquad \eta_t \overset{iid}{\sim} N(0,1), \qquad \sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2. \tag{8}$$

The comparison is made under two parametric setups:

- High persistence: $\alpha = 0.05$, $\beta = 0.90$, $\omega = 1 - \alpha - \beta$
- Low persistence: $\alpha = 0.30$, $\beta = 0.20$, $\omega = 1 - \alpha - \beta$

The specifications of h_t that we consider are:

$$\text{Model 1:} \quad \ln h_t = \omega + \alpha \ln y_{t-1}, \tag{9}$$
$$\text{Model 2:} \quad \ln h_t = \omega + \alpha \ln y_{t-1} + \beta_5 \ln \mathrm{EqWMA}_{5,t-1}, \tag{10}$$
$$\text{Model 3:} \quad \ln h_t = \omega + \alpha \ln y_{t-1} + \sum_{j \in q} \beta_j \ln \mathrm{EqWMA}_{j,t-1}, q = \{5, 20, 60\}, \tag{11}$$
$$\text{Model 4:} \quad \ln h_t = \omega + \alpha \ln y_{t-1} + \sum_{j \in q} \beta_j \ln \mathrm{EqWMA}_{j,t-1}, q = \{5, 20, 60, 120\}. \tag{12}$$

The first is a log-ARCH(1) specification. In the high-persistence setup, we would expect it to do poorly. Models 2 – 4 contain one or more of the weekly ($j = 5$), monthly ($j = 20$), quarterly ($j = 60$) and half-yearly ($j = 120$) log-volatility proxies. To measure the discrepancy between h_t and σ_t^2, we compute three statistics:

$$\text{Mean:} \quad \frac{1}{T} \sum_{t=1}^{T} h_t - \sigma_t^2$$
$$\text{MAE:} \quad \frac{1}{T} \sum_{t=1}^{T} \left| h_t - \sigma_t^2 \right|$$
$$\text{MSE:} \quad \frac{1}{T} \sum_{t=1}^{T} \left(h_t - \sigma_t^2 \right)^2$$

In the high-persistence setup, Tables 1, the overall pattern is clear according to all three statistics: The more log-volatility proxies, the better fit. The exception is Model 4, which performs slightly worse than Model 3. This shows that using more past information does not always improve the forecast precision. In the low persistence setup, Table 2, Model 1 performs best according to Mean, whereas Model 2 performs best according to MAE and MSE.

4 Empirical illustration

In this section, we use daily financial close-to-close returns to illustrate our results. We fit several specifications of h_t with alternative variable structures. The benchmark model to beat

Table 1: Results under high persistence

Models	Mean	Absolute	MSE
Model 1	0.0008	0.1694	0.0514
Model 2	0.0022	0.1137	0.0236
Model 3	**0.0006**	**0.0530**	**0.0055**
Model 4	-0.0011	0.0549	0.0057

Table 2: Results under low persistence

Models	Mean	Absolute	MSE
Model 1	**0.0000**	0.3045	0.3069
Model 2	0.0015	**0.2734**	**0.2645**
Model 3	-0.0002	0.3096	0.2948
Model 4	-0.0024	0.3120	0.2971

in our comparisons is the GARCH(1,1) specification. The financial returns in our study are the S&P500 index, the Shanghai Shenzhen CSI Index (SHSZ300), the Apple (AAPL) stock and the Equinor (EQNR) stock. These provide four fairly different financial returns. The S&P500 is a key indicator of the US stock market, comprising 500 leading companies and covering approximately 80% of available market capitalization. The SHSZ300 is a significant benchmark for the Chinese stock market. Apple is a major player in the global electronics market and a constituent of the S&P500. Equinor ASA, a mjor player in the energy sector, especially in oil and gas production, represents a significant portion of the Norwegian stock market. In total, these returns provide insights into market and sector-specific performances, and contribute to a comprehensive analysis of our models's performance. The source of all of our data is Bloomberg.

The models we include in our comparison are:

$$\text{No. 1:}\quad \ln h_t = \omega, \tag{13}$$

$$\text{No. 2:}\quad h_t = \omega + \alpha r_t^2 + \sigma_{t-1}^2, \tag{14}$$

$$\text{No. 3:}\quad \ln h_t = \omega + \alpha y_{t-1}, \tag{15}$$

$$\text{No. 4:}\quad \ln h_t = \omega + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{j\in q} \beta_j \ln \mathrm{EqWMA}_{j,t-1} + \lambda_k y_{t-k} I_{\{r_{t-i}<0\}}, \tag{16}$$

$$\text{No. 5:}\quad \ln h_t = \omega + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{j\in q} \beta_j \ln \mathrm{EqWMA}_{j,t-1} + \lambda_k y_{t-k} I_{\{r_{t-i}<0\}} + \sum_{l=1}^{s} \delta_l x_{l,t-1}, \tag{17}$$

where $p = 5$ and $q = \{5, 20, 60, 120\}$. Model No. 1 is a constant model in logs. Model No. 2 is a GARCH(1,1). Its parameters are estimated with the standard Quasi Maximum Likelihood Estimator (QMLE). Model No. 3 is a log-ARCH(1). Model No. 4 is a log-ARCH(5) specification augmented with log-volatility proxies and asymmetry terms. Finally, model No. 5 contains two covariates, $x_{1,t-1}$ and $x_{2,t-1}$, in addition to the terms of model No. 4. The first covariate is a measure of the volume on day $t-1$, whereas the second is log-range between the maximum and minimum prices during the day (i.e. a volatility proxy). Note that both series have been transformed to approximate stationarity by subtracting a 20 day moving average.

Tables 3 – 6 contains the estimation results. For all four returns the ARCH and GARCH coefficients in the GARCH(1,1) model are in the usual range (0.9 for the GARCH coefficient and 0.05 for the ARCH coefficient). For model No. 3, the log-ARCH(1) specification, the log-ARCH(1) is significant at the usual significance levels for all four returns. When additional terms are added in models No. 4 and No. 5, however, it is no longer always significant. Similarly, additional lags of the log-ARCH terms are not always significant. The $\ln \text{EqWMA}_{j,t-1}$ terms are always significant at usual significance levels when $j = 20$, i.e. the monthly term. However, they are not always significant for $j = 5$, $j = 60$ and $j = 120$. For SP500, for example, they are significant for $j = 5$, $j = 20$ and $j = 120$. For EQNR, by contrast, they are only significant for $j = 20$ and $j = 60$. This suggests the persistence dynamics may differ substantially.

Table 7 contains diagnostics and goodness-of-fit measures of the models. For the log-volatility models to be weakly identified as volatility models, $E(z_t^2)$ should equal 1. As is clear from the diagnostics column labelled $E(\widehat{z}_t^{\,2})$, the results show that all models are identified. We use two measures of fit. The first is MSE, which is computed as $T^{-1}\sum_{t=1}^{T}(r_t^2 - \widehat{h}_t)^2$. The second is the QLIKE, which is computed as $T^{-1}\sum_{t=1}^{T} r_t^2/\widehat{h}_t + \ln \widehat{h}_t$. According to MSE, model No. 5 provides the best fit for SP500 and EQNR. For SHSZ300, by contrast, the GARCH(1,1) provides the best fit, whereas for AAPL it is the log-ARCH(1) that provides the best fit. The results according to QLIKE are clearer. There, model No. 5 performs best for SHSZ300, AAPL and EQNR. For SP500 it is the GARCH(1,1) that performs best. It is slightly surprising that it does not perform best for more than one return, since the GARCH(1,1) estimates are obtained by minimising QLIKE.

Variable	No. 1	No. 2	No. 3	No. 4	No. 5
ω	0.0378	0.0171	0.3845	0.1654	0.1330
	(0.0124)	(0.0000)	(0.0329)	(0.0505)	(0.0506)
σ^2_{t-1}		0.8859			
		(0.0001)			
y_{t-1}		0.1001	0.1204	-0.0243	-0.0444
		(0.0001)	(0.0108)	(0.0147)	(0.0152)
y_{t-2}				-0.0014	0.0004
				(0.0121)	(0.0121)
y_{t-3}				0.0254	0.0288
				(0.0121)	(0.0121)
y_{t-4}				-0.0015	0.0037
				(0.0121)	(0.0121)
y_{t-5}				0.0193	0.0271
				(0.0121)	(0.0122)
$y_{t-1}I_{\{r_{t-1}<0\}}$				-0.0298	-0.0303
				(0.0170)	(0.0169)
$\ln \mathrm{EqWMA}_{5,t-1}$				0.2619	0.1800
				(0.0490)	(0.0511)
$\ln \mathrm{EqWMA}_{20,t-1}$				0.3720	0.4482
				(0.0665)	(0.0678)
$\ln \mathrm{EqWMA}_{60,t-1}$				0.0989	0.1172
				(0.0964)	(0.0963)
$\ln \mathrm{EqWMA}_{120,t-1}$				0.2374	0.2369
				(0.0790)	(0.0789)
$x_{1,t-1}$					0.3871
					(0.1360)
$x_{2,t-1}$					17.7203
					(5.0066)

Table 3: Coefficient estimates with standard errors in parentheses for models of SP500 return volatility

Variable	No.1	No.2	No.3	No.4	No.5
ω	0.0432	0.0119	1.0069	0.2422	0.2175
	(0.0242)	(0.0000)	(0.0392)	(0.0761)	(0.0762)
σ^2_{t-1}		0.9335			
		(0.0001)			
y_{t-1}		0.0640	0.0811	-0.0262	-0.0458
		(0.0001)	(0.0148)	(0.0208)	(0.0214)
y_{t-2}				0.0065	0.0133
				(0.0165)	(0.0165)
y_{t-3}				0.0207	0.0261
				(0.0165)	(0.0165)
y_{t-4}				0.0336	0.0381
				(0.0165)	(0.0165)
y_{t-5}				0.0069	0.0123
				(0.0167)	(0.0167)
$y_{t-1}I_{\{r_{t-1}<0\}}$				0.0187	0.0097
				(0.0270)	(0.0269)
$\ln \mathrm{EqWMA}_{5,t-1}$				-0.1143	-0.1783
				(0.0690)	(0.0720)
$\ln \mathrm{EqWMA}_{20,t-1}$				0.8101	0.8439
				(0.1031)	(0.1042)
$\ln \mathrm{EqWMA}_{60,t-1}$				-0.0844	-0.0581
				(0.1472)	(0.1470)
$\ln \mathrm{EqWMA}_{120,t-1}$				0.2122	0.2158
				(0.1212)	(0.1208)
$x_{1,t-1}$					-0.6249
					(0.1502)
$x_{2,t-1}$					26.5274
					(4.5016)

Table 4: Coefficient estimates with standard errors in parentheses for models of SHSZ300 return volatility

Variable	No.1	No.2	No.3	No.4	No.5
ω	0.1131	0.0247	1.8928	0.2712	0.2033
	(0.0293)	(0.0002)	(0.0242)	(0.0718)	(0.0717)
σ^2_{t-1}		0.9537			
		(0.0001)			
y_{t-1}		0.0452	0.1790	0.0183	-0.0113
		(0.0001)	(0.0106)	(0.0155)	(0.0160)
y_{t-2}				0.0099	0.0074
				(0.0124)	(0.0124)
y_{t-3}				0.0156	0.0234
				(0.0123)	(0.0123)
y_{t-4}				0.0142	0.0268
				(0.0123)	(0.0124)
y_{t-5}				0.0205	0.0335
				(0.0123)	(0.0123)
$y_{t-1}I_{\{r_{t-1}<0\}}$				0.0327	0.0315
				(0.0203)	(0.0203)
$\ln \text{EqWMA}_{5,t-1}$				0.1117	0.0008
				(0.0420)	(0.0446)
$\ln \text{EqWMA}_{20,t-1}$				0.2454	0.3310
				(0.0520)	(0.0531)
$\ln \text{EqWMA}_{60,t-1}$				0.1368	0.1684
				(0.0875)	(0.0874)
$\ln \text{EqWMA}_{120,t-1}$				0.4083	0.4024
				(0.0820)	(0.0818)
$x_{1,t-1}$					0.2918
					(0.0741)
$x_{2,t-1}$					7.3143
					(1.9257)

Table 5: Coefficient estimates with standard errors in parentheses for models of AAPL return volatility

Variable	No.1	No.2	No.3	No.4	No.5
ω	0.0476	0.0365	1.3044	0.3387	0.3044
	(0.0259)	(0.0001)	(0.0272)	(0.0710)	(0.0713)
σ^2_{t-1}		0.9370			
		(0.0001)			
y_{t-1}		0.0536	0.1329	-0.0059	-0.0227
		(0.0001)	(0.0132)	(0.0194)	(0.0199)
y_{t-2}				0.0153	0.0129
				(0.0154)	(0.0154)
y_{t-3}				0.0146	0.0173
				(0.0153)	(0.0152)
y_{t-4}				0.0066	0.0115
				(0.0152)	(0.0153)
y_{t-5}				0.0269	0.0329
				(0.0152)	(0.0153)
$y_{t-1}I_{\{r_{t-1}<0\}}$				0.0674	0.0637
				(0.0257)	(0.0257)
$\ln \text{EqWMA}_{5,t-1}$				0.0559	0.0047
				(0.0555)	(0.0570)
$\ln \text{EqWMA}_{20,t-1}$				0.3223	0.3735
				(0.0777)	(0.0788)
$\ln \text{EqWMA}_{60,t-1}$				0.3808	0.3940
				(0.1205)	(0.1204)
$\ln \text{EqWMA}_{120,t-1}$				0.0283	0.0295
				(0.0992)	(0.0991)
$x_{1,t-1}$					0.1868
					(0.0805)
$x_{2,t-1}$					7.1866
					(3.3129)

Table 6: Coefficient estimates with standard errors in parentheses for models of EQNR return volatility

Asset	Model	$E(\hat{z}_t^2)$	MSE	QLIKE
SPX	No.1	1.0011	21.2797	1.2720
	No.2	0.9995	16.2942	**0.8314**
	No.3	1.0000	20.8571	1.1786
	No.4	1.0000	17.4157	0.8515
	No.5	1.0000	**16.0721**	0.8388
SHSZ300	No.1	1.0007	42.3892	1.9770
	No.2	1.0015	**37.4145**	1.6783
	No.3	1.0000	41.7368	1.9338
	No.4	1.0000	37.6204	1.6830
	No.5	1.0000	37.9433	**1.6721**
AAPL	No.1	1.0017	1378.6830	2.9926
	No.2	0.9999	1370.6587	2.7377
	No.3	0.9997	**1367.4685**	2.9295
	No.4	0.9997	1373.4137	2.7242
	No.5	0.9997	1387.7167	**2.6883**
EQNR	No.1	1.0006	91.9901	2.3318
	No.2	1.0015	81.4642	2.1392
	No.3	0.9993	89.9500	2.2795
	No.4	0.9991	74.5133	2.1192
	No.5	0.9991	**74.0925**	**2.1122**

Table 7: Diagnostics and goodness-of-fit measures

5 Conclusions

The volatility models used in practice are unlikely to equal the Data Generating Process (DGP). Accordingly, models that are valid under misspecification is of great importance. We establish exact, general and mild conditions under which a large class of volatility prediction specifications exists. Crucially, the specifications within the class generate volatility predictions that are weakly identified for volatility under misspecification. Next, we derive a consistent and asymptotically normal estimator that is valid under dependence of unknown form. The volatility prediction specifications we consider in more detail are modifications of the log-ARCH-X model. The specifications are highly interpretable and versatile, and accommodate zero returns (in contrast to the classic log-ARCH specification), short-term and long-term persistence, asymmetry, volatility proxies and additional covariates. Since the volatility specifications are in logs, inference is standard under nullity of the parameters, and positivity of the volatility predictions are guaranteed. In our simulation experiments the predictions are both unbiased and identified for the benchmark model, whereas in our empirical illustration the volatility predictions compare well with those of the benchmark volatility model.

References

Corsi, F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics 7*, 174–196.

Corsi, F., F. Audrino, and R. Reno (2012). HAR Modeling for Rrealized Volatility Forecasting. In L. Bauwens, C. Hafner, and S. Laurent (Eds.), *Handbook of Volatility Models and Their Applications*, pp. 363–382. New Jersey: Wiley.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance 48*, 1779–1801.

Newey, W. and K. West (1987). A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica 55*, 703–708.

Sucarrat, G. (2021). Identification of Volatility Proxies as Expectation of Squared Financial Return. *Forthcoming in International Journal of Forecasting*. DOI: https://doi.org/10.1016/j.ijforecast.2021.03.008.

Sucarrat, G. and Á. Escribano (2012). Automated Model Selection in Finance: General-to-Specific Modelling of the Mean and Volatility Specifications. *Oxford Bulletin of Economics and Statistics 74*, 716–735.

A Proofs

A.1 Proof of Proposition 1

a) From A 1, we know that the series $\{(y_t, \boldsymbol{x}_t')'\}$ is strictly stationary and ergodic. A 2 thus implies that $E(y_t^2)$ is finite for all t, that $E(\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}')$ has finite elements for all t and that it is invertible for all t. By the Cauchy-Shwarz inequality, we have

$$E(\boldsymbol{x}_{t-1}y_t) \leq \sqrt{E(\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}')E(y_t^2)} < \infty \qquad \text{for all } t. \tag{18}$$

So $\boldsymbol{\theta}_0^* \equiv E(\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}')^{-1}E(\boldsymbol{x}_{t-1}y_t) \leq \infty$, and the linear prediction $\boldsymbol{\theta}_0^{*\prime}\boldsymbol{x}_{t-1}$ and the error term u_t both exist.

b) We prove this by formulating an optimization problem and solve for the First Order Condition(FOC). We have

$$\min_{\boldsymbol{\theta}^* \in \boldsymbol{\Theta}^*} E(u_t^2) = \min_{\boldsymbol{\theta}^* \in \boldsymbol{\Theta}^*} E[(y_t - \boldsymbol{\theta}^{*\prime}\boldsymbol{x}_{t-1})^2] \tag{19}$$

The FOC implies

$$\frac{\partial E(u_t^2)}{\partial \boldsymbol{\theta}^*} = E\big[2(y_t - \boldsymbol{\theta}^{*\prime}\boldsymbol{x}_{t-1}) \cdot -\boldsymbol{x}_{t-1}\big] = \mathbf{0}, \tag{20}$$

$$= -2E(\boldsymbol{x}_{t-1}y_t) + 2E(\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}')\boldsymbol{\theta}^* = \mathbf{0}, \tag{21}$$

and therefore that

$$\boldsymbol{\theta}^* = E(\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}')^{-1}E(\boldsymbol{x}_{t-1}y_t).$$

So $E(\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}')^{-1}E(\boldsymbol{x}_{t-1}y_t)$ is a stationary point for the unconditional MSE. It is also straightforward to verify that $\boldsymbol{\theta}^*$ is the minimiser by inspecting the second derivative.

c) By the ergodic and continuous mapping theorems, it follows that the moments in $\boldsymbol{\theta}_0^*$ can be consistently estimated by their sample moments. Accordingly, $\widehat{\boldsymbol{\theta}}^* \underset{p}{\rightarrow} \boldsymbol{\theta}_0^*$ as $T \rightarrow \infty$.

d) Since $u_t = \boldsymbol{\theta}_0^{*\prime}\boldsymbol{x}_{t-1}$, it follows from the FOC that

$$-E\big(u_t \cdot \boldsymbol{x}_{t-1}\big) = \mathbf{0} \tag{22}$$

By assumption A 1, the first entry of $\boldsymbol{x}_{t-1}$ is a constant equal to 1. Accordingly,

$$E(x_{1,t-1}u_t) = E(1 \cdot u_t) = 0 \tag{23}$$

for all t.

A.2 Proof of Proposition 2

a) We can rewrite the log-volatility prediction as

$$\ln h_t = \boldsymbol{\theta}_0'\boldsymbol{x}_{t-1} = \boldsymbol{\theta}_0^{*\prime}\boldsymbol{x}_{t-1} + \ln E(z_t^{*2})x_{1,t-1} \tag{24}$$

where $x_{1,t-1} = 1$ for all t. Then, by Proposition 1 and A 3, we know the RHS of the equation is finite, thus, the log-volatility exists.

b) By definition in (2) and (5), we have

$$\begin{aligned}
E(z_t^2) &= E(r_t^2/h_t) \\
&= E\Big(\frac{r_t^2}{\exp(\ln h_t)}\Big) \\
&= E\Big(\frac{r_t^2}{\exp(\theta_1 + \theta_2 x_{2,t-1} + \ldots + \theta_k x_{k,t-1})}\Big) \\
&= E\Big(\frac{r_t^2}{\exp[(\theta_1^* + \ln E(z_t^{*2})) + \theta_2 x_{2,t-1} + \ldots + \theta_k x_{k,t-1}]}\Big) \\
&= E\Big(\frac{r_t^2}{\exp[(\theta_1^* + \theta_2 x_{2,t-1} + \ldots + \theta_k x_{k,t-1})]\exp\left[\ln E(z_t^{*2})\right]}\Big) \\
&= E\Big(\frac{r_t^2}{\exp(\boldsymbol{\theta}^{*\prime}\boldsymbol{x}_{t-1})}\frac{1}{E(z_t^{*2})}\Big) \\
&= E\Big(z_t^{*2}\frac{1}{E(z_t^{*2})}\Big) \\
&= 1
\end{aligned} \tag{25}$$

c) From c) in Proposition 1, we know the OLS estimator $\widehat{\boldsymbol{\theta}}^*$ consistently estimates $\boldsymbol{\theta}_0^*$. The sample mean of residuals, $\left(T^{-1}\sum_{t=1}^{T}\widehat{z}_t^{*2}\right)$, is defined as a function of $\widehat{\boldsymbol{\theta}}^*$. We want to show that it converges to the population counterpart in probability.

The idea of the proof is transferring the consistency of OLS estimator to the its function form. We will use the mean value theorem in calculus as the bridge for the transfer (Hayashi (2000) pp. 470–471).

The mean value theorem in calculus allows for the mean value expansion of the continuously differentiable function $\boldsymbol{g}()$ as,

$$\underbrace{\boldsymbol{g}(\boldsymbol{x})}_{q\times 1} = \underbrace{\boldsymbol{g}(\boldsymbol{x}_0)}_{q\times 1} + \underbrace{\frac{\partial \boldsymbol{g}(\bar{\boldsymbol{x}})}{\partial \boldsymbol{x}'}}_{q\times p}\underbrace{(\boldsymbol{x}-\boldsymbol{x}_0)}_{p\times 1} \tag{26}$$

where $\boldsymbol{g}() : \mathbb{R}^p \to \mathbb{R}^q$ and $\bar{\boldsymbol{x}}$ is a value between $\boldsymbol{x}$ and $\boldsymbol{x}_0$.

We set $\boldsymbol{x} = \widehat{\boldsymbol{\theta}}^*$, so that

$$\boldsymbol{g}(\widehat{\boldsymbol{\theta}}^*) = T^{-1}\sum_{t=1}^{T} r_t^2/\exp\left(\widehat{\boldsymbol{\theta}}^{*\prime}\boldsymbol{x}_{t-1}\right) = T^{-1}\sum_{t=1}^{T} m(\widehat{\boldsymbol{\theta}}^*) \tag{27}$$

In this setup, $\boldsymbol{g}(\widehat{\boldsymbol{\theta}}^*)$ is affected by sample size T in two ways, direct effect of $\frac{1}{T}$ and indirect effect of $\widehat{\boldsymbol{\theta}_T^*}$. For the simplicity of notations, we still write the function as $\boldsymbol{g}()$, even though its a function of T intrinsically. The mean value expansion around $\boldsymbol{\theta}_0^*$ is,

$$\underbrace{\boldsymbol{g}(\widehat{\boldsymbol{\theta}}^*)}_{1\times 1} = \underbrace{\boldsymbol{g}(\boldsymbol{\theta}_0^*)}_{1\times 1} + \underbrace{\frac{\partial \boldsymbol{g}(\bar{\boldsymbol{\theta}}^*)}{\partial \widehat{\boldsymbol{\theta}}^{*\prime}}}_{1\times k}\underbrace{(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^*)}_{k\times 1} \tag{28}$$

Then, we put all function value terms to LHS to get:

$$\boldsymbol{g}(\widehat{\boldsymbol{\theta}}^*) - \boldsymbol{g}(\boldsymbol{\theta}_0^*) = \frac{\partial \boldsymbol{g}(\bar{\boldsymbol{\theta}}^*)}{\partial \widehat{\boldsymbol{\theta}}^{*\prime}}(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^*) \tag{29}$$

If we can show the RHS is converging to 0 in probability, given that $\boldsymbol{g}(\boldsymbol{\theta}_0^*) \underset{p}{\to} E(z_t^{*2})$, we can prove that $\left(T^{-1}\sum_{t=1}^{T}\widehat{z}_t^{*2}\right) \underset{p}{\to} E(z_t^{*2})$ by the Ergodic and continuous mapping theorems. We decompose this into two steps. First, we show the consistency of the mean value estimator by proving $(\bar{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^*) \underset{p}{\to} \mathbf{0}$. Then, we apply the uniform law of large numbers to the function $\frac{\partial \boldsymbol{g}(\bar{\boldsymbol{\theta}}^*)}{\partial \widehat{\boldsymbol{\theta}}^{*\prime}} = -m(\overline{\boldsymbol{\theta}}^*)\boldsymbol{x}_{t-1}'$, and show the convergence of this vector random function.

First, since the OLS estimator $\widehat{\boldsymbol{\theta}}^*$ consistently estimates $\boldsymbol{\theta}_0^*$. Then, we have $(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^*) \underset{p}{\to} \mathbf{0}$. This indicates for all the components i (where $i = 1, \ldots, k$) in the vector $\widehat{\boldsymbol{\theta}}^*$, we will have:

$$\lim_{T\to\infty} \mathcal{P}\left(|\widehat{\boldsymbol{\theta}_T^*} - \boldsymbol{\theta}_0^*| > \boldsymbol{\epsilon}\right) = 0, \forall\, \boldsymbol{\epsilon} > \mathbf{0} \tag{30}$$

$\bar{\boldsymbol{\theta}}^*$ is a vector lying between $\widehat{\boldsymbol{\theta}_T^*}$ and $\boldsymbol{\theta}_0^*$. Therefore, we can compare the relation of the difference as,

$$|\bar{\boldsymbol{\theta}}_T^* - \boldsymbol{\theta}_0^*| \le |\widehat{\boldsymbol{\theta}_T^*} - \boldsymbol{\theta}_0^*| \tag{31}$$

Combining the two conditions, we can prove that $(\bar{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^*) \underset{p}{\rightarrow} \mathbf{0}$ since

$$\lim_{T\to\infty} \mathcal{P}\left(|\bar{\boldsymbol{\theta}}_T^* - \boldsymbol{\theta}_0^*| > \boldsymbol{\epsilon}\right) = 0, \forall\, \boldsymbol{\epsilon} > \mathbf{0} \tag{32}$$

. We initiate the second part of the proof from the uniform law of large numbers from Lemma 7.2, Hayashi(2000). This lemma ensures the uniform convergence of random functions according to 4 conditions.

(i) The parameter space $\boldsymbol{\Theta}^*$ is compact.

(ii) The vector random function $-m_t(\boldsymbol{\theta}^*)\boldsymbol{x}_{t-1}'$ is continuous in $\boldsymbol{\theta}^*$ for all $\boldsymbol{w}_t \equiv \{(r_t^2, \boldsymbol{x}_t')'\}$.

(iii) The function $-m_t(\boldsymbol{\theta}^*)\boldsymbol{x}_{t-1}'$ is measurable in $\boldsymbol{w}_t$ for all the $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}^*$.

(iv) The dominance condition is satisfied:

$$E\Big[\sup_{\boldsymbol{\theta}^*\in\boldsymbol{\Theta}^*} \|-m_t(\boldsymbol{w}_t;\boldsymbol{\theta}^*)\boldsymbol{x}_{t-1}'\|\Big] < \infty$$

.

From A 1, we know that series $\boldsymbol{w}_t \equiv \{(r_t^2, \boldsymbol{x}_t')'\}$ is strictly stationary and ergodic. We also have the compactness assumption of the parameter space in A 4. Then, the first condition (i) is staified. The vector random function

$$\begin{aligned}\frac{\partial \boldsymbol{g}(\bar{\boldsymbol{\theta}}^*)}{\partial \widehat{\boldsymbol{\theta}}^{*\prime}} &= -m(\bar{\boldsymbol{\theta}}^*)\boldsymbol{x}_{t-1}' \\ &= -\frac{r_t^2}{\exp(\bar{\boldsymbol{\theta}}^{*\prime}\boldsymbol{x}_{t-1})}\boldsymbol{x}_{t-1}'\end{aligned} \tag{33}$$

is continuous in $\bar{\boldsymbol{\theta}}^*$ for all $\boldsymbol{w}_t$, which fulfills condition (ii). $m(\bar{\boldsymbol{\theta}}^*)\boldsymbol{x}_{t-1}'$ is measurable in $\{(r_t^2, \boldsymbol{x}_t')'\}$ for all $\bar{\boldsymbol{\theta}}^*$in parameter space $\boldsymbol{\Theta}^*$. This verifies condition (iii). Since the parameter space is compact, we know

$$\begin{aligned}\frac{r_t^2}{\exp(\bar{\boldsymbol{\theta}}^{*\prime}\boldsymbol{x}_{t-1})}\boldsymbol{x}_{t-1}' &\le \frac{r_t^2}{\exp(\bar{\boldsymbol{\theta}}^{*\prime}\boldsymbol{x}_{t-1})}\|\boldsymbol{x}_{t-1}'\| \\ &\le \sup_{\boldsymbol{\theta}^*\in\boldsymbol{\Theta}^*} \frac{r_t^2}{\exp(\boldsymbol{\theta}^{*\prime}\boldsymbol{x}_{t-1})}\|\boldsymbol{x}_{t-1}'\|\end{aligned} \tag{34}$$

From A 4, we know $E\Big(\sup_{\boldsymbol{\theta}^*\in\boldsymbol{\Theta}^*} \frac{r_t^2}{\exp(\boldsymbol{\theta}^{*\prime}\boldsymbol{x}_{t-1})}\Big) = E\Big(\sup_{\boldsymbol{\theta}^*\in\boldsymbol{\Theta}^*} m(\boldsymbol{\theta}^*)\Big) < \infty$. Combining this with A 1, we can decompose and show the finiteness by

$$\begin{aligned}E\Big(\sup_{\boldsymbol{\theta}^*\in\boldsymbol{\Theta}^*} \frac{r_t^2}{\exp(\boldsymbol{\theta}^{*\prime}\boldsymbol{x}_{t-1})}\|\boldsymbol{x}_{t-1}'\|\Big) &\le E\Big(\sup_{\boldsymbol{\theta}^*\in\boldsymbol{\Theta}^*} \frac{r_t^2}{\exp(\boldsymbol{\theta}^{*\prime}\boldsymbol{x}_{t-1})}\Big)E(\|\boldsymbol{x}_{t-1}'\|) \\ &< \infty\end{aligned} \tag{35}$$

Then, we have the dominance condition (iv) satisfied.

Then $E(m(\boldsymbol{\theta}^*)\boldsymbol{x}'_{t-1})$ is a continuous function of $\boldsymbol{\theta}^*$ and $\frac{1}{T}\Big(\sum_{t=1}^{T}\frac{r_t^2}{\exp(\bar{\boldsymbol{\theta}}^{*\prime}\boldsymbol{x}_{t-1})}\boldsymbol{x}'_{t-1}\Big)$ will converge uniformly in probability to its population mean $E\Big(\frac{r_t^2}{\exp\big(\boldsymbol{\theta}^{*\prime}\boldsymbol{x}_{t-1}\big)}\boldsymbol{x}'_{t-1}\Big)$, which is a vector with each element finite. Thus, the RHS will converge to $\mathbf{0}$ in probability. And we have,

$$\Big(T^{-1}\sum_{t=1}^{T}\widehat{z}_t^{*2}\Big) \underset{p}{\rightarrow} E(z_t^{*2}) \tag{36}$$

d) Notice that the only difference between $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}^*$ is in the first entry.

$$\theta_{01} = \theta_{01}^* + \ln E(z_t^{*2}), \qquad \theta_{0j} = \theta_{0j}^* \qquad \text{for} \qquad j > 1. \tag{37}$$

From c) in Proposition 2, $E(z_t^{*2})$ can be consistently estimated. Then, with the continuous mapping theorem, we can also consistently estimate $\ln E(z_t^{*2})$ from its sample mean. This implies that our estimator of θ_{01} is also consistent. We have proved that $\hat{\boldsymbol{\theta}}^* \underset{p}{\rightarrow} \boldsymbol{\theta}_0^*$, and this also implies,

$$\hat{\theta_{0j}^*} \underset{p}{\rightarrow} \theta_{0j}^* = \theta_{0j} \qquad \text{for} \qquad j > 1.$$

Combining the two parts, we conclude that $\widehat{\boldsymbol{\theta}} \underset{p}{\rightarrow} \boldsymbol{\theta}_0$. ■

A.3 Proof of Proposition 3

The sampling error is defined as

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^* &= \left[\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1}\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{t-1}y_t\right] - \boldsymbol{\theta}_0^* \\
&= \left[\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1}\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{t-1}u_t\right] \\
&= \left[\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1}\right]^{-1}\bar{\boldsymbol{s}}_t
\end{aligned} \tag{38}$$

where $\bar{\boldsymbol{s}}_t \equiv \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{t-1}u_t$. Multiplying through by $\sqrt{T}$, we obtain,

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^*) = \left[\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1}\right]^{-1}\sqrt{T}\bar{\boldsymbol{s}}_t \tag{39}$$

From A 1, we have that $\boldsymbol{w}_t \equiv \{(r_t^2, \boldsymbol{x}'_t)'\}$ is strictly stationary and ergodic. Then we have,

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1} \underset{p}{\rightarrow} E(\boldsymbol{x}_t\boldsymbol{x}'_t). \tag{40}$$

We apply continuous mapping theorem and get the convergence of the reverse form,

$$\left[\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1}\right]^{-1} \underset{p}{\rightarrow} E(\boldsymbol{x}_t\boldsymbol{x}'_t)^{-1}. \tag{41}$$

From A 5, it follows that,

$$\begin{aligned}\sqrt{T}\bar{\boldsymbol{s}}_t &= \sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{s}_t \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{s}_t \underset{d}{\rightarrow} N(\boldsymbol{0},\boldsymbol{S})\end{aligned} \tag{42}$$

By Slutsky's theorem, cf. Lemma 2.4 (c) in Hayashi(2000), we obtain,

(i) $\sqrt{T}\bar{\boldsymbol{s}}_t \underset{d}{\rightarrow} N(\boldsymbol{0},\boldsymbol{S})$.

(ii) $\left[\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1}\right]^{-1} \underset{p}{\rightarrow} E(\boldsymbol{x}_t\boldsymbol{x}'_t)^{-1}$.

Then,

$$\left[\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1}\right]^{-1}\sqrt{T}\bar{\boldsymbol{s}}_t \underset{d}{\rightarrow} N(\boldsymbol{0}, E(\boldsymbol{x}_t\boldsymbol{x}'_t)^{-1}\boldsymbol{S}E(\boldsymbol{x}_t\boldsymbol{x}'_t)^{-1}) \tag{43}$$

So $\sqrt{T}(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^*) \underset{d}{\rightarrow} N(\boldsymbol{0},\boldsymbol{\Sigma}^*)$, where $\boldsymbol{\Sigma}^* = E(\boldsymbol{x}_t\boldsymbol{x}'_t)^{-1}\boldsymbol{S}E(\boldsymbol{x}_t\boldsymbol{x}'_t)^{-1}$. ■

A.4 Proof of Proposition 4

Define $\tau_0 \equiv \ln E(z_t^{*2})$ and $\widehat{\tau} \equiv \Big(\ln T^{-1}\sum_{t=1}^{T}\widehat{z}_t^{*2}\Big)$. Since $T^{-1}\sum_{t=1}^{T}\widehat{z}_t^{*2}$ is consistent for $E(z_t^{*2})$, from the continuous mapping theorem it follows that $(\widehat{\tau} - \tau_0) \underset{p}{\rightarrow} 0$. We now derive

$$\sqrt{T}(\widehat{\tau} - \tau_0) \underset{d}{\rightarrow} N(\boldsymbol{0},\boldsymbol{\Sigma}_\tau). \tag{44}$$

To this end, define

$$\widehat{\boldsymbol{\phi}} \equiv \begin{bmatrix}\widehat{\boldsymbol{\theta}}^* \\ \widehat{\tau}\end{bmatrix} \tag{45}$$

and derive the asymptotic distribution of $\widehat{\boldsymbol{\phi}}$.

We first derive the structure of the variance-covariance matrix $\boldsymbol{\Sigma}_\tau$ by applying the mean value expansion to $\widehat{\tau}$. We use the definition of $m_t()$ in A 3 and rewrite the equation using $m_t(\boldsymbol{\theta}_0^*) \equiv r_t^2/\exp(\boldsymbol{\theta}_0^{*\prime}\boldsymbol{x}_{t-1})$,

$$
\begin{aligned}
\widehat{\tau}(\widehat{\boldsymbol{\theta}}^*) &= \widehat{\tau}(\boldsymbol{\theta}_0^*) + \frac{\partial \widehat{\tau}(\bar{\boldsymbol{\theta}}^*)}{\partial \boldsymbol{\theta}^{*\prime}}(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^*) \\
&= \widehat{\tau}(\boldsymbol{\theta}_0^*) + \frac{1}{\frac{1}{T}\sum_{t=1}^T m_t(\bar{\boldsymbol{\theta}}^*)} \frac{1}{T}\sum_{t=1}^T \frac{\partial m_t(\bar{\boldsymbol{\theta}}^*)}{\partial \boldsymbol{\theta}^{*\prime}}(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^*)
\end{aligned} \tag{46}
$$

where $\bar{\boldsymbol{\theta}}^*$ is a value between $\widehat{\boldsymbol{\theta}}^*$ and $\boldsymbol{\theta}_0^*$. Then, we subtract τ_0 from both sides of the equation and get,

$$
\begin{aligned}
\widehat{\tau}(\widehat{\boldsymbol{\theta}}^*) - \tau_0 &= \ln \frac{1}{T}\sum_{t=1}^T \widehat{z}_t^{*2} - \tau_0 \\
&= (\tau(\boldsymbol{\theta}_0^*) - \tau_0) + \frac{\frac{1}{T}\sum_{t=1}^T \frac{\partial m_t(\bar{\boldsymbol{\theta}}^*)}{\partial \boldsymbol{\theta}^{*\prime}}}{\frac{1}{T}\sum_{t=1}^T m_t(\bar{\boldsymbol{\theta}}^*)}(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^*)
\end{aligned} \tag{47}
$$

We scale both sides by $\sqrt{T}$

$$
\sqrt{T}(\widehat{\tau}(\widehat{\boldsymbol{\theta}}^*) - \tau_0) = \sqrt{T}(\tau(\boldsymbol{\theta}_0^*) - \tau_0) + \frac{\frac{1}{T}\sum_{t=1}^T \frac{\partial m_t(\bar{\boldsymbol{\theta}}^*)}{\partial \boldsymbol{\theta}^{*\prime}}}{\frac{1}{T}\sum_{t=1}^T m_t(\bar{\boldsymbol{\theta}}^*)}\sqrt{T}(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^*), \tag{48}
$$

use a add-and-subtract trick in the second term of RHS to get,

$$
\begin{aligned}
\frac{\frac{1}{T}\sum_{t=1}^T \frac{\partial m_t(\bar{\boldsymbol{\theta}}^*)}{\partial \boldsymbol{\theta}^{*\prime}}}{\frac{1}{T}\sum_{t=1}^T m_t(\bar{\boldsymbol{\theta}}^*)}\sqrt{T}(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^*) &= \Big(\frac{\frac{1}{T}\sum_{t=1}^T \frac{\partial m_t(\bar{\boldsymbol{\theta}}^*)}{\partial \boldsymbol{\theta}^{*\prime}}}{\frac{1}{T}\sum_{t=1}^T m_t(\bar{\boldsymbol{\theta}}^*)} + \frac{E(z_t^{*2}\boldsymbol{x}_{t-1}')}{E(z_t^{*2})} - \frac{E(z_t^{*2}\boldsymbol{x}_{t-1}')}{E(z_t^{*2})}\Big)\sqrt{T}(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^*) \\
&= \Big[-\frac{E(z_t^{*2}\boldsymbol{x}_{t-1}')}{E(z_t^{*2})} + \Big(\frac{\frac{1}{T}\sum_{t=1}^T \frac{\partial m_t(\bar{\boldsymbol{\theta}}^*)}{\partial \boldsymbol{\theta}^{*\prime}}}{\frac{1}{T}\sum_{t=1}^T m_t(\bar{\boldsymbol{\theta}}^*)} + \frac{E(z_t^{*2}\boldsymbol{x}_{t-1}')}{E(z_t^{*2})}\Big)\Big]\sqrt{T}(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^*)
\end{aligned} \tag{49}
$$

For the second part of the RHS, with the property of strictly stationary and egordic and the uniform law of large numbers, we have,

$$
\frac{1}{T}\sum_{t=1}^T m_t(\bar{\boldsymbol{\theta}}^*) \underset{p}{\to} E(z_t^{*2}) \tag{50}
$$

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^T \frac{\partial m_t(\bar{\boldsymbol{\theta}}^*)}{\partial \widehat{\boldsymbol{\theta}}^{*\prime}} &= -\frac{1}{T}\sum_{t=1}^T m_t(\bar{\boldsymbol{\theta}}^*)\boldsymbol{x}_{t-1}' \\
&\underset{p}{\to} -E(z_t^{*2}\boldsymbol{x}_{t-1}')
\end{aligned} \tag{51}
$$

and we also have

$$
\frac{\frac{1}{T}\sum_{t=1}^T \frac{\partial m_t(\bar{\boldsymbol{\theta}}^*)}{\partial \boldsymbol{\theta}^{*\prime}}}{\frac{1}{T}\sum_{t=1}^T m_t(\bar{\boldsymbol{\theta}}^*)} \underset{p}{\to} -\frac{E(z_t^{*2}\boldsymbol{x}_{t-1}')}{E(z_t^{*2})} \tag{52}
$$

and we can rewrite the equation as

$$
\sqrt{T}(\widehat{\tau}(\widehat{\boldsymbol{\theta}}^*) - \tau_0) = \sqrt{T}(\tau(\boldsymbol{\theta}_0^*) - \tau_0) - \frac{E(z_t^{*2}\boldsymbol{x}_{t-1}')}{E(z_t^{*2})}\sqrt{T}(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^*) + o_p(1) \tag{53}
$$

For the part $\widehat{\tau}(\boldsymbol{\theta}_0^*)$ on the RHS, we apply the mean value theorem on the $\ln x$ function,

$$\ln x = \ln x_0 + \frac{\partial \ln \overline{x}}{\partial x}(x - x_0) \tag{54}$$

where $\overline{x}$ is a value between x and x_0, and

$$x = \frac{1}{T}\sum_{t=1}^{T} z_t^{*2} \tag{55}$$

$$x_0 = E(z_t^{*2}) \tag{56}$$

$$\begin{aligned}\frac{\partial \ln \overline{x}}{\partial x} &= \frac{\partial \overline{\tau}}{\partial e^{\tau}} \\ &= e^{-\overline{\tau}}\end{aligned} \tag{57}$$

Recall that $e^{\tau_0} = E(z_t^{*2})$, we can have the mean value expansion as,

$$\begin{aligned}\ln \frac{1}{T}\sum_{t=1}^{T} z_t^{*2} &= \ln E(z_t^{*2}) + \frac{\partial \overline{\tau}}{\partial e^{\tau}}\Big(\frac{1}{T}\sum_{t=1}^{T} z_t^{*2} - E(z_t^{*2})\Big) \\ &= \ln E(z_t^{*2}) + e^{-\overline{\tau}}\Big(\frac{1}{T}\sum_{t=1}^{T} z_t^{*2} - E(z_t^{*2})\Big)\end{aligned} \tag{58}$$

where $e^{-\overline{\tau}}$ is a value between $\left(\frac{1}{T}\sum_{t=1}^{T} z_t^{*2}\right)^{-1}$ and $E(z_t^{*2})^{-1}$.

$$\begin{aligned}\ln \frac{1}{T}\sum_{t=1}^{T} z_t^{*2} &= \ln E(z_t^{*2}) + e^{-\overline{\tau}}\Big(\frac{1}{T}\sum_{t=1}^{T} z_t^{*2} - E(z_t^{*2})\Big) \\ &= \ln E(z_t^{*2}) + \Big(e^{-\overline{\tau}} - E(z_t^{*2})^{-1} + E(z_t^{*2})^{-1}\Big)\Big(\frac{1}{T}\sum_{t=1}^{T} z_t^{*2} - E(z_t^{*2})\Big) \\ &= \ln E(z_t^{*2}) + E(z_t^{*2})^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T} z_t^{*2} - E(z_t^{*2})\Big) + o_p(1)\end{aligned} \tag{59}$$

Combining the two equations, we get,

$$\begin{aligned}\sqrt{T}(\widehat{\tau}(\widehat{\boldsymbol{\theta}}^*) - \tau_0) &= \sqrt{T}(\tau(\boldsymbol{\theta}_0^*) - \tau_0) - \frac{E(z_t^{*2}\boldsymbol{x}_{t-1}')}{E(z_t^{*2})}\sqrt{T}(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^*) + o_p(1) \\ &= \frac{1}{E(z_t^{*2})}\sqrt{T}\Big(\frac{1}{T}\sum_{t=1}^{T} z_t^{*2} - E(z_t^{*2})\Big) - \frac{E(z_t^{*2}\boldsymbol{x}_{t-1}')}{E(z_t^{*2})}\sqrt{T}(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^*) + o_p(1)\end{aligned} \tag{60}$$

From earlier, we know that

$$\begin{aligned}\sqrt{T}(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^*) &= \left[\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'\right]^{-1}\sqrt{T}\bar{\boldsymbol{s}}_t \\ &= E(\boldsymbol{x}_t\boldsymbol{x}_t')^{-1}\sqrt{T}\bar{\boldsymbol{s}}_t + o_p(1)\end{aligned} \tag{61}$$

In order to get the joint distribution of the estimators, we stack the estimators on top of each other and get,

$$\begin{aligned}
\begin{bmatrix} \sqrt{T}(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^*) \\ \sqrt{T}(\widehat{\tau} - \tau_0) \end{bmatrix} &= \begin{bmatrix} E(\boldsymbol{x}_t\boldsymbol{x}_t')^{-1}\sqrt{T}\bar{\boldsymbol{s}}_t \\ -\frac{E(z_t^{*2}\boldsymbol{x}_{t-1}')}{E(z_t^{*2})}E(\boldsymbol{x}_t\boldsymbol{x}_t')^{-1}\sqrt{T}\bar{\boldsymbol{s}}_t + \frac{1}{E(z_t^{*2})}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(z_t^{*2} - E(z_t^{*2})\Big) \end{bmatrix} + \begin{bmatrix} o_p(1) \\ o_p(1) \end{bmatrix} \\
&= \begin{bmatrix} E(\boldsymbol{x}_t\boldsymbol{x}_t')^{-1} & \mathbf{0} \\ -\frac{E(z_t^{*2}\boldsymbol{x}_{t-1}')}{E(z_t^{*2})}E(\boldsymbol{x}_t\boldsymbol{x}_t')^{-1} & \frac{1}{E(z_t^{*2})} \end{bmatrix} \sqrt{T} \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{s}_t \\ \frac{1}{T}\sum_{t=1}^{T}\Big(z_t^{*2} - E(z_t^{*2})\Big) \end{bmatrix} + \begin{bmatrix} o_p(1) \\ o_p(1) \end{bmatrix} \\
&= \begin{bmatrix} E(\boldsymbol{x}_t\boldsymbol{x}_t')^{-1} & \mathbf{0} \\ -\frac{E(z_t^{*2}\boldsymbol{x}_{t-1}')}{E(z_t^{*2})}E(\boldsymbol{x}_t\boldsymbol{x}_t')^{-1} & E(z_t^{*2})^{-1} \end{bmatrix} \sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{w}_t + \begin{bmatrix} o_p(1) \\ o_p(1) \end{bmatrix} \\
&= \boldsymbol{G}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{w}_t + \begin{bmatrix} o_p(1) \\ o_p(1) \end{bmatrix}
\end{aligned} \tag{62}$$

From A 5, we know the joint distribution of $\boldsymbol{w}_t$, and we apply the Slutsky's theroem to get,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{w}_t \underset{d}{\to} N(\mathbf{0}, \boldsymbol{\Sigma}_w)$$

$$\boldsymbol{G}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{w}_t \underset{d}{\to} N(\mathbf{0}, \boldsymbol{G}\boldsymbol{\Sigma}_w\boldsymbol{G}') \tag{63}$$

A.5 Proof of Corollary 1

The result follows from applying the delta method on the result of Proposition 4. We have the representation of $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\phi}}$.

$$\widehat{\boldsymbol{\theta}} = \begin{bmatrix} \widehat{\theta}_1 \\ \widehat{\theta}_2 \\ \vdots \\ \widehat{\theta}_K \end{bmatrix} = \begin{bmatrix} \widehat{\theta}_1^* + \widehat{\tau} \\ \widehat{\theta}_2^* \\ \vdots \\ \widehat{\theta}_K^* \end{bmatrix} \tag{64}$$

$$\widehat{\boldsymbol{\phi}} = \begin{bmatrix} \widehat{\theta}_1^* \\ \widehat{\theta}_2^* \\ \vdots \\ \widehat{\theta}_K^* \\ \widehat{\tau} \end{bmatrix} \tag{65}$$

We create a function $\boldsymbol{a}(\boldsymbol{\phi}_0) : \mathbb{R}^{K+1} \to \mathbb{R}^K$ to specify the relation between the true value of the parameters.

$$\begin{aligned}\boldsymbol{\theta}_0 &= \boldsymbol{a}(\boldsymbol{\phi}) \\ &= \begin{pmatrix} \phi_{0,1} + \phi_{0,K+1} \\ \phi_{0,2} \\ \vdots \\ \phi_{0,K} \end{pmatrix} \\ &= \begin{pmatrix} \theta_{0,1} + \tau \\ \theta_{0,2} \\ \vdots \\ \theta_{0,K} \end{pmatrix}\end{aligned} \tag{66}$$

By taking the first derivative of the $\boldsymbol{a}()$ function, we have the following derivatives,

$$\begin{aligned}\boldsymbol{A}(\boldsymbol{\phi}) &= \frac{\partial \boldsymbol{a}(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}'} \\ &= \left(\frac{\partial \boldsymbol{a}(\boldsymbol{\phi})}{\partial \phi_{0,1}}, \dots, \frac{\partial \boldsymbol{a}(\boldsymbol{\phi})}{\partial \phi_{0,K}} \right) \\ &= \begin{bmatrix} 1 & 0 & \dots & 0 & 1 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}\end{aligned} \tag{67}$$

This partial derivative $\boldsymbol{A}(\boldsymbol{\phi})$ is a $K \times (K+1)$ matrix, with the constant elements. Then, we apply Lemma 2.5 in Hayashi(2000), which is the "delta method" to get,

$$\sqrt{T}(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) \underset{d}{\rightarrow} N(\mathbf{0}, \boldsymbol{G}\boldsymbol{\Sigma}_w\boldsymbol{G}') \tag{68}$$

$$\sqrt{T}(\boldsymbol{a}(\widehat{\boldsymbol{\phi}}) - \boldsymbol{a}(\boldsymbol{\phi}_0)) \underset{d}{\rightarrow} N(\mathbf{0}, \boldsymbol{A}(\boldsymbol{\phi})\boldsymbol{G}\boldsymbol{\Sigma}_w\boldsymbol{G}'\boldsymbol{A}(\boldsymbol{\phi})') \tag{69}$$

$$\sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \underset{d}{\rightarrow} N(\mathbf{0}, \boldsymbol{A}(\boldsymbol{\phi})\boldsymbol{\Sigma}_{\boldsymbol{\phi}}\boldsymbol{A}(\boldsymbol{\phi})') \tag{70}$$

The result follows from applying the delta method on the result of Proposition 4.