Estimation and Inference in Adaptive Learning Models with Slowly Decreasing Gains

Alexander Mayer*

April 24, 2019

Abstract

This paper develops techniques of estimation and inference in a prototypical macroeconomic adaptive learning model with slowly decreasing gains. A sequential three-step procedure based on a 'super-consistent' estimator of the rational expectations equilibrium parameter is proposed. This procedure is asymptotically equivalent to first estimating the structural parameters jointly via ordinary least-squares and then using the so-obtained estimates to form a plug-in estimator of the rational expectations equilibrium parameter. In spite of failing Grenander's conditions for well-behaved data, limiting normality of all estimators centered at their true parameter values is established. It is then shown that, notwithstanding potential threats to inference arising from non-standard convergence rates and a singular variance-covariance matrix, classical hypothesis tests involving single, as well as joint restrictions remain valid. Monte-Carlo evidence confirms the accuracy of the asymptotic theory for the finite sample behaviour of the estimators and test statistics discussed here.

Keywords: *Adaptive learning, rational expectations, asymptotic collinearity, non-stationary regression, degenerate variances, co-integration, strong mixing, (generalized) Wald-statistic.*

1 Introduction

This paper is concerned with estimation and inference procedures in a stylized macroeconomic learning model

$$y_t = \beta y^e_{t|t-1} + \delta x_t + \varepsilon_t, \quad \text{for} \quad t = 1, 2, \ldots, T, \tag{1.1}$$

where $y^e_{t|t-1}$ represents agents' (potentially non-rational) expectation of y_t formed in period $t-1$, x_t is a strictly exogenous regressor and ε_t represents the disturbance term, with properties to be discussed below. Furthermore, it is assumed that $|\beta| < 1$, while δ is allowed to

*WHU – Otto Beisheim School of Management, Germany; e-mail: alexander.mayer@whu.edu. The author is grateful to Michael Massmann, Jörg Breitung, Frank Kleibergen and Andrew Rosalsky for helpful discussions and comments. Suggestions made by participants of the IAAE conference 2018 in Montreal and the SMYE 2019 in Brussels are highly appreciated. All remaining errors are my own.

be any real number. Various economic models, like the classical cobweb model or the New Keynesian Phillips curve, can be cast in form of (1.1); for a more detailed account of examples encompassed by (1.1) see Christopeit and Massmann (2018). The crucial feature of model (1.1) is the expectation term $y^e_{t|t-1}$. Under rational expectations (RE), economic agents are presumed to incorporate the information $\mathcal{I}_{t-1} := \sigma(y_s, s < t;\, x_s,\, s \leq t)$ in an optimal manner so that $y^e_{t|t-1} = E[y_t \mid \mathcal{I}_{t-1}]$, thereby yielding the RE equilibrium

$$y_t = \alpha x_t + \varepsilon_t, \quad \text{with} \quad \alpha := \delta/(1-\beta). \tag{1.2}$$

The plausibility of the traditional RE approach to modeling expectations has, however, been contested in recent years (see, e.g., Evans and Honkapohja (2001)). According to the macroeconomic learning literature, economic agents depart in many situations from RE by behaving 'boundedly rational': rather than presupposing complete knowledge of $E[y_t \mid \mathcal{I}_{t-1}]$, the economic agent acts like an econometrician forecasting α recursively. Specifically, the agent is assumed to update her expectations according to the adaptive scheme

$$y^e_{t|t-1} = a_{t-1} x_t, \tag{1.3}$$

where the point forecast a_t of α is obtained from some recursive procedure; see, e.g., Sargent (1993) or Evans and Honkapohja (2001). In line with this literature, the present paper assumes the learning scheme to take the form of a least-squares type stochastic approximation algorithm

$$\begin{aligned} a_t &= a_{t-1} + \gamma_t \frac{x_t}{r_t}(y_t - x_t a_{t-1}) \\ r_t &= r_{t-1} + \gamma_t (x_t^2 - r_{t-1}), \end{aligned} \tag{1.4}$$

whereby, in addition to estimating α by a_t, a further 'normalization' step based on r_t is taken by the agent to estimate the regressor second moment $E[x_t^2]$. The so-called 'gain' γ_t reflects the agents' responsiveness to previous forecast errors. Several empirical applications have confirmed the plausibility of (1.4) as a model for expectation formation of macroeconomic variables (see, e.g., Chakraborty and Evans (2008), Markiewicz and Pick (2014), Berardi and Galimberti (2014) or Berardi and Galimberti (2017)); a further illustration will be discussed below.

With boundedly rational agents, the actual law of motion is therefore given by

$$y_t = \beta a_{t-1} x_t + \delta x_t + \varepsilon_t. \tag{1.5}$$

A comparison between the preceding display and equation (1.2) reveals that the informational friction associated with bounded rationality facilitates identification of the 'structural' parameters $\lambda := (\beta,\, \delta)'$. This, in turn, raises the question of whether λ can be estimated and, if so, what can be said about the properties of any such estimator? As has been demonstrated by Christopeit and Massmann (2018), finding an answer to this question is a non-trivial undertaking because the data-generating process resulting from (1.4) and (1.5) is highly non-linear and subject to asymptotic collinearity caused by a non-stationary and self-referential regressor (see also Chevillon and Mavroeidis (2017)). Christopeit and Massmann (2018) thoroughly study the statistical properties of the joint OLS estimator of λ for constant gains (i.e. $\gamma_t = \gamma$) and recursive least-squares learning (i.e. $\gamma_t = 1/t$). Yet despite its great theoretical value, the work of the aforementioned authors is hardly applicable to the study of macroeconomic phenomena, which constitute the very motivation of the authors's undertaking. Specifically, in order to keep their asymptotic analysis tractable, the restriction $x_t = 1$ is imposed, an assumption that is difficult to justify on practical grounds. However, by slightly changing the specification of γ_t, the current exposition demonstrates that time-varying, and possibly weakly dependent, regressors can be allowed for.[1] Besides this gain in generality, the framework discussed here offers several other advantages: First, instead of logarithmic convergence rates as established by Christopeit and Massmann (2018), estimators of the structural parameters are found to converge at a rate that facilitates a practical application to real-world data. Specifically, the estimators of β and δ converge at a polynomial rate to its joint normal distribution, with the rate of convergence being inversely related to the learning rate of the agent. Second, the asymptotic properties of estimators of the equilibrium parameter α are established. Similar to estimating co-integrating relations, a linear combination of the joint estimator of the 'short-run' coefficients β and δ, which govern the actual law of motion (1.5), converges at a faster rate to the 'long-run' RE equilibrium parameter α. This estimation framework is then complemented by an asymptotically equivalent three-step procedure that bears clear similarities to the two-step approach to estimating co-integrating vectors and associated error correction models popularized by Engle and Granger (1987). Finally, the question of hypothesis testing is addressed comprehensively, including a detailed treatment of the properties of Wald(-type) test statistics in the presence of a singular variance-covariance matrix.

The remainder of the paper is organized as follows. Section 2 lays out assumptions and discusses estimation and inference procedures. Monte Carlo evidence is reported in section 3 while section 4 concludes. All proofs are provided in the appendix. Additional results are provided in the supplementary material.

2 Estimation and inference

2.1 Assumptions

The first assumption specifies the nature of the gain sequence used by the agent in her updating scheme (1.4). Specifically, γ_t is assumed to be of polynomial form:

Assumption A *The sequence $(\gamma_t)_{t\geq 1}$ of positive real numbers satisfies*

$$\gamma_t := \gamma/t^\eta,$$

where $\eta \in (1/2, 1)$ and $\gamma > 0$.

Hence, the present specification of the gain sequence covers the intermediate case lying on a continuum between least-squares learning ($\eta = 1$) and constant-gain learning ($\eta = 0$) considered by Christopeit and Massmann (2018). Following the terminology used in the stochastic

[1] It is noteworthy that the proposed econometric framework requires the development of an asymptotic theory which is distinct from the treatment in Christopeit and Massmann (2018).

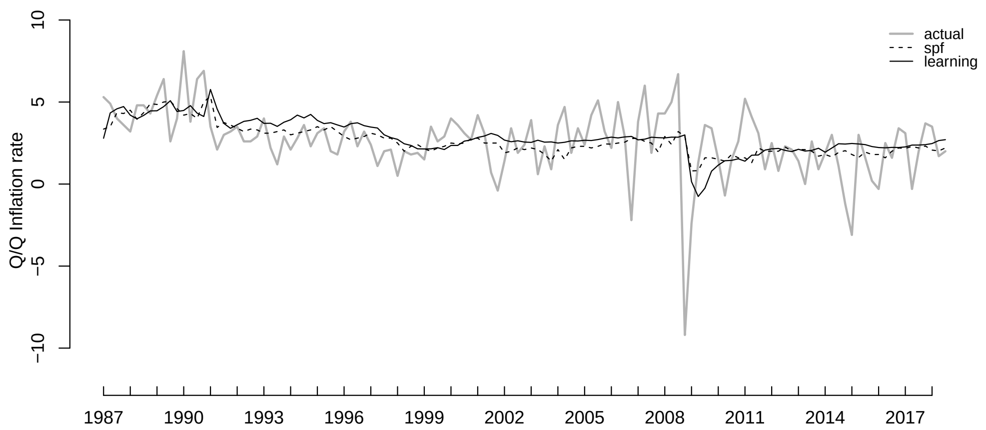
approximation literature, this polynomial gain with $\eta \in (1/2, 1)$ considered here will be referred to as 'slowly decreasing' (cf. Polyak and Juditsky (1992), Kushner and Yan (1993) or Chen (1993)). The gain parameter η plays a key role in the current analysis of estimation and testing procedures based on model (1.5). Specifically, the agent learns the rational expectations parameter α the faster the larger η; an intuitive idea formalized by the following L^2 estimate

$$E[(a_t - \alpha)^2] \sim \gamma_t,$$

cf. appendix A.3. One immediate consequence of the preceding display is that the model of interest (1.5) is subject to asymptotic collinearity as $t \to \infty$.

Imposing the restriction $\eta > 1/2$ in assumption A ensures the square summability of the gain sequence $(\gamma_k,\, k \geq 1)$, which turns out to be a crucial condition for the development of an appropriate asymptotic theory; see also the treatment in Kuan and White (1994) and the discussion in Christopeit and Massmann (2018, pp. 5). Besides this theoretical argument, there is also an empirical motivation behind assumption A, namely the ability of the learning scheme (1.4) with slowly decreasing gain to replicate the expectation formation of professional forecasters. As an illustration, figure 1 depicts the historical US inflation rate (grey line;

Figure 1: Actual and expected inflation (1986Q3-2018Q3)



'actual') together with the median of one-quarter ahead forecasts of the survey of professional forecasters (dotted black line; 'spf'); both taken from the Federal Reserve Bank of Philadelphia. Based on a mean-square error criterion, the solid black line represents that learning rule $y_t^e = a_{t-1}' x_t$, which best fits the 'true' SPF expectations $\widehat{y}_t$, say.[2] Motivated by a Phillips curve relationship, x_t in (1.4) contains an unrestricted constant and an output gap measure taken from the Congressional Budget Office. The optimal pair $(\gamma,\, \eta)$ that minimizes $T^{-1} \sum_{t=1}^{T} (\widehat{y}_t - y_t^e)^2$ over $[0, 1] \times [0, 2]$ is $\gamma \simeq 1$ and $\eta \simeq 0.75$, thereby providing empirical evidence in support of assumption A.[3]

[2]Remark 2.2 below demonstrates how the case of a single regressor x_t can be extended to the case where x_t represents a $k \times 1$ $(k \geq 1)$ dimensional column vector of covariates.

[3]Note, that the mean-square error is evaluated over a grid which encompasses the gain sequences considered in Christopeit and Massmann (2018). The initial values $(a_0,\, r_0)'$ are estimated via OLS based on a pre-sample of length 20.

The following set of assumptions specifies the distributional characteristics of $(x_t,\, \varepsilon_t)'$ and $(a_t,\, r_t)'$:

Assumption B

(B1) *The process* $(\varepsilon_t,\, t \geq 1)$ *forms a martingale difference sequence with respect to*

$$\mathcal{G}_t := \sigma(\{x_k,\, k \geq 1\} \cup \{\varepsilon_s,\, s \leq t\}),$$

and $\kappa_\varepsilon^{(m)} := E[\varepsilon_t^m \mid \mathcal{G}_{t-1}]$ $(m = 2, 3, 4)$ *are finite constants almost surely* (*a.s.*).

(B2) *The process* $(x_k,\, k \geq 1)$ *is strictly stationary and fulfills one of the following conditions:*

(1) *The adapted sequence* $(\{(x_k - E[x_1]), \mathcal{F}_k\}, k \geq 1)$ *forms a martingale difference sequence. The following conditional expectations are, for all* t*, equivalent and equal to finite constants a.s.*

$$\kappa_x^{(n)} := E[x_t^n \mid \mathcal{V}_{t-1}] = E[x_t^n \mid \mathcal{F}_{t-1}],\ n = 2, \ldots, 8, \tag{i}$$

with

$$\mathcal{V}_t := \sigma(\{(x_s,\, \varepsilon_s),\, s \leq t\}) \text{ and } \mathcal{F}_t := \sigma(\{x_s,\, s \leq t\}).$$

(2) *The process* $(x_t,\, t \geq 1)$ *is* α*-mixing with mixing coefficient satisfying*

$$\alpha(m) = O(m^{-(1+\varrho)}), \tag{ii}$$

for some $\varrho > 0$ *and* $\|x_1\|_\infty < \infty$.[4] *The following (not necessarily constant) conditional expectations are equal a.s.*

$$E[x_t^2 \mid \mathcal{V}_{t-1}] = E[x_t^2 \mid \mathcal{F}_{t-1}]. \tag{iii}$$

(B3) *The initial value* a_0 *is independently distributed of* $(\{\varepsilon_t, x_t\},\, t \geq 1)$ *and* $\|a_0\|_4 < \infty$.

(B4) *For all* t, $r_t = \kappa_x^{(2)}$.

While assumptions (B1) and (B2) specify the properties of $(x_t,\, \varepsilon_t)'$, assumptions (B3) and (B4) impose restrictions on the recursive estimates $(a_t,\, r_t)'$ used by the agent to learn about $(\alpha,\, \kappa_x^{(2)})'$. Assumption (B1) rules out predetermined regressors by requiring x_t to be strictly exogeneous with respect to the error term, while the latter is assumed to be homoskedastic with finite homokurtosis. Conditions (i) and (iii) of assumption (B2) require the conditional mean of certain polynomials of x_t to be independent of ε_t. If $(x_t, t \geq 1)$ is independently distributed of $(\varepsilon_t, t \geq 1)$, conditions ($i$) and ($iii$) are satisfied. When assumptions (B1), part (1) of (B2), (B3) and (B4) are met, the agents' updating scheme (1.4) satisfies the set of so-called *Robbins-Monro* conditions. Recursive algorithms of this class have a long tradition in the stochastic

[4]It is understood that $\|X\|_p := E[|X|^p]^{1/p}$ $(p \geq 1)$ for any random variable with $E[|X|^p] < \infty$. For $p = \infty$: $\|X\|_\infty :=$ ess sup X, i.e. the essential supremum of a random variable that is bounded *a.s.*; see, e.g., Davidson (1994, p. 132).

approximation literature, where they are usually analysed in terms of associated ordinary-differential equations; for more details see, for example, Benveniste et al. (1990, chap. 1.10.1). The present paper builds on-, and extends known results from this literature in order to develop an appropriate asymptotic theory for the estimation and inference procedures discussed below. A simple example of a process satisfying these conditions is where the elements of the triple $(a_0,\, x_t,\, \varepsilon_t)$ are jointly as well as individually independent, identically distributed (*i.i.d.*) with $\|a_0\|_4 < \infty$, $\|\varepsilon_1\|_4 < \infty$ and $\|x_1\|_8 < \infty$. Part (2) of assumption (B2) allows the regressor to be weakly dependent at the expense of an almost sure bound on x_t. Here, weak dependence is measured using the notion of strong mixing[5], a type of asymptotic independence defined in the following way: assume that $(x_t,\, t \geq 1)$ is defined on the probability space $(\Omega, \mathcal{F}, \mathrm{P})$, then the process $(x_t,\, t \geq 1)$ is called strong mixing if $\alpha(m) \to 0$ as $m \to \infty$, where

$$\alpha(m) := \alpha(\mathcal{F}_1^k, \mathcal{F}_{k+m}^\infty), \text{ with } \alpha(\mathcal{A}, \mathcal{B}) := \sup |\mathrm{P}(A \cap B) - \mathrm{P}(A)\mathrm{P}(B)|, \tag{2.6}$$

for any $A \in \mathcal{A} \subseteq \mathcal{F}_1^k$, $B \in \mathcal{B} \subseteq \mathcal{F}_{k+m}^\infty$, and $\mathcal{F}_a^b := \sigma(\{x_t,\, a \leq t \leq b\})$; see, for example, Bradley (2005). Examples of strong mixing include, among others, m-dependent processes, Gaussian processes with a continuous spectral density that is bounded away from zero and ARMA processes under certain smoothness and summability conditions; see, for example, Ibragimov and Linnik (1971) and the discussion in Davidson (1994, pp. 209). While the regressor might exhibit time-series dependence, serially uncorellatedness of the error term is a necessary condition for consistent estimability of the least-squares estimators of $(\beta,\, \delta)'$ discussed below. The reason is that their estimation involves moment conditions of the form $E[(a_{t-1} - \alpha)x_t\varepsilon_t] = 0$. Since a_t contains the complete history of the innovation at time t, the preceding moment conditions are violated whenever ε_t is serially correlated. Finally, assumption (B4) requires that $r_t = \kappa_x^{(2)}$ for all t, a restriction motivated by $r_t \to \kappa_x^{(2)}$ *a.s.*; cf. lemma A.6. Although the Monte-Carlo experiment in section 3 suggests that condition (B4) might be relaxed, it will nevertheless be retained for reasons of analytical tractability.

2.2 Joint estimation of β, δ and a plug-in estimator of α

Consider the joint OLS estimator of $\lambda = (\beta,\, \delta)'$ given by

$$\widehat{\lambda} := \left(\sum_{t=1}^T w_t w_t'\right)^{-1} \sum_{t=1}^T w_t y_t, \quad \text{with} \quad w_t := (a_{t-1}x_t, x_t)'. \tag{2.7}$$

As a starting point for a discussion of the statistical properties of $\widehat{\lambda}$, observe that the regressor w_t fails Grenander's conditions (see, e.g., Hannan (1970, p. 215)), as both the sample second

[5]The technical reason for adopting the concept of strong mixing is twofold: First, measurable transformations of mixing processes are known to be themselves mixing (see, e.g., Davidson (1994, theorem 14.1)); a particularly useful property given the highly nonlinear structure of the underlying data-generating process. Second, a crucial part of the asymptotic theory is concerned with evaluating moments of products of functions of $(x_t,\, t \geq 1)$. As shown in the appendix, this non-trivial problem can by solved using mixingale laws of McLeish (1975), for which assumption (B2) constitutes a sufficient condition.

moment matrix of the regressor

$$M_T \coloneqq \sum_{t=1}^{T} w_t w_t' \tag{2.8}$$

as well as its inverse, both suitably normalized, are asymptotically singular, i.e.

$$T^{-1}M_T \xrightarrow{p} \kappa_x^{(2)} \begin{bmatrix} \alpha^2 & \alpha \\ \alpha & 1 \end{bmatrix} \quad \text{and} \quad T^b M_T^{-1} \xrightarrow{p} \frac{2cb}{\gamma\sigma^2} \begin{bmatrix} 1 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix}, \tag{2.9}$$

where $c \coloneqq 1 - \beta$ and $b \coloneqq 1 - \eta$; the proof of the preceding display can be found in appendix B.1. Intuitively, the singularity of the empirical second moment matrix stems from the fact that the entries of w_t are asymptotically collinear, since a_t converges in mean-square to α as t tends to infinity; cf. appendix A.3. Another crucial aspect of (2.9) is the slower convergence rate of the inverse as compared to the regressor sample moment matrix. To shed more light at this behaviour, define the non-singular transformation matrix

$$G \coloneqq \begin{bmatrix} 1 & 0 \\ -\alpha & 1 \end{bmatrix}, \tag{2.10}$$

which can then be used to rephrase the actual law of motion (1.5):

$$y_t = w_t'\lambda + \varepsilon_t = w_t'GG^{-1}\lambda + \varepsilon_t = \beta x_t(a_{t-1} - \alpha) + \alpha x_t + \varepsilon_t. \tag{2.11}$$

The second component in the transformed regression is well-behaved, while the first regressor involves a random, evaporating trend, i.e. $x_t(a_{t-1} - \alpha) = o_p(1)$ as $t \to \infty$; a setting similar to that of regression models with slowly varying- or evaporating (deterministic) trends discussed by Phillips (2007, 2016). Hence, the convergence rate in the direction of the regular regression component dominates that of the evaporating regressor. Specifically,

$$D_T^{-1/2} G' M_T G D_T^{-1/2} \xrightarrow{p} \begin{bmatrix} \sigma^2\gamma/(2cb) & 0 \\ 0 & \kappa_x^{(2)} \end{bmatrix}, \tag{2.12}$$

where $D_T = \operatorname{diag}(T^b, T)$; cf. appendix B.2.[6] The above implies the following asymptotic representation of the suitably normalized OLS estimator

$$T^{b/2}(\widehat{\lambda} - \lambda) \overset{a}{\sim} \frac{2cb}{\gamma\sigma^2} \begin{bmatrix} 1 \\ -\alpha \end{bmatrix} \frac{1}{T^{b/2}} \sum_{t=1}^{T} (a_{t-1} - \alpha) x_t \varepsilon_t. \tag{2.13}$$

An intriguing feature of the partial sum $\sum_t (a_{t-1} - \alpha) x_t \varepsilon_t$ is that the variances of its sequence coordinates are proportional to γ_t. Thus, in order to deliver the asymptotic distribution of the estimator, one has to appeal to a CLT that allows for potentially degenerate variances. This can be achieved by resorting to a result by Davidson (1993, corollary 2.2), thereby yielding

[6] Put differently, the largest eigenvalue of M_T converges in probability to a finite constant when scaled by T^{-1}, while the second has to be re-normalized by T^{-b} to ensure convergence to a non-degenerate probability limit.

the following singular limiting distribution:

Proposition 2.1 *Suppose assumptions* A *and* B *hold. Then,*

$$T^{b/2}(\widehat{\lambda} - \lambda) \xrightarrow{d} \mathcal{N}\big(0_2, V\big), \quad with \;\; V := \frac{2cb}{\gamma}\begin{bmatrix} 1 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix},$$

where 0_k *denotes a* k*-dimensional column vector of zeros,* $c = 1 - \beta$ *and* $b = 1 - \eta$*.* V *can be estimated consistently by* $T^b V_T$*, where* $V_T := \widehat{\sigma}^{\,2} M_T^{-1}$ *with*

$$\widehat{\sigma}^{\,2} := \frac{1}{T}\sum_{t=1}^{T} \widehat{\varepsilon}_t^{\,2} \quad and \;\; \widehat{\varepsilon}_t := y_t - \widehat{\lambda}' w_t.$$

Proof. See appendix B.1. □

As in the context of a linear regression model, one would expect the asymptotic distribution of the preceding display to depend on the error variance and/or the regressor second moment. Interestingly, neither is the case here with λ, γ and η completely determining the limiting variance-covariance matrix V. This raises the question of whether its consistent estimability requires a priori knowledge of these model parameters. As shown by proposition 2.1 however, the usual variance-covariance estimator V_T is consistent, i.e. given a sample of $(y_t,\, w_t')'$ is at hand, one does not need to know the aforementioned model parameters for estimation purposes. This, in turn, validates the use of the classical standard errors for the elements of the joint estimator $\widehat{\lambda} := (\widehat{\lambda}_\beta,\, \widehat{\lambda}_\delta)'$, i.e.

$$\mathrm{SE}(\widehat{\lambda}_\beta) := \sqrt{m^{11}\,\widehat{\sigma}^{\,2}} \text{ and } \mathrm{SE}(\widehat{\lambda}_\delta) := \sqrt{m^{22}\,\widehat{\sigma}^{\,2}}, \tag{2.14}$$

where m^{ii} represents the i^{th} diagonal element of M_T^{-1}, with the sample second moment matrix $M_T = \sum_{t=1}^{T} w_t w_t'$ being defined in equation (2.8); see also remark B.1 to appendix B.1 for a proof of consistency. Using a different approach and imposing the assumption that $x_t = 1$, Christopeit and Massmann (2018) establish a similar result in their theorem 4 for the case of recursive-least squares learning (i.e. $\eta = 1$). The (singular) variance covariance matrix in Christopeit and Massmann (2018) is equivalent to that stated above up to a factor of proportionality, which stems from the different specification of the gain. The crucial difference between their result and proposition 2.1 is the rate of convergence: while Christopeit and Massmann (2018) show that $\widehat{\lambda}$ converges at a logarithmic rate, it is seen that for slowly decreasing gains $\widehat{\lambda}$ converges at a faster, polynomial rate. A trade-off between the rate at which the agent learns α (increasing in η) and the convergence rate of $\widehat{\lambda}$ (decreasing in η) becomes thus apparent –with the limiting case of $\eta = 1$ treated by Christopeit and Massmann (2018).

The singularity of the limiting distribution of $\widehat{\lambda}$ means that a linear combination of its entries, namely $V^\perp \widehat{\lambda}$, with $V^\perp := (\alpha,\, 1)$, converges at a rate higher[7] than $T^{b/2}$. Note that

[7] For another example from the econometric literature where the singularity of the limiting distribution arises due to the 'super-consistency' of a linear-combination of the joint estimator, see remark 2.1 below or the discussion in Lütkepohl and Burda (1997) on testing non-causality restrictions using vector autoregressions.

$V^{\perp}$, which constitutes the second row of the inverted transformation matrix

$$G^{-1} = \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix}, \tag{2.15}$$

is the orthogonal complement of the limiting variance-covariance V, i.e. $V^{\perp}V = 0_k'$. As summarized by corollary 2.1 below, the linear combination of the joint OLS estimator formed by $V^{\perp}$ converges at the 'standard' rate of $T^{1/2}$ to the RE equilibrium parameter $\alpha = V^{\perp}\lambda$.

Corollary 2.1 *Define $\tau^2 := \sigma^2/\kappa_x^{(2)}$. Then,*

$$D_T^{1/2} G^{-1}(\widehat{\lambda} - \lambda) \xrightarrow{d} \mathcal{N}(0_2, V_1), \quad \textit{with } V_1 := \begin{bmatrix} 2cb/\gamma & 0 \\ 0 & \tau^2 \end{bmatrix}.$$

Proof. See appendix B.2. □

Hence, even though the limiting distribution of the joint OLS estimator is singular, a suitable rotation converges in distribution to a normal random-vector with non-degenerate variance-covariance matrix. Clearly, $V^{\perp}\widehat{\lambda}$ does not present a feasible strategy for estimating α. Instead, consider the simple plug-in estimator

$$\widehat{\lambda}_\alpha := \widehat{\lambda}_\delta/(1 - \widehat{\lambda}_\beta), \tag{2.16}$$

with $\widehat{\lambda}_\beta$ and $\widehat{\lambda}_\delta$ denoting respectively the two components of $\widehat{\lambda} := (\widehat{\lambda}_\beta,\, \widehat{\lambda}_\delta)'$.[8] Since $\widehat{\lambda}_\alpha - \alpha = V^{\perp}(\widehat{\lambda} - \lambda)/(1 - \widehat{\lambda}_\beta)$, the asymptotic normality of (2.16) follows as a by-product of the preceding corollary, i.e.

$$T^{1/2}V^{\perp}(\widehat{\lambda} - \lambda) \xrightarrow{d} \mathcal{N}(0, \tau^2) \text{ and } T^{1/2}(\widehat{\lambda}_\alpha - \alpha) \xrightarrow{d} \mathcal{N}(0, (\tau/c)^2); \tag{2.17}$$

cf. appendix B.2. The limiting variance[9] $(\tau/c)^2$ is consistently estimated by $T\mathrm{SE}(\widehat{\lambda}_\alpha)^2$, where

$$\mathrm{SE}(\widehat{\lambda}_\alpha) := \widehat{\tau}(1 - \widehat{\lambda}_\beta)^{-1}, \text{ with } \widehat{\tau} := \widehat{\sigma}\left(\sum_{t=1}^{T} x_t^2\right)^{-1/2}. \tag{2.18}$$

Since the plug-in estimator (2.16) equals $\phi(x, y) = y/(1 - x)$ ($|x| < 1$, $y \in \mathbb{R}$) evaluated at $\widehat{\lambda}$, the Δ-method provides an alternative standard-error

$$\mathrm{SE}_\Delta(\widehat{\lambda}_\alpha) := \sqrt{\nabla\phi(\widehat{\lambda})'\, V_T\, \nabla\phi(\widehat{\lambda})}, \tag{2.19}$$

with $\nabla\phi$ denoting the gradient vector of ϕ. Interestingly, whereas $\widehat{\lambda}_\alpha$ is strictly less efficient than the infeasible estimator $V^{\perp}\widehat{\lambda}$ for $0 < \beta < 1$, there exist efficiency gains for using $\widehat{\lambda}_\alpha$ when $-1 < \beta < 0$.

[8] A simple calculation shows that $\widehat{\lambda}_\alpha = (\widehat{\lambda}_\alpha,\, 1)\widehat{\lambda}$.

[9] Note that $(\tau/c)^2$ coincides with the lower bound established in the literature on stochastic approximation for the 'averaged iterates estimator' of α (see, e.g., Polyak and Juditsky (1992), Kushner and Yan (1993) or Chen (1993)).

Remark 2.1 *Concerning the different convergence rates of* $\widehat{\lambda}$ *on the one hand, and* $\widehat{\lambda}_\alpha$ *on the other, one is reminded of estimating co-integrating vectors based on autoregressive distributed lag* (ADL) *models. Specifically, consider the* ADL(1,0) *model*

$$y_t = \beta y_{t-1} + \delta x_t + u_t, \tag{ADL}$$

where $x_t \sim I(1)$, $y_t - \alpha x_t \sim I(0)$ *for some* α, *and* u_t *is white noise. An error-correction reparametrization of equation* (ADL), *namely,*

$$\Delta y_t = -(1-\beta)(y_{t-1} - \alpha x_{t-1}) + \delta \Delta x_t + u_t \;\; \textit{with} \;\; \alpha = \delta/(1-\beta), \tag{ECM}$$

suggests the following two-step approach dating back to Stock (1987): first estimate the coefficients of the short-run dynamics $(\beta, \delta)'$ *based on equation* (ADL) *and then use these estimates,* $\widehat{\beta}$ *and* $\widehat{\delta}$, *say, to form the plug-in estimator* $\widehat{\alpha} = \widehat{\delta}/(1-\widehat{\beta})$ *of the co-integrating coefficient* α. *Pesaran and Shin (1998) have shown that the joint OLS estimator of the short-run dynamics is root-T consistent with* (*singular*) *asymptotic normal distribution, while the plug-in estimator converges in distribution when scaled by* T*; see also Wickens and Breusch (1988), Banerjee et al. (1993) and Hassler and Wolters (2006). This approach to estimating co-integrating vectors bears thus considerable similarity to the estimation procedure described before: in order to estimate the parameter* α *governing the long-run RE equilibrium* (1.2), *one first estimates the parameters* β *and* δ *of the actual law-of-motion, prescribing the short-run deviations from this equilibrium:*

$$y_t = \beta a_{t-1} x_t + \delta x_t + \varepsilon_t,$$

and then uses the so-obtained estimates to form a 'super-consistent' plug-in estimator of the long-run coefficient.

Remark 2.2 *The preceding limiting theory extends naturally from the single variable case to a more general setting in which* $\boldsymbol{x}_t$ *represents a column vector of* $k \geq 1$ *regressors, so that the actual law of motion is given by*

$$y_t = \beta \boldsymbol{a}_{t-1}' \boldsymbol{x}_t + \boldsymbol{\delta}' \boldsymbol{x}_t + \varepsilon_t, \quad \text{for} \quad t = 1, 2, \ldots, T, \tag{2.20}$$

with $|\beta| < 1$ *and* $\boldsymbol{\delta} \in \mathbb{R}^k$. *In order to form the forecast of the RE equilibrium vector* $\boldsymbol{\alpha} = \boldsymbol{\delta}/(1-\beta)$ *at time* t, *the agent is assumed to use the following multivariate recursive least-squares type algorithm*

$$\begin{aligned} \boldsymbol{a}_t &= \boldsymbol{a}_{t-1} + \gamma_t \mathbf{R}_t^{-1} \boldsymbol{x}_t (y_t - \boldsymbol{a}_{t-1}' \boldsymbol{x}_t) \\ \mathbf{R}_t &= \mathbf{R}_{t-1} + \gamma_t (\boldsymbol{x}_t \boldsymbol{x}_t' - \mathbf{R}_{t-1}). \end{aligned} \tag{2.21}$$

Now, let $\widehat{\boldsymbol{\lambda}}$ *denote the* $(k+1)$*-dimensional OLS estimator of* $\boldsymbol{\lambda} := (\beta, \boldsymbol{\delta}')'$ *based on* (2.20) *and define* $\mathbf{V}^{\perp} := (\boldsymbol{\alpha}, \mathbf{I}_k)$, *a* $k \times (k+1)$ *matrix, which results from aligning* $\boldsymbol{\alpha}$ *and* $\mathbf{I}_k$ *columnwise so that* $\mathbf{V}^{\perp}\boldsymbol{\lambda} = \boldsymbol{\alpha}$. *Just as in the scalar case treated above, it follows that* $T^{b/2}(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})$ *is asymptotically normal while the linear combination* $\mathbf{V}^{\perp}(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})$ *converges to its limiting normal* *distribution at the faster rate $T^{1/2}$. Specifically, define the $(k+1)\times(k+1)$ matrix*

$$\mathbf{G} \coloneqq \begin{bmatrix} 1 & \mathbf{0}_k' \\ -\boldsymbol{\alpha} & \mathbf{I}_k \end{bmatrix}, \tag{2.22}$$

which has the property $\mathbf{G}^{-1}\boldsymbol{\lambda} = (\beta,\, \boldsymbol{\alpha}')'$; and introduce the multivariate scaling matrix of dimension $(k+1)\times(k+1)$

$$\mathbf{D}_T \coloneqq \begin{bmatrix} T^b & \mathbf{0}_k' \\ \mathbf{0}_k & T\mathbf{I}_k \end{bmatrix}. \tag{2.23}$$

Under the assumption that $(\varepsilon_t,\, \boldsymbol{x}_t')'$ is i.i.d. and regularity conditions which extend those imposed by assumption (B2), *it is shown in appendix* S.1 *that*

$$T^{b/2}(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}_{k+1}, \frac{2cb}{k\gamma}\begin{bmatrix} 1 & -\boldsymbol{\alpha}' \\ \boldsymbol{\alpha} & \boldsymbol{\alpha}\boldsymbol{\alpha}' \end{bmatrix}\right) \tag{2.24}$$

and

$$\mathbf{D}_T^{1/2}\mathbf{G}^{-1}(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}_{k+1}, \begin{bmatrix} 2cb/(k\gamma) & \mathbf{0}_k' \\ \mathbf{0}_k & \sigma^2\mathbf{Q}^{-1} \end{bmatrix}\right). \tag{2.25}$$

The author believes that the preceding display remains to hold if the elements of $\boldsymbol{x}_t'$ are allowed to exhibit serial dependence as specified in part (2) *of assumption* (B2). *Again, it is seen that $\mathbf{V}^{\perp}$ is the orthogonal complement of the rank-one variance-covariance matrix $\mathbf{V}$, i.e. $\mathbf{V}^{\perp}\mathbf{V} = \mathbf{O}'$. The only material difference between the preceding display and its counterpart of corollary* 2.1 *is the appearance of $k = \dim(\boldsymbol{x}_t)$ in the denominator of the limiting variance-covariance matrix $\mathbf{V}$. Equation* (2.24) *can be used to carry out inference about $(\beta,\, \boldsymbol{\delta}')'$ and to construct a plug-in estimator of $\boldsymbol{\alpha}$ analogously to* (2.17).

2.3 Hypothesis testing

Suppose you are interested in testing single hypotheses of the form H_0: $\beta = \beta_0$ or H_0: $\delta = \delta_0$ using the joint OLS estimator $\widehat{\lambda} = (\widehat{\lambda}_\beta,\, \widehat{\lambda}_\delta)'$. As a natural starting-point, consider the textbook t statistics

$$t_\beta \coloneqq \frac{\widehat{\lambda}_\beta - \beta_0}{\mathrm{SE}(\widehat{\lambda}_\beta)} \quad \text{and} \quad t_\delta \coloneqq \frac{\widehat{\lambda}_\delta - \delta_0}{\mathrm{SE}(\widehat{\lambda}_\delta)}, \tag{2.26}$$

where the standard errors have been defined in equation (2.14). As shown in remark B.1 to appendix B.1, t_β and t_δ are asymptotically standard normally distributed, which allows us to draw inferences from $\widehat{\lambda}$ about the structural parameters β or δ. For a null hypothesis H_0: $\alpha = \alpha_0$, the corresponding t statistic based on $\widehat{\lambda}_\alpha$ is similarly defined as

$$t_\alpha \coloneqq \frac{\widehat{\lambda}_\alpha - \alpha_0}{\mathrm{SE}(\widehat{\lambda}_\alpha)}. \tag{2.27}$$

A standard normal null-distribution of (2.27) results directly from equations (2.17) and (2.18) as $t_\alpha = T^{1/2}V_0^\perp(\widehat{\lambda} - \lambda_0)/\widehat{\tau}$, where $V_0^\perp = (\alpha_0,\, 1)'$ and $\lambda_0 \in \mathbb{R}^2$ is such that $V_0^\perp \lambda_0 = \alpha_0$. The case of joint hypotheses about λ is slightly more involved due to the singularity of the limiting distribution of the joint OLS estimator $\widehat{\lambda}$. In order to fix ideas, consider the classical Wald-statistic

$$\mathcal{W} := (\widehat{\lambda} - \lambda_0)' V_T^{-1} (\widehat{\lambda} - \lambda_0) \tag{2.28}$$

for a null hypothesis H_0: $\lambda = \lambda_0 \in \mathbb{R}^2$; recall that $V_T = \widehat{\sigma}^2 M_T^{-1}$ is the usual variance-covariance estimator defined in proposition 2.1. As summarized by proposition 2.1, $T^{b/2}(\widehat{\lambda} - \lambda_0)$ is asymptotically normal under the null with singular variance-covariance matrix

$$V_0 = \frac{2cb}{\gamma} \begin{bmatrix} 1 & -\alpha_0 \\ -\alpha_0 & \alpha_0^2 \end{bmatrix}, \tag{2.29}$$

which is consistently estimated by $T^b V_T$. If V_0 were non-singular, straightforward application of Slutzky's theorem and the continuous mapping theorem would reveal the asymptotic Chi-square distribution of $\mathcal{W}$. Here, with V_0 being of reduced rank, this line of reasoning fails as V^{-1} is not defined. Invertibility of the limiting variance-covariance matrix is, however, not a necessary condition for $\mathcal{W}$ to have an asymptotic Chi-square distribution; see, for example, Andrews (1987, Comment no. 7, p. 353). Indeed, as shown by corollary 2.1, the limiting distribution of the joint OLS estimator is non-singular *upon transformation.* Since the Wald-statistic is not invariant to transformations, i.e.

$$\mathcal{W} = \frac{(\widehat{\lambda} - \lambda_0)' G_0^{-1\prime} D_T^{1/2} (D_T^{-1/2} G_0' M_T G_0 D_T^{-1/2}) D_T^{1/2} G_0^{-1} (\widehat{\lambda} - \lambda_0)}{\widehat{\sigma}^2}, \tag{2.30}$$

with

$$G_0 := \begin{bmatrix} 1 & 0 \\ -\alpha_0 & 1 \end{bmatrix}, \tag{2.31}$$

the following is an immediate consequence of corollary 2.1:

Corollary 2.2 *Suppose assumptions* A *and* B *hold. Then, for sequences of local alternatives given by*

$$H_T\colon \lambda = \lambda_T, \ \ \textit{with } \lambda_T := \lambda_0 + T^{-1/2}\mu \ \textit{ and } \mu := (\mu_\beta,\, \mu_\delta)' \in \mathbb{R}^2, \tag{2.32}$$

it follows

$$\mathcal{W} \xrightarrow{d} \chi^2(2, \xi), \ \ \textit{with } \xi = \frac{\mu' M_0 \mu}{\sigma^2} \ \textit{ and } M_0 = \kappa_x^{(2)} \begin{bmatrix} \alpha_0^2 & \alpha_0 \\ \alpha_0 & 1 \end{bmatrix}, \tag{2.33}$$

where $\chi^2(m, n)$ denotes a Chi-square distribution with $m \in \mathbb{Z}$ degrees of freedom and non-centrality parameter $n \geq 0$.

Proof. See appendix B.3. □

Another cure to the problem of hypothesis tests in the presence of a singular variance-covariance matrix V lies in replacing the proper inverse with a generalized inverse; see, for example, Andrews (1987) and Lütkepohl and Burda (1997). To be more specific, recall that for any two-dimensional, singular matrix B the (Moore-Penrose) generalized inverse is defined via

$$B^+ := A \begin{bmatrix} 1/a & 0 \\ 0 & 0 \end{bmatrix} A', \tag{2.34}$$

with A and a denoting respectively the 2×2 matrix of eigenvectors and the largest eigenvalue of B; see, for example, Seber (2008, chapter 7). Hence,

$$V_0^+ = \frac{\gamma}{2cb(1+\alpha_0^2)^2} \begin{bmatrix} 1 & -\alpha_0 \\ -\alpha_0 & \alpha_0^2 \end{bmatrix}. \tag{2.35}$$

Yet the pseudo-inverse (2.34) is not continuous and V_T violates the rank condition of Andrews (1987): $\mathrm{rk}(T^b V_T) > \mathrm{rk}(V)$ uniformly in T with probability one. In consequence,

$$(T^b V_T)^+ = (T^b V_T)^{-1} \overset{p}{\not\to} V_0^+,$$

even though $T^b V_T$ is consistent for V. However, taking (2.35) and the results of the preceding paragraph into account, V^+ can be direclty estimated via

$$V_T^- := \frac{\gamma}{2b(1-\widehat{\lambda}_\beta)(1+\widehat{\lambda}_\alpha^2)^2} \begin{bmatrix} 1 & -\widehat{\lambda}_\alpha \\ -\widehat{\lambda}_\alpha & \widehat{\lambda}_\alpha^2 \end{bmatrix}. \tag{2.36}$$

Since $V_T^- - V^+ = o_p(1)$, theorem 1 in Andrews (1987) applies, so that, for a sequence of local alternatives given by

$$H_T\colon \lambda = \lambda_T, \ \text{ with } \lambda_T := \lambda_0 + T^{-b/2}\mu \ \text{ and } \ \mu := (\mu_\beta,\, \mu_\delta)' \in \mathbb{R}^2, \tag{2.37}$$

it follows that

$$\mathcal{W}_+ := T^b(\widehat{\lambda} - \lambda_0)' V_T^- (\widehat{\lambda} - \lambda_0) \xrightarrow{d} \chi^2(1, \kappa), \ \text{ with } \kappa = \mu' V^+ \mu; \tag{2.38}$$

see also the discussion in Lütkepohl and Burda (1997). Hence, the classical Wald statistic as well as its generalized counterpart equipped with a consistent estimator of V^+ can be used to test joint hypotheses. The results of the local power analysis indicate, however, the superiority of the classical Wald statistic: comparing (2.32) to (2.37), it is evident that $\mathcal{W}_+$ detects violations only if deviations from the null are relatively large (depending on η), while the local power of $\mathcal{W}$ remains non-trivial also against alternatives in much narrower neighborhoods of the null. Furthermore, with M_0 and V_0^+ being positive semi-definite, there exist non-trivial local alternatives for which the non-centrality parameters ξ and κ are zero.

Specifically, as

$$\xi = \left(\frac{\alpha_0\mu_\beta + \mu_\delta}{\tau}\right)^2 \text{ and } \kappa = \frac{\gamma}{2cb}\left(\frac{\mu_\beta - \alpha_0\mu_\delta}{1+\alpha_0^2}\right)^2, \tag{2.39}$$

$\mathcal{W}$ and $\mathcal{W}_+$ are seen to lack power in the direction of the orthogonal complements of M_0 and V_0^+, respectively: that is, $\mathcal{W}$ has no power against $\mu = c(-1,\, \alpha_0)'$, while $\mathcal{W}_+$ has no power against $\mu = c(\alpha_0,\, 1)' = cV_0^{+\perp}$, for some constant $c \neq 0$.

2.4 An asymptotically equivalent three-step estimator

This section presents an asymptotically equivalent approach to the estimation procedure for $(\alpha,\, \beta,\, \delta)$ discussed at the beginning of this paragraph. Suppose, to begin with, that the true RE equilibrium parameter α is known and the interest lies in estimating β. Then, after a little rearrangement, the actual law of motion $y_t = \beta a_{t-1}x_t + \delta x_t + \varepsilon_t$ is rewritten as

$$y_t - \alpha x_t = \beta(a_{t-1} - \alpha)x_t + \varepsilon_t. \tag{2.40}$$

It is thus intuitively appealing to estimate β using the OLS estimator

$$\widehat{\beta}_0 := \frac{\sum_{t=1}^{T}(y_t - \alpha x_t)(a_{t-1} - \alpha)x_t}{\sum_{t=1}^{T}((a_{t-1} - \alpha)x_t)^2}, \tag{2.41}$$

which, indeed, possesses a limiting normal distribution:

Proposition 2.2 *Suppose that assumptions* A *and* B *hold. Then,*

$$T^{b/2}(\widehat{\beta}_0 - \beta) \xrightarrow{d} \mathcal{N}(0, 2cb/\gamma).$$

Proof. See appendix B.4. □

A comparison between proposition 2.1 and 2.2 reveals the asymptotic equivalence between $\widehat{\beta}_0$ and $\widehat{\lambda}_\beta$ (the first element of the joint OLS estimator $\widehat{\lambda}$). Since $\widehat{\beta}_0$ depends on the unknown RE equilibrium parameter, this approach is infeasible unless a suitable estimator of α is available. Under RE, one might resort to estimating α based on the RE equilibrium (1.2), yielding the OLS estimator

$$\widehat{\alpha} := \frac{\sum_{t=1}^{T} y_t x_t}{\sum_{t=1}^{T} x_t^2}. \tag{2.42}$$

But with agents' beliefs departing from RE, it is not clear why $\widehat{\alpha}$ should possess any desirable statistical properties. It can be shown, however, that although y_t is generated by (1.5)−rather than by the RE equilibrium (1.2)−the limiting distribution of (2.42) is normal with mean α when suitably scaled:

Proposition 2.3 *Let the conditions of proposition 2.2 hold. Then,*

$$T^{1/2}(\widehat{\alpha} - \alpha) \xrightarrow{d} \mathcal{N}(0, (\tau/c)^2).$$

Proof. See appendix B.5. □

Again, the limiting distributions of $\widehat{\lambda}_\alpha$ (cf. corollary 2.1) and $\widehat{\alpha}$ coincide while the normalization of $T^{1/2}$ implies a faster convergence rate than that of $\widehat{\beta}_0$. This 'super-consistency' of $\widehat{\alpha}$ allows the estimation of β in a second step by

$$\widehat{\beta} \coloneqq \frac{\sum_{t=1}^T \tilde{y}_t \tilde{x}_t}{\sum_{t=1}^T \tilde{x}_t^2}, \tag{2.43}$$

where $\tilde{y}_t \coloneqq y_t - \widehat{\alpha} x_t$ and $\tilde{x}_t \coloneqq (a_{t-1} - \widehat{\alpha}) x_t$. As shown in remark B.2 to appendix B.5, estimation and inference based on $\widehat{\beta}$ is not contaminated by a 'generated regressor' problem known from the discussion in Pagan (1984), i.e.

$$T^{b/2}(\widehat{\beta} - \beta) = T^{b/2}(\widehat{\beta}_0 - \beta) + o_p(1). \tag{2.44}$$

Put differently, knowledge of α does not improve the estimation of β. Returning thus to the co-integration analogy drawn in remark 2.1, this estimation procedure is clearly reminiscent of the two-step approach to estimating co-integrating vectors and associated error correction models popularized by Engle and Granger (1987). Finally, consider the 'three-step' estimator

$$\widehat{\delta} \coloneqq \widehat{\alpha}(1 - \widehat{\beta}). \tag{2.45}$$

Exploiting the difference in convergence rates between $\widehat{\alpha}$ on the one hand, and $\widehat{\beta}$ on the other, it follows that

$$T^{b/2}(\widehat{\delta} - \delta) = -\alpha T^{b/2}(\widehat{\beta} - \beta) + o_p(1); \tag{2.46}$$

cf. appendix B.5 (remark B.2). The result in (2.46), together proposition 2.3 and equation (2.44), shows for $\widetilde{\lambda} \coloneqq (\widehat{\beta}, \widehat{\delta})'$ that

$$T^{b/2}(\widetilde{\lambda} - \lambda) \overset{a}{\sim} T^{b/2}(1, -\alpha)'(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0_2, V), \tag{2.47}$$

i.e., the vector of sequential estimators $\widetilde{\lambda}$ and the joint OLS estimator $\widehat{\lambda}$ share the same limiting distribution; compare proposition 2.1.[10]

3 Monte-Carlo experiment

Consider a data-generating process (DGP) given by the actual law of motion

$$y_t = \beta a_{t-1} x_t + \delta x_t + \varepsilon_t, \tag{3.48}$$

with $\delta = 1/2$ and $\beta = -2/3$, implying for the RE equilibrium that $\alpha = 0.3$. The regressor evolves over time according to the first-order autoregression $x_t = 1/5 - 4/5\, x_{t-1} + u_t$, where u_t is identically and independently drawn from a mean-zero t-distribution with ten degrees

[10] As a result of a purely algebraic exercise, the joint OLS estimator $\widehat{\lambda}$ can be recovered by replacing α with $\widehat{\lambda}_\alpha$ in (2.40).

of freedom; $\text{var}[u_t]$ is chosen such that $E[x_t^2] = 1$; x_0 is set equal to the unconditional mean $E[x_t] = 1/9$. The innovation ε_t is an *i.i.d.* standardized t-distributed random variable with five degrees of freedom. The agent updates her beliefs about α according to the learning-scheme (1.4), which is reproduced here for convenience:

$$a_t = a_{t-1} + \gamma_t \frac{x_t}{r_t}(y_t - x_t a_{t-1}), \text{ with } r_t = r_{t-1} + \gamma_t (x_t^2 - r_{t-1}).$$

The initial values $(a_0, r_0)'$ are randomly drawn according to $a_0 \sim \mathcal{U}[-2, 2]$ and $r_0 \sim \mathcal{U}[1/5, 3]$, with $\mathcal{U}[a, b]$ $(a \leq b)$ denoting the uniform distribution. It is noteworthy that assumption (B4) (i.e. $r_t = E[x_t^2]$) is not imposed.[11] The gain sequence is chosen as $\gamma_t = t^{-\eta}$, with $\eta \in \{0.6, 0.7, 0.8, 0.9\}$. Monte-Carlo estimates (using 10,000 iterations) of 100 times the bias and size of a two-sided t-test (the t-statistics are defined in equations (2.26) and (2.27)) at the 5% significance level are reported in table 1 for the joint OLS estimator $\widehat{\lambda} = (\widehat{\lambda}_\beta, \widehat{\lambda}_\delta)'$ of $\lambda = (\beta, \delta)'$ (cf. equation (2.7)) and the plug-in estimator $\widehat{\lambda}_\alpha$ of α (cf. equation (2.16)). Results for the (asymptotically equivalent) three-step procedure do not differ substantially and are therefore omitted for the sake of better clarity and readability.

Table 1: Simulation results – Joint OLS estimator[a]

		$\eta = 0.6$		$\eta = 0.7$		$\eta = 0.8$		$\eta = 0.9$	
	T	bias	size	bias	size	bias	size	bias	size
β	250	-2.45	4.62	-2.80	4.35	-3.19	4.45	-4.40	4.25
	500	-1.99	4.53	-2.92	4.50	-3.03	4.34	-4.21	4.44
	1,000	-1.25	4.99	-1.92	4.89	-2.92	4.64	-3.99	4.58
δ	250	0.76	4.85	0.87	4.50	1.20	4.65	1.32	4.16
	500	0.51	4.70	0.86	4.79	1.13	4.48	1.24	4.10
	1,000	0.50	5.01	0.68	4.81	0.93	4.85	1.04	4.38
α	250	0.02	5.71	0.02	6.01	-0.04	6.35	-0.05	5.82
	500	-0.05	5.42	-0.01	5.75	0.01	5.93	-0.02	6.01
	1,000	0.00	5.33	0.01	5.50	-0.00	5.76	0.02	5.60

[a] Bias is multiplied by 100. Size refers to rejection frequencies (%) under the null of a two-sided t-test at the 5% significance level using the 0.975 percentile from the standard normal distribution. The construction of the standard errors is outlined in section 2.3.

In case of $\widehat{\lambda}$, the asymptotic approximations provide a more accurate description of the finite sample behaviour the smaller η. As discussed in the preceding chapter, this is due to the fact that the convergence rate of the joint OLS estimator $\widehat{\lambda}$ is inversely related to η. Observe that the bias of the $\widehat{\lambda}_\delta$ is approximately equal to $-\alpha$ $(= -0.3)$ times the bias of $\widehat{\lambda}_\beta$, which reflects the degenerate limiting distribution of $\widehat{\lambda}$ as summarized by proposition 2.1. On the other hand, the superior performance of the plug-in estimator $\widehat{\lambda}_\alpha$ of α in terms of bias demonstrates its 'super-consistency': i.e., independently of η, $\widehat{\lambda}_\alpha$ converges at the usual root-T rate to its limiting normal distribution (cf. equation (2.17)). Next, size and (local) power properties of the Wald statistic $\mathcal{W}$ (cf. equation (2.28)) and its counterpart $\mathcal{W}_+$ (cf. equation (2.38)) equipped with the Moore-Penrose inverse are reported in table 2. For a null hypothesis H_0:

[11] The following results of the small-sample experiment don't differ substantially from those where assumption (B4) is imposed.

$\lambda = \lambda_0$, with $\lambda_0 = (1/2,\ -3/2)'$, two local alternatives are considered:

$$(i)\ \lambda_T = \lambda_0 + \iota_2 T^{-1/2} \quad \text{and} \quad (ii)\ \lambda_T = \lambda_0 + \iota_2 T^{-b/2},$$

where ι_2 denotes the 2-dimensional vector of ones.

Table 2: Simulation results – Wald tests[a]

		$\eta = 0.6$		$\eta = 0.7$		$\eta = 0.8$		$\eta = 0.9$	
	T	size	power $(i)/(ii)$	size	power $(i)/(ii)$	size	power $(i)/(ii)$	size	power $(i)/(ii)$
$\mathcal{W}$	250	4.83	12.34/94.98	4.82	13.80/95.00	4.95	11.99/95.00	4.79	11.04/95.00
	500	4.96	12.46/95.00	4.96	14.44/95.00	4.86	12.03/95.00	4.43	12.96/95.00
	1,000	5.04	14.21/95.00	4.84	14.21/95.00	4.91	14.26/95.00	4.80	13.53/95.00
$\mathcal{W}_+$	250	4.97	0.60/3.80	4.69	0.46/2.10	4.61	0.30/0.99	4.31	-0.04/7.20
	500	4.70	0.01/3.60	4.66	-0.30/2.30	4.50	-0.08/5.20	4.12	-0.45/8.51
	1,000	4.90	-0.21/3.30	4.70	-0.67/4.54	4.61	-0.72/8.21	4.48	-0.49/9.01

[a] Size refers to rejection frequencies (%) under the null at the 5% significance level using the 0.95 percentile from the Chi-square distribution with one degree of freedom and two degrees of freedom for $\mathcal{W}_+$ and $\mathcal{W}$, respectively. Power refers to size-corrected rejection frequencies (%) for the local alternatives (i) and (ii).

As can be inferred from corollary 2.2, size-corrected (at a 5% significance level) asymptotic local power of the textbook Wald-statistic $\mathcal{W}$ is non-trivial against both sequence of alternatives: 14.61% under (i) and 95% under (ii). In contrast, the generalized Wald-statistic $\mathcal{W}_+$ possesses non-trivial power only against sequences that are relatively far away from the null: i.e. 0% under (i), while 3.61% ($\eta = 0.6$), 4.84% ($\eta = 0.7$), 7.33% ($\eta = 0.8$), and 14.94% ($\eta = 0.9$) under (ii); cf. equation (2.38). As can be seen in table 2, the small sample rejection frequencies closely match the theoretical size and power.

4 Conclusion

This paper establishes the asymptotic equivalence between a newly proposed 'three-step' procedure and the joint OLS estimator of the structural parameters β and δ in a stylized moacroeconomic model of the form

$$y_t = \beta y^e_{t|t-1} + \delta x_t + \varepsilon_t, \tag{4.49}$$

where the agents' expectation formation is boundedly rational in the sense that $y^e_{t|t-1}$ obeys a stochastic approximation algorithm (cf. equation (1.4)). In contrast to Christopeit and Massmann (2018) who set $x_t = 1$, the current exposition treats the regressor as a time-varying random variable. Moreover, the agent is assumed to use slowly-decreasing gains, i.e. $\gamma_t \sim t^{-\eta}$ with $\eta \in (1/2, 1)$; the intermediate case lying on a continuum between least-squares learning ($\eta = 1$) and constant-gain learning ($\eta = 0$) considered by Christopeit and Massmann (2018). The estimators of the 'structural' parameters β and δ converge at a polynomial rate to their joint, singular asymptotic normal distribution. A trade-off between the rate at which the agent learns and the convergence rate of the estimators becomes apparent. Furthermore, two estimators of the RE equilibrium parameter $\alpha = \delta(1-\beta)^{-1}$ emerge as a by-product of the two estimation frameworks. Reminiscent of estimating co-integrating relationships, these estimators converge at the 'standard' (faster) convergence rate $T^{1/2}$, and are thus not subject to the aforementioned trade-off. Finally, it is shown how the limiting results can be used to draw inferences about single and joint hypotheses.

References

Abadir, K. M. and J. R. Magnus (2006). *Matrix Algebra*. Econometric Exercises, Volume 1. Cambridge: Camebridge Univerity Press.

Andrews, D. W. K. (1987). Asymptotic results for generalized Wald tests. *Econometric Theory 3*, 348–358.

Apostol, T. M. (1976). *Introduction to Analytic Number Theory*. Undergraduate texts in mathematics. New York: Springer.

Apostol, T. M. (1999). An elementary view of Euler's summation formula. *The American Mathematical Monthly 106*, 409–418.

Banerjee, A., J. J. Dolado, J. W. Galbraith and D. F. Hendry (1993). *Co-integration, Error Correction, and the Econometric Analysis of Non-stationary Data*. Advanced Texts in Econometrics. Oxford: Oxford University Press.

Benveniste, A., M. Métivier and P. Priouret (1990). *Adaptive Algorithms and Stochastic Approximation*. Applications of Mathematics, Volume 22. Berlin: Springer.

Berardi, M. and J. K. Galimbeti (2014). A note on the representative adaptive learning algorithm. *Economic Letters 124*, 104–107.

Berardi, M. and J. K. Galimbeti (2014). Empirical calibration of adaptive learning. *Journal of Economic Behavior & Organization 144*, 219–237.

Bradley, R.-C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys 2*, 107–144.

Chakraborty, A. and G. W. Evans (2008). Can perpetual learning explain the forward premium puzzle? *Journal of Monetary Economics 55*, 77–490.

Chen, H.-F. (1993). Asymptotically efficient stochastic approximation. *Stochastics and Stochastic Reports 45*, 1–16.

Chen, H.-F. (2002). *Stochastic Approximation and Its Applications*. Nonconvex Optimization and Its Applications, Volume 64. Dordrecht: Kluwer Academic Publishers.

Chung, K. L. (2001). *A Course in Probability Theory*. Third edition. San Diego: Academic Press.

Chevillon, G. and S. Mavroeidis (2017). Learning can generate long memory. *Journal of Econometrics 198*, 1–7.

Christopeit, N. and M. Massmann (2018). Estimating structural parameters in regression models with adaptive learning. *Econometric Theory 34*, 68–111.

Davidson, J. (1993). The central limit theorem for globally nonstationary near-epoch dependent functions of mixing processes: the asymptotically degenerate case. *Econometric Theory 9*, 402–412.

Davidson, J. (1994). *Stochastic Limit Theory*. Advanced Texts in Econometrics. Oxford: Oxford University Press.

Engle, F. E. and C. W. J. Granger (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica 55*, 251–276.

Evans, G. W. and S. Honkapohja (2001). *Learning and Expectations in Macroeconomics*. Princeton: Princeton University Press.

Hannan, E. J. (1970). *Multiple Time Series*. New York: Wiley.

Hassler, U. and J. Wolters (2006). Autoregressive distributed lag models cointegration. *Advances in Statistical Analysis 90*, 59–74.

Ibragimov, I. A. and Y. V. Linnik (1971). *Independent and stationary sequences of random variables*. Groningen: Wolters-Noordhoff.

Jameson, G. J. O. (2013). The incomplete gamma functions. *The Mathematical Gazette 100*, 298–306.

Joag-Dev, K. and F. Proschan (1993). Negative association of random variables with applications. *The Annals of Statistics 11*, 286–295.

Kuan, C. H. and H. White (1994). Adaptive learning with nonlinear dynamics driven by dependent processes. *Econometrica 62*, 1087–1114.

Kushner, H. J. and H. Huang (1981). Asymptotic properties of stochastic approximations with constant coefficients. *Journal on Control and Optimization 19*, 87–105.

Kushner, H. J. and J. Yan (1993). Stochastic approximation with averaging of the iterates: optimal asymptotic rate of convergence for general proceses. *Journal on Control and Optimization 31*, 1045–1062.

Lampret, V. (2001). The EulerMaclaurin and Taylor formulas: twin, elementary derivations. *Mathematics Magazine 74*, 109–122.

Ltkephol, H. and M. M. Burda (1997). Modified Wald tests under nonregular conditions. *Journal of Econometrics 78*, 315–332.

Markiewicz, A. and A. Pick (2014). Adaptive learning and survey data. *Journal of Economic Behavior & Organization 107*, 685–707.

McLeish, D. L. (1975). A maximal inequality and dependent strong laws. *The Annals of Probability 3*, 829–839.

Miao, Y. and S. Xu (2013). Almost sure convergence of weighted sums. *Miskolc Mathematical Notes 14*, 173–181.

Pagan, A. (1984). Econometric issues in the analysis of regressions with generated regressors. *International Economic Review 25*, 221–247.

Pesaran, H. M. and Y. Shin (1998). An autoregressive distributed lag modelling approach to cointegration analysis. In: S. Strom (Ed.), *Econometrics and Economic Theory in the 20th Century. The Ragnar Frisch Centennial Symposium*, 371–413. Camebridge: Cambridge University Press.

Phillips, C. B. B. (2007). Regression with slowly varying regressors and nonlinear trends. *Econometric Theory 23*, 557–614.

Phillips, C. B. B. (2016). Inference in near-singular regression. In: G. González-Rivera, R. Carter Hill and T.-H. Lee (Eds.) *Essays in Honor of Aman Ullah* (Advances in Econometrics, Volume 36). Bingley: Emerald Publishing Ltd.

Polyak, B. T. and A. B. Juditsky (1992). Acceleration of stochastic approximation by averaging. *Journal on Control and Optimization 30*, 838–855.

Rio, E. (1995). A maximal inequality and dependent Marcinkiewicz-Zygmund strong laws. *Annals of Probability 23*, 918–937.

Rosalsky, A. and J. Rosenblatt (1998). On convergence of series of random variables with applications to martingale convergence and to convergence of series with orthogonal summands. *Stochastic Analysis and Applications 16*, 553–566.

Sargent, T. J. (1993). *Bounded Rationality in Macroeconomics*. Oxford: Oxford University Press.

Seber, G. A. F. (2008). *A Matrix Handbook for Statisticians*. Hoboken: Wiley.

Silva, J. L. (2015). Remarks on Yu Miao and Shoufang Xu's paper 'Almost sure convergence of weighted sums'. *Miskolc Mathematical Notes 16*, 503–506.

Stock, J. H. (1987). Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica 55*, 1035–1056.

Stoica, G. (1987). A note on the rate of convergence in the strong law of large numbers for martingales. *Journal of Mathematical Analysis and Applications 381*, 910–913.

Temme, N. M. (2008). *Asymptotic Methods for Integrals*. Singapore: World Scientific Publishing.

Thompson, I. (2013). Algorithm 026: incomplete gamma functions with negative arguments. *ACM Transactions on Mathematical Software 39*, 1–9.

Wickens, M. R. and T. S. Breusch (1988). Dynamic specification, the long-run and the estimation of transformed regression models. *The Economic Journal 98*, 198–205.

Winitzki, S. (2003). Computing the incomplete gamma function to arbitrary precision. In G. Goos, J. Hartmanis and J. van Leeuwen (Eds.), *Proceedings of the International Conference on Computational Science and Its Applications*, 790–798. New York: Springer.

General remarks: The appendix consists of three main parts: appendix A contains six auxiliary lemmata; appendix B contains the proofs of the main results; appendix S contains supplementary material. Throughout, let $\kappa_a^{(\ell)}$, $\kappa_\varepsilon^{(\ell)}$ and $\kappa_x^{(\ell)}$ denote the ℓ^{th} moment of $a_0 - \alpha$, ε_t and x_t; set $\sigma^2 \coloneqq \kappa_\varepsilon^{(2)}$. Furthermore, the auxiliary assumption that

$$0 < c\gamma_i < 1, \text{ for all } i \tag{C}$$

is made. By assumption, $|\beta| < 1$, which implies that $c\gamma > 0$ (recall that $c = 1 - \beta$). Hence, if $c\gamma < 1$, assumption C is trivially satisfied. If, on the other hand, $c\gamma > 1$, suppose that $i \geq i_0$, where $i_0 \in \mathbb{N}_1$ is such that $c\gamma_i < 1$ for all $i \geq i_0$. This is innocous as, by assumption A, $\gamma_i \to 0$ for $i \to \infty$. In order to keep the notation simple during the proofs, it is, *w.l.o.g.*, assumed that $c\gamma < 1$; i.e. $i_0 = 1$.

A Auxiliary results

Note. This section derives the working formula for a_t and collects some helpful auxiliary lemmata.

Working formula for a_t

Taking assumption (B4) into account, one gets from equation (1.4):

$$a_t^* = a_{t-1}^*(1 - c\gamma_t x_t^{*2}) + \gamma_t u_t, \tag{A.1}$$

where $a_t^* \coloneqq a_t - \alpha$ and $u_t \coloneqq x_t^* \varepsilon_t^*$, with $x_t^* \coloneqq \kappa_x^{(2)-1/2} x_t$ and $\varepsilon_t^* \coloneqq \kappa_x^{(2)-1/2} \varepsilon_t$. Let

$$f_i(x_i) \coloneqq 1 - c\gamma_i x_i^{*2}, \tag{A.2}$$

and set

$$\vartheta_{m,n} \coloneqq \prod_{i=n}^{m} f_i(x_i), \ 1 \leq n \leq m+1, \tag{A.3}$$

where it is understood that $\vartheta_m \coloneqq \vartheta_{m,1}$, $\vartheta_{m,m+1} \coloneqq 1$. Applying recurrence to (A.1), yields

$$a_t^* = \xi_t + a_0^* \vartheta_t, \text{ with } \xi_t \coloneqq \sum_{k=1}^{t} \gamma_k \vartheta_{t,k+1} u_k. \tag{A.4}$$

Remark A.1 *The properties of the weighted sum ξ_t play a crucial role for the subsequent development of an asymptotic theory. For example, in order to resort to limiting laws, $E[\xi_t^2]$ and $E[\xi_t^4]$ need to be evaluated. The main idea is to replace the double-indexed random weights $\vartheta_{m,n}$ with the product of the expected functions $f_i(x_i)$*

$$\Phi_{m,n} \coloneqq \prod_{i=n}^{m} E[f_i(x_i)] = \prod_{i=n}^{m} (1 - c\gamma_i), \ 1 \leq n \leq m+1, \tag{A.5}$$

and to show that the approximation error is negligible. As with $\vartheta_{m,n}$, the notational conventions $\Phi_m \coloneqq \Phi_{m,1}$, $\Phi_{m,m+1} \coloneqq 1$ are used.

A.1 Lemma A.1

Note. This lemma summarizes properties of partial sums of the coefficients

$$g_{k,t} \coloneqq \gamma_k \Phi_{t,k+1},\ 1 \le k \le t, \tag{A.6}$$

which will be frequently used in the subsequent proofs.

Lemma A.1 *Define*

$$\phi_t^i \coloneqq \sum_{k=1}^{t} g_{k,t},\ \phi_t^{ii} \coloneqq \sum_{k=1}^{t} g_{k,t}^2 \ \text{ and } \phi_t^{iii} \coloneqq \sum_{k=1}^{t-1} \sum_{s=1}^{k} g_{s,t} g_{s,k}.$$

Then,

$$\begin{aligned}
&(a) \qquad && \phi_t^i = 1/c + o(1),\\
&(b) \qquad && \phi_t^{ii} = \gamma_t/(2c) + o(\gamma_t),\\
&(c) \qquad && \phi_t^{iii} = 1/(2c^2) + o(1),
\end{aligned}$$

and for $\alpha > 0$ and $\beta \ge 0$

$$(d) \qquad \sum_{k=1}^{t} \Phi_{t,k+1}^{\alpha} \gamma_k^{\beta} = O(\gamma_t^{\beta-1}).$$

Remark A.2 *By the Cesàro mean convergence theorem (cf., e.g., Davidson (1994, theorem 2.26)), it follows that $\bar{\phi}^i = 1/c + o(1)$, $\bar{\phi}^{ii} = o(1)$ and $2\bar{\phi}^{iii} = 1/c^2 + o(1)$, where $\bar{\phi}^{\ell} \coloneqq T^{-1} \sum_{t=1}^{T} \phi_t^{\ell}$ for $\ell \in \{i, ii, iii\}$.*

Proof of lemma A.1 (a): Throughout, the conventions $\Phi_t \coloneqq \Phi_{t,1}$ and $\Phi_{t,t+1} \coloneqq 1$ are used. Note that

$$c\phi_t^i + \Phi_t = 1. \tag{A.7}$$

It thus remains to show that $\Phi_t = o(1)$. Using Euler summation (see, e.g., Apostol (1999, theorem 2)), the generalized harmonic number

$$\mathcal{H}_n(\eta) \coloneqq \sum_{i=1}^{n} 1/i^{\eta} \tag{A.8}$$

can be written as

$$\mathcal{H}_t(\eta) = \frac{t^{1-\eta}}{1-\eta} + \zeta(\eta) + \eta R_t^{\infty}(\eta), \text{ with } R_a^b(\eta) \coloneqq \int_a^b \frac{u - \lfloor u \rfloor}{u^{1+\eta}} du, \tag{A.9}$$

where $\zeta(\cdot)$ denotes the Riemann zeta function

$$\zeta(\eta) \coloneqq \lim_{t \to \infty} \left(\mathcal{H}_t(\eta) - \frac{t^{1-\eta}}{1-\eta} \right), \tag{A.10}$$

$\lfloor x \rfloor$ is the greatest integer smaller or equal x and

$$0 \le R_t^{\infty}(\eta) \le \int_t^{\infty} \frac{1}{u^{1+\eta}} du = \frac{1}{\eta t^{\eta}}; \tag{A.11}$$

see, e.g., Apostol (1976, theorem 3.2). For completeness, note that

$$\mathcal{H}_t(1) = O(\ln(t)) \text{ and, for } \epsilon > 0,\ \mathcal{H}_t(1+\epsilon) = O(1). \tag{A.12}$$

In order to keep the notation simple, the dependence of $\mathcal{H}_t(\eta)$ and $R_a^b(\eta)$ on η will be hereafter implicitly understood; i.e. $\mathcal{H}_t \coloneqq \mathcal{H}_t(\eta)$ and $R_a^b \coloneqq R_a^b(\eta)$. The claim is a consequence of the following: for any $r \ge 1$ and $0 \le k \le t+1$, there exists a finite constant C such that

$$\Phi_{t,k+1}^r \le C \exp\{-ar(t^b - k^b)\}, \tag{A.13}$$

where

$$a \coloneqq \bar{c}/b, \tag{A.14}$$

$$b \coloneqq 1 - \eta, \tag{A.15}$$

$$\bar{c} \coloneqq c\gamma, \text{ with } c = 1 - \beta. \tag{A.16}$$

A first-order Taylor series expansion of $\ln(1-x)$ around $x = 0$ yields

$$\ln(1 - c\gamma_j) = -c\gamma_j - (c\gamma_j/\xi_j)^2, \tag{A.17}$$

where $\xi_j \in (\sqrt{2}(1 - c\gamma_j), \sqrt{2})$ (Lagrange form of the remainder); clearly, $\xi_j > 0$ uniformly in j. Suppose first that $k = 0$. Then, the elementary inequality $\exp\{x\} \le 1$ $(x \le 0)$ gives

$$\begin{aligned}\Phi_t^r = \exp\{r \sum_{j=1}^t \ln(1 - c\gamma_j)\} &= \exp\{-cr \sum_{j=1}^t \gamma_j\} \exp\{-r \sum_{j=1}^t (c\gamma_j/\xi_j)^2\} \\ &\le \exp\{-cr \sum_{j=1}^t \gamma_j\} = \exp\{-\bar{c}r\mathcal{H}_t\}. \end{aligned} \tag{A.18}$$

But, by equations (A.9) and (A.11),

$$\exp\{-\bar{c}r\mathcal{H}_t\} = \exp\{-\bar{c}r\zeta(\eta)\} \exp\{-art^b\} \exp\{-\bar{c}r\eta R_t^{\infty}\} \le C \exp\{-art^b\}, \tag{A.19}$$

where $C \coloneqq \exp\{-\bar{c}r\zeta(\eta)\}$ (note that $\zeta(\eta) < 0$ for $\eta < 1$); thereby proving the claim. Turning to the case where $k > 0$, note that

$$\sum_{j=k+1}^{t} \gamma_j = \gamma(\mathcal{H}_t - \mathcal{H}_k) = (\gamma/b)(t^b - k^b) - \gamma\eta R_k^t, \tag{A.20}$$

using equation (A.9) and $R_t^\infty = R_k^\infty - R_k^t$. Next, by equation (A.17),

$$\begin{aligned}\Phi_{t,k+1}^r &= \exp\{-r\bar{c}(H_t - H_k)\}\exp\{-r\sum_{i=k+1}^{t}(c\gamma_i/\xi_i)^2\} \\ &\le \exp\{-r\bar{c}(H_t - H_k)\} = \exp\{-ra(t^b - k^b)\}\exp\{r\bar{c}\eta R_k^t\}.\end{aligned} \tag{A.21}$$

But

$$\exp\{r\bar{c}\eta R_k^t\} \le \exp\{r\bar{c}\eta R_k^\infty\} \le \exp\{rc\gamma_k\} \to 1, \tag{A.22}$$

as $k \to \infty$, thereby showing that $\exp\{r\bar{c}\eta R_k^t\}$ is uniformly bounded over k and t. The claim thus follows by putting C in (A.13) as $C \coloneqq \sup_{k\le t}\exp\{r\bar{c}\eta R_k^t\}$. □

Proof of lemma A.1-(b): The main idea of this proof is to approximate the partial sum

$$c^2\phi_t^{ii} = \sum_{k=1}^{t}(cg_{k,t})^2 \tag{A.23}$$

by the integral of the function

$$f(k,t) \coloneqq (c\gamma_k)^2\exp\{-2a(t^b - k^b)\}. \tag{A.24}$$

To begin with, let us rewrite (A.23) as

$$c^2\phi_t^{ii} = \int_1^t f(k,t)dk + A_t + B_t, \tag{A.25}$$

where

$$A_t \coloneqq \sum_{k=1}^{t} f(k,t) - \int_1^t f(k,t)dk \quad \text{and} \quad B_t \coloneqq \sum_{k=1}^{t}((cg_{k,t})^2 - f(k,t)). \tag{A.26}$$

The remainder of this proof can be structured in three steps:

(1) show that $\int_1^t f(k,t)dk = (c\gamma_t/2)(1 + o(1))$;

(2) show that $A_t = o(\gamma_t)$;

(3) show that $B_t = o(\gamma_t)$.

Proof of step (1): This part of the proof aims at establishing

$$\int_1^t f(k,t)dk = \frac{c\gamma_t}{2}(1 + o(1)). \tag{A.27}$$

More generally, it will be shown that that for any $s, r, z > 0$

$$\lim_{t \to \infty} t^{\eta(s-1)} \int_1^t \exp\{-(rz/b)(t^b - k^b)\}(z/k^{1-b})^s dk = \frac{z^{s-1}}{r}. \tag{A.28}$$

Assuming (A.28) is true, (A.27) follows, by hypothesis, from (A.28) with $s = r = 2$ and $z = \bar{c}$. In order to verify the result in (A.28), recall the definition of the (upper) incomplete gamma function

$$\Gamma(s,x) := \int_x^\infty t^{s-1} \exp\{-t\} dt, \tag{A.29}$$

see, e.g., Jameson (2016) for details. The 'complete' gamma function is obtained as $\Gamma(s) := \Gamma(s, 0)$. Note that the definition of $\Gamma(s, x)$ extends to arbitrary (possibly complex) values of s and x; see, e.g., Winitzki (2003) or Thompson (2013). The proof makes repeatedly use of the following lemma.

Lemma A.1*a*. *For any $x, v > 0$*

$$\lim_{t \to \infty} t^{1+v}\Gamma(-1/v, -xt^v) \exp\{-xt^v\} = -x^{(1-v)/v} \exp\{-i\pi/v\}, \tag{G-a}$$

$$\int \exp\{-xt^v\}dt = -\frac{\Gamma(1/v, xt^v)}{vx^{1/v}} + C, \tag{G-b}$$

$$\int \exp\{xt^v\}dt = -\frac{\Gamma(1/v, -xt^v)}{vx^{1/v}} \exp\{-i\pi/v\} + C, \tag{G-c}$$

with i and C denoting the imaginary number and a suitable constant, respectively.[12]

Remark. *The above implies together with equation* (A.13) *for any $r > 0$*

$$\int_1^\infty \Phi_t^r dt = O(1), . \tag{G-d}$$

Proof of lemma A.1*a*: Define $w(t) := \Gamma(-1/v, -xt^v)$ and $q(t) := t^{-1-v} \exp\{xt^v\}$ and denote the first derivatives with respect to t by

$$w'(t) := -\frac{v \exp\{xt^v - i\pi/v\}}{t^2 x^{1/v}} \quad \text{and} \quad q'(t) := t^{-v-2} \exp\{xt^v\}(v(xt^v - 1) - 1), \tag{A.30}$$

so that

$$\frac{w'(t)}{q'(t)} = -\frac{v \exp\{-i\pi/v\}}{x^{1/v}(vx - (1+v)/t^v)}. \tag{A.31}$$

Then, by L'Hôspital's rule,

$$\lim_{t \to \infty} t^{1+v}\Gamma(-1/v, -xt^v) \exp\{-xt^v\} = \lim_{t \to \infty} \frac{w'(t)}{q'(t)} = -x^{(1-v)/v} \exp\{-i\pi/v\}. \tag{A.32}$$

This completes the proof of part (G-a). Turning to part (G-b), use the u-substitution $u := xt^v$

and observe that $t^{1-v} = u^{1/v-1}x^{1-1/v}$. Hence,

$$\int \exp\{-xt^v\}dt = \frac{1}{vx^{1/v}} \int \exp\{-u\}u^{1/v-1}du$$
$$= -\frac{1}{vx^{1/v}} \int_u^\infty \exp\{-k\}k^{1/v-1}dk + C, \tag{A.33}$$

where the last equality uses that for any integrable function $f(\cdot)$

$$\int f(x)dx = \int_a^x f(t)dt + C,$$

with a so that the integral converges; thereby proving the claim. Part (G-c) follows directly from (G-b) upon using the u-substitution $u := ((-1)x)^{1/v}t$. □

Back to the proof of (A.28)**:** To begin with, note that (ignoring the constant of integration)

$$\int \exp\{-(rz/b)(t^b - k^b)\}(z/k^{1-b})^s dk = C_0 m(k,t), \tag{A.34}$$

where

$$C_0 := -b^{-1}(b/r)^{\frac{1-s\eta}{b}} z^{\frac{s-1}{b}} \exp\{-i\pi(1-s\eta)/b\} \tag{A.35}$$
$$m(k,t) := \exp\{-(rz/b)t^b\}\Gamma\big((1-s\eta)/b, -k^b rz/b\big). \tag{A.36}$$

Derivation of equation (A.34): Use the u-substitution

$$u := (rz/b)^{\frac{1-s\eta}{b}} k^{1-s\eta}, \tag{A.37}$$

so that

$$dk = (b/(rz))^{\frac{1-s\eta}{b}} \frac{k^{s\eta}}{1-s\eta} du.$$

Note that $\exp\{-(rz/b)(t^b - k^b)\} = \exp\{u^{\frac{b}{1-s\eta}}\}\exp\{-rz/bt^b\}$. Hence, (ignoring the constant of integration)

$$\int \exp\{-(rz/b)(t^b - k^b)\}(z/k^{1-b})^s dk = \frac{(b/r)^{\frac{1-s\eta}{b}} z^{\frac{s-1}{b}} \exp\{-(rz/b)t^b\}}{1-s\eta}$$
$$\times \int \exp\{u^{\frac{b}{1-s\eta}}\}du. \tag{A.38}$$

Now, (G-c) implies that with $x = 1$, $v = b/(1-s\eta)$ and $t = u$

$$\int \exp\{u^{\frac{b}{1-s\eta}}\}du = -\frac{1-s\eta}{b} \exp\{-\frac{i\pi(1-s\eta)}{b}\}\Gamma((1-s\eta)/b, -u^{\frac{b}{1-s\eta}}). \tag{A.39}$$

Equation (A.34) follows upon undoing the substitution and collecting terms.

Since, $m(1,t) = O(\exp\{-(rz/b)t^b\})$,

$$\int_1^t \exp\{-rz/b\left(t^b - k^b\right)\}(z/k^{1-b})^s dk = C_0 m(t,t) + O(\exp\{-\lambda t^b\}), \tag{A.40}$$

for some $\lambda > 0$. Next, it will be shown that

$$\lim_{t\to\infty} t^{\eta(s-1)} m(t,t) = \frac{C_1}{rz}, \text{ with } C_1 := -b^{\eta(s-1)/b}(rz)^{(1-s\eta)/b} \exp\{i\pi(1-s\eta)/b\}. \tag{A.41}$$

Because $m(t,t) = \exp\{-(rz/b)t^b\}\Gamma((1-s\eta)/b, -t^b rz/b)$, by L'Hôspital's rule,

$$\lim_{t\to\infty} t^{\eta(s-1)} m(t,t) = \lim_{t\to\infty} \frac{h_1'(t)}{h_2'(t)}, \tag{A.42}$$

where $h_1(t) := \Gamma((1-s\eta)/b, -t^b rz/b)$ and $h_2(t) := \exp\{t^b rz/b\}t^{-\eta(s-1)}$; the corresponding partial derivatives with respect to t are respectively given by

$$h_1'(t) := C_1 t^{-s\eta} \exp\{t^b rz/b\} \text{ and } h_2'(t) := rzt^{-s\eta} \exp\{t^b rz/b\}(1 - \frac{\eta(s-1)}{rzt^b}). \tag{A.43}$$

This verifies equation (A.41). Some simple calculations reveal further that $C_0 C_1/(rz) = z^{s-1}/r$. □

Proof of step (2): Note that one version of the Euler-Maclaurin formula (see, e.g., Lampret (2001)) states that

$$\begin{aligned}\sum_{k=1}^{t} f(k,t) - \int_1^t f(k,t)dk = &\frac{1}{2}(f(1,t) + f(t,t)) \\ &+ \frac{1}{12}(f^{(1)}(t,t) - f^{(1)}(1,t)) + \rho(t;f),\end{aligned} \tag{A.44}$$

where $f^{(\ell)}(k,t) := \partial^\ell/\partial k^\ell f(k,t)$ denotes the ℓ^{th} partial derivative of $f(k,t)$ with respect to k and

$$|\rho(t;f)| \le \frac{1}{120}\int_1^t |f^{(3)}(k,t)|dk; \tag{A.45}$$

cf. Lampret (2001, p. 109). It will be shown that the three terms on the right-hand side of (A.44) are of order $O(\gamma_t^2)$. Clearly, $f(1,t) + f(t,t) = O(\gamma_t^2)$. From

$$f^{(1)}(k,t) = 2(ck^b - \eta)k^{-1} f(k,t), \tag{A.46}$$

one gets $f^{(1)}(t,t) - f^{(1)}(1,t) = o(\gamma_t^2)$. Furthermore,

$$\begin{aligned}f^{(3)}(k,t) = 2\Big(&4(ck^b)^3 - 18(ck^b)^2(1-b) - ck^b(1-b)(19b-26) \\ &- 2(2-b)(1-b)(3-2b)\Big)\frac{f(k,t)}{k^3}\end{aligned} \tag{A.47}$$

implies the existence of a finite constant $C_0 > 0$ such that $|f^{(3)}(k,t)| \le C_0\,(c\gamma_k)^3 f(k,t)$. Hence, by equation (A.28), $|\rho(t;f)| = O(\gamma_t^4)$ – thereby showing that $A_t = O\left(\gamma_t^2\right)$. □

Proof of step (3): Using equations (A.9), (A.17) and (A.20), it follows that

$$
\begin{aligned}
(cg_{k,t})^2 &= (c\gamma_k)^2 \exp\{2 \sum_{j=k+1}^{t} \ln(1 - c\gamma_j)\} \\
&= (c\gamma_k)^2 \exp\{-2\bar{c}(\mathcal{H}_t - \mathcal{H}_k)\} \exp\{-2 \sum_{j=k+1}^{t} (c\gamma_j/\xi_j)^2\} \\
&= (c\gamma_k)^2 \exp\{-2\bar{c}(\mathcal{H}_t - \mathcal{H}_k)\} + (c\gamma_k)^2 \exp\{-2\bar{c}(\mathcal{H}_t - \mathcal{H}_k)\}(\exp\{-2 \sum_{j=k+1}^{t} (c\gamma_j/\xi_j)^2\} - 1) \\
&= f(k,t) + f(k,t)(\exp\{2\bar{c}\eta R_k^t\} - 1) + f(k,t)(\exp\{-2 \sum_{j=k+1}^{t} (c\gamma_j/\xi_j)^2\} - 1) \\
&\qquad + f(k,t)(\exp\{2\bar{c}\eta R_k^t\} - 1)(\exp\{-2 \sum_{j=k+1}^{t} (c\gamma_j/\xi_j)^2\} - 1) \\
&=: f(k,t) + B^i(k,t) + B^{ii}(k,t) + B^{iii}(k,t), \qquad \text{(A.48)}
\end{aligned}
$$

say. It thus suffices to show that $\sum_k B^\ell(k,t) = O(\gamma_t^a)$ $(a > 1)$ for $\ell \in \{i, ii, iii\}$. Begin with $B^i(k,t)$. Since $f(k,t)$ is non-negative, monotonically decreasing in $k \leq t$ and $\exp\{2\bar{c}\eta R_k^t\} \geq 1$, it follows that $B^i(k,t) \geq 0$. Infer from the discussion surrounding (A.9) that $\eta R_k^t \leq k^{-\eta}$ and thus $B^i(k,t) \leq f(k,t)(\exp\{2c\gamma_k\} - 1)$. Next, suppose *w.l.o.g.* (as $\gamma_k \to 0$) that $2c\gamma_k \leq 1$. The inequality $|\exp\{x\} - 1| \leq (7/4)|x|$ $(|x| \leq 1)$ yields $B^i(k,t) \leq (7/4) f(k,t) c\gamma_k$. By (A.28) and the integral comparison test, $\sum_k B^i(k,t) = O(\gamma_t^2)$. Next, consider $B^{ii}(k,t)$, which is non-positive so that

$$
\left|B^{ii}(k,t)\right| = f(k,t)(1 - \exp\{-2 \sum_{j=k+1}^{t} (c\gamma_j/\xi_j)^2\}). \qquad \text{(A.49)}
$$

The definition of the remainder ξ_j (cf. equation (A.17)) implies

$$
\sum_{j=k+1}^{t} (c\gamma_j/\xi_j)^2 \leq C \sum_{j=k+1}^{t} \gamma_j^2 \leq C \sum_{j=k}^{\infty} \gamma_j^2, \qquad \text{(A.50)}
$$

with $C := (1/2)(c/(1-\bar{c}))^2$. But, by theorem 3.2 (*c*) in Apostol (1976),

$$
\sum_{i=k}^{\infty} i^{-2\eta} = O(k^{1-2\eta}) \;\Leftrightarrow\; \sum_{i=k}^{\infty} \gamma_i^2 = O(\gamma_k^{2-1/\eta}); \qquad \text{(A.51)}
$$

where $O(\gamma_k^{2-1/\eta})$ denotes a positive quantity. Hence, putting the above together yields

$$
1 - \exp\{-2 \sum_{j=k+1}^{t} (c\gamma_j/\xi_j)^2\} \leq 1 - \exp\{-O(\gamma_k^{2-1/\eta})\} \leq O(\gamma_k^{2-1/\eta}), \qquad \text{(A.52)}
$$

using that $1-\exp\{-x\} < x$ $(x > -1)$. Therefore, by (A.28) and the integral comparison test

$$\sum_{k=1}^{t} |B^{ii}(k,t)| = \sum_{k=1}^{t} f(k,t) O(\gamma_k^{2-1/\eta}) = O(\gamma_t^{3-1/\eta}) = o(\gamma_t), \tag{A.53}$$

where the final equality uses that, by assumption A, $3-1/\eta \in (1,2)$. Finally, $\sum_k B^{iii}(k,t) = O(\gamma_t^2)$ is deduced from the analysis of $\sum_k B^{ii}(k,t)$ as

$$\left| 1 - \exp\{-2 \sum_{j=k+1}^{t} (c\gamma_j/\xi_j)^2\} \right| \leq 1 \Rightarrow |B^{iii}(k,t)| \leq O(|B^{ii}(k,t)|). \tag{A.54}$$

This completes the proof of lemma A.1-(b). □

Proof of lemma A.1 (c): For $1 \leq s \leq k \leq t$,

$$g_{s,t} g_{s,k} = g_{s,k}^2 \Phi_{t,k+1}, \tag{A.55}$$

so that

$$\begin{aligned} \phi_t^{iii} &= \sum_{k=1}^{t-1} \sum_{s=1}^{k} g_{s,t} g_{s,k} \\ &= \sum_{k=1}^{t-1} \Phi_{t,k+1} \phi_k^{ii} = 1/(2c) \sum_{k=1}^{t-1} g_{k,t}(1+o(1)) \to 1/(2c^2), \end{aligned} \tag{A.56}$$

by part (b). □

Proof of lemma A.1 (d): With $r=\alpha$, $s=\beta$ and $z=\bar{c}$ it follows from (A.28) that

$$\lim_{t\to\infty} \gamma_t^{1-\beta} \int_1^t \exp\{-\alpha a(t^b - k^b)\} \gamma_k^{\beta} dk = (c\alpha)^{-1}. \tag{A.57}$$

□

A.2 Lemma A.2, lemma A.3

Note. Lemma A.2 and A.3 are central for the asymptotic analysis under strongly mixing regressors (part (2) of assumption (B2)). The following two mixing inequalities will frequently be used: If part (2) of assumption (B2) holds true, then

$$|\mathrm{cov}[x_1^2, x_k^2]| \leq 6 \, \|x_1\|_\infty^4 \, \alpha(k), \tag{mix-i}$$

$$\left\| E[x_{t+k}^2 \mid \mathcal{F}_t] - \kappa_x^{(2)} \right\|_p \leq 2(2^{1/p}+1) \, \|x_1\|_\infty^2 \, \alpha(k)^{1/p}, \tag{mix-ii}$$

for $p > 0$. For a proof of (mix-i) and (mix-ii) see, for example, theorems 14.1, 14.2 and 14.3 in Davidson (1994).

A.2.1 Lemma A.2

Lemma A.2 *Assume that part* (2) *of assumption* (B2) *holds true and define* $V_{t,k+1} \coloneqq \Phi_{t,k+1}^{-1}\vartheta_{t,k+1}$, *for* $0 \le k \le t+1$. *Then,*

$$E[|V_{t,k+1} - 1|] = O(\gamma_k^{1-1/(2\eta)}). \tag{a}$$

Furthermore, there exists a constant $\lambda > 0$ *such that*

$$\vartheta_t \le \exp\{-\lambda t^b\} \ a.s.. \tag{b}$$

Remark A.3 *It follows immediately from the proof of lemma A.2-*(a)*, that also*

$$E[|V_{t,k+1}^n - 1|] = O(\gamma_k^{1-1/(2\eta)}) \ (n > 0) \ , \tag{a-i}$$
$$E[|V_{t,k+1}V_{t,s+1} - 1|] = O(\gamma_k^{1-1/(2\eta)}) \ (k \le s \le t). \tag{a-ii}$$

Proof of lemma A.2-(a)**:** Define the random variable $X_k \coloneqq \mu_k(x_k^2 - \kappa_x^{(2)})$, where

$$\mu_k \coloneqq \frac{c\gamma_k}{\kappa_x^{(2)}(1 - c\gamma_k)}. \tag{A.58}$$

Because $\mu_k \to 0$ and $\|x_1\|_\infty < \infty$, there exists a $k_0 \in \mathbb{N}_1$ such that $|X_k| < 1$ *a.s.* for all $k \ge k_0$. Assume, *w.l.o.g.*, that $k_0 = 1$. Then,

$$V_{t,k+1} = \prod_{i=k+1}^{t} \frac{1 - c\gamma_i x_i^{*2}}{1 - c\gamma_i} = \prod_{i=k+1}^{t} \frac{(1 - c\gamma_i) - c\gamma_i(x_i^{*2} - 1)}{1 - c\gamma_i}$$
$$= \prod_{i=k+1}^{t} (1 - X_i) = \exp\{\sum_{i=k+1}^{t} \ln(1 - X_i)\}. \tag{A.59}$$

In order to further manipulate (A.59), the following will be shown:

(1) The adapted sequence $(\{X_k, \mathcal{F}_k\},\ k \ge 1)$ is as a mixingale of size $-1/2$ so that

$$S_t \xrightarrow{a.s.} S_\infty, \tag{i}$$

where $S_t \coloneqq \sum_{k=1}^{t} X_k$, and $|S_\infty| < \infty$ *a.s.*.

(2) Define the tail-sequence $T_k \coloneqq S_\infty - S_k$. Then,

$$E[\sup_{j \ge k} |T_j|] = O(\gamma_k^{1-1/(2\eta)}). \tag{ii}$$

***Proof of* (1):** According to McLeish (1975), the adapted sequence $(\{X_k, \mathcal{F}_k\},\ k \ge 1)$ is a mixingale of size $-1/2$ if, for sequences of non-negative constants c_n and $\psi_m = O(m^{-(1/2+\varrho)})$ ($\varrho > 0$), one has for all $n \ge 1$ and $m \ge 0$,

$$\|E[X_n \mid \mathcal{F}_{n-m}]\|_2 \le c_n\psi_m. \tag{A.60}$$

Since $\mu_n \leq (1-c\gamma)^{-1}\kappa_x^{(2)-1}\gamma_n$, one has, by property (mix-ii),

$$\|E[X_n \mid \mathcal{F}_{n-m}]\|_2 \leq c_n\psi_m, \text{ with } c_n := C\gamma_n,$$
$$C := (1-c\gamma)^{-1}\kappa_x^{(2)-1}(8^{1/2}+2)\,\|x_1\|_\infty^2\,,$$
$$\psi_m := \alpha(m)^{1/2}. \tag{A.61}$$

By assumption B-(ii), $\alpha(m)^{1/2} = O(m^{-(1+\varrho)/2})$ ($\varrho > 0$), and $(1+\varrho)/2 > 1/2$. Since $\sum_{i=1}^{\infty}\gamma_i^2 < \infty$, corollary 1.8 in McLeish (1975) reveals that the partial sum S_t converges $a.s.$ to a finite limit, say, S_∞, which proves equation (i). ***Proof of* (2):** Turning to equation (ii), it suffices, by Lyapunov's inequality, to verify that

$$\gamma_k^{1/\eta-2}E[\sup_{j\geq k}|T_j|^2] = O(1). \tag{A.62}$$

From the proof of theorem 1-(i) in Rosalsky and Rosenblatt (1998, pp. 557-558) one deduces that

$$E[\sup_{j\geq k}|T_j|^2] \leq C\lim_{n\to\infty}E[\max_{k\leq i\leq n}(\sum_{j=k}^{i}X_j)^2]. \tag{A.63}$$

On the other hand, theorem 1.6 in McLeish (1975) in conjunction with lemma A.2-(i) ensures the existence of a finite constant $C > 0$ such that

$$E[\max_{k\leq i\leq n}(\sum_{j=k}^{i}X_j)^2] \leq C\sum_{j=k}^{n}\gamma_j^2. \tag{A.64}$$

By theorem 3.2 (c) in Apostol (1976),

$$\sum_{i=k}^{\infty} i^{-2\eta} = O(k^{1-2\eta}) \;\Leftrightarrow\; \sum_{i=k}^{\infty}\gamma_i^2 = O(\gamma_k^{2-1/\eta}). \tag{A.65}$$

Hence, putting (A.63), (A.64) and (A.65) together yields for some constant $C > 0$

$$\gamma_k^{1/\eta-2}E[\sup_{j\geq k}|T_j|^2] \leq C\gamma_k^{1/\eta-2}\sum_{i=k}^{\infty}\gamma_i^2 = O(1), \tag{A.66}$$

thereby proving equation (ii).

Back to equation (A.59)**:** A first-order Taylor series expansion of $\ln(1-x)$ around zero yields

$$\ln(1-X_i) = -X_i - (X_i/\xi_i)^2, \tag{A.67}$$

where ξ_i lies on the line segment connecting $\sqrt{2}(1-X_i)$ and $\sqrt{2}$ (Lagrange form of the remainder). Since $|X_i| < 1$ $a.s.$, there exists some C such that $\inf_i|\xi_i| \geq C > 0$ $a.s.$; i.e. (A.67) is well-defined. Because $\sum_{i=k+1}^{t}X_i = S_t - S_k = (S_\infty - S_k) - (S_\infty - S_t) = T_k - T_t$, it follows from the above that $\sum_{i=k+1}^{t} \ln(1 - X_i) = T_k - T_t + n_{k,t}$; or, equivalently,

$$V_{t,k+1} = \exp\{T_k - T_t + n_{k,t}\}, \text{ with } n_{k,t} \coloneqq \sum_{i=k+1}^{t} (X_i/\xi_i)^2. \tag{A.68}$$

Set $M_{k,t} \coloneqq \exp\{|T_t - T_k + n_{k,t}|\}$. The inequality $|\exp\{x\} - 1| \leq |x| \exp\{|x|\}$ yields

$$\begin{aligned}|V_{t,k+1} - 1| = |\exp\{T_t - T_k + n_{k,t}\} - 1| &\leq |T_t - T_k + n_{k,t}| \exp\{|T_t - T_k + n_{k,t}|\} \\ &\leq (\sup_{j \geq k} |T_j| + |n_{k,t}|) M_{k,t}. \end{aligned} \tag{A.69}$$

It follows from Hölder's inequality that

$$\begin{aligned} E[|n_{k,t}|] &\leq \sum_{i=k+1}^{\infty} E[(X_i/\xi_i)^2] \leq C^{-2} \sum_{i=k+1}^{\infty} E[X_i^2] \\ &= C^{-2} \sum_{i=k+1}^{\infty} O(\gamma_i^2) = O(\gamma_k^{2-1/\eta}), \end{aligned} \tag{A.70}$$

where C is such that $\inf_i |\xi_i| \geq C > 0$ *a.s.*. Because, $E[|n_{k,t}|] = o(E[\sup_{j \geq k} |T_j|])$ and $M_{k,t}$ is almost surely bounded in k and t, it follows from Hölder's inequality that

$$E[|V_{t,k+1} - 1|] \leq CE[\sup_{j \geq k} |T_j|](1 + o(1)) = O(\gamma_k^{1-1/2\eta}). \tag{A.71}$$

This proves part (a) of this lemma. □

Proof of lemma A.2-(b): Since, by assumption (B2), $(x_k,\, k \geq 1)$ is bounded almost surely and $\gamma_k \to 0$, there exists some (random) number $k_0 \in \mathbb{N}_1$ so that $\gamma_k x_k^{*2} < 1/c$ *a.s.* and $\ln(1 - c\gamma_k x_k^{*2})$ is well defined for all $k \geq k_0$ (recall that $x_k^* = \kappa_x^{(2)-1/2} x_k$). For simplicity, assume $k_0 = 1$. (Otherwise, apply the following to ϑ_{t,k_0}, where $\vartheta_t = \vartheta_{k_0-1}\vartheta_{t,k_0}$.) Hence, a first order Taylor-series expansion of $\ln(1 - x)$ around zero yields

$$\ln(1 - c\gamma_k x_k^{*2}) = -c\gamma_k x_k^{*2} - (c\gamma_k)^2 (x_k^{*2}/\xi_k)^2, \tag{A.72}$$

where ξ_k lies on the line segment connecting $\sqrt{2}(1 - c\gamma_k x_k^{*2})$ and $\sqrt{2}$ (Lagrange form of the remainder). Since $c\gamma_k x_k^{*2} < 1$ *a.s.*, there exists some constant C such that $\inf_k |\xi_k| \geq C > 0$

$a.s.$, i.e. (A.72) is well-defined. Taking equation (A.72) into account, one obtains

$$\begin{aligned} t^{-b}\sum_{k=1}^{t}\ln(1-c\gamma_k x_k^{*2}) &= -\,t^{-b}\sum_{k=1}^{t}c\gamma_k x_k^{*2} - t^{-b}\sum_{k=1}^{t}(c\gamma_k)^2(x_k^{*2}/\xi_k)^2 \\ &= -\,t^{-b}\sum_{k=1}^{t}c\gamma_k \\ &\quad -\,t^{-b}\sum_{k=1}^{t}c\gamma_k(x_k^{*2}-1) \\ &\quad -\,t^{-b}\sum_{k=1}^{t}(c\gamma_k)^2(x_k^{*2}/\xi_k)^2. \end{aligned} \tag{A.73}$$

By equation (A.9), the first term on the right-hand side of (A.73) obeys,

$$t^{-b}\sum_{k=1}^{t}c\gamma_k = c\gamma t^{-b}\mathcal{H}_t(\eta) \longrightarrow c\gamma/b =: a. \tag{A.74}$$

The summands of the second term in (A.73) are readily recognized as $(1-c\gamma_k)X_k$, where, as shown in lemma A.2-(i), $S_t=\sum_{k=1}^{t}X_k<\infty$ $a.s.$ as $t\to\infty$. Hence,

$$t^{-b}\sum_{k=1}^{t}c\gamma_k(x_k^{*2}-1) \simeq t^{-b}S_t = O_{a.s}(t^{-b}). \tag{A.75}$$

Finally, turn to the third expression on the right-hand side of (A.73) and note that

$$t^{-b}\sum_{k=1}^{t}(c\gamma_k)^2(x_k^{*2}/\xi_k)^2 = O_{a.s.}(t^{-b}). \tag{A.76}$$

To see this, note that, by Hölder's inequality,

$$\sum_{k=1}^{\infty}(c\gamma_k)^2E[(x_k^{*2}/\xi_k)^2] \le (c\left\|x_1\right\|_\infty^2/C\kappa_x^{(2)})^2\sum_{k=1}^{\infty}\gamma_k^2<\infty, \tag{A.77}$$

where C is such that $\inf_t|\xi_t|\ge C>0$, and it has been used that, by assumption A, $\sum_{k\ge 1}\gamma_k^2<\infty$. Hence, a (sufficient) condition for $\sum_{k\ge 1}(c\gamma_k)^2(x_k^{*2}/\xi_k)^2$ to converge $a.s.$ is satisfied. Putting thus the preceding steps together yields

$$t^{-b}\sum_{k=1}^{t}\ln(1-c\gamma_k x_k^{*2}) = -c\gamma t^{-b}\mathcal{H}_t(\eta)+o_{a.s}(1) \longrightarrow -a<0\ \ a.s., \tag{A.78}$$

or, equivalently,

$$\vartheta_t^{1/t^b} = \exp\{t^{-b}\sum_{k=1}^{t}\ln(1-c\gamma_k x_k^{*2})\} \longrightarrow \exp\{-a\}\ \ a.s., \tag{A.79}$$

with ϑ_t defined in (A.3). This, in turn, ensures the existence of a (random) number $t_0\in\mathbb{N}_1$

so that $\vartheta_t \le \delta^{t^b}$ *a.s.* for some $\delta \in (\exp\{-a\}, 1)$ and all $t \ge t_0$. This proves lemma A.2-(b). □

A.2.2 Lemma A.3

Note. One major purpose of the following lemma is to verify sufficient conditions of a weak LLN applied to $T^{-b}\sum_{t=1}^{T} a_t^{*2}$. In particular, this requires the evaluation of the covariances $\text{cov}[a_t^*, a_{t+m}^*]$ ($m \ge 0$); see lemma A.5 below.

Lemma A.3 *Suppose assumption part* (2) *of assumption* (B2) *holds true and set* $w_t \coloneqq x_t^2 - \kappa_x^{(2)}$. *Then, for* $m \ge 0$*:*

$$\begin{aligned} E[w_{t+m}a_{t+m-1}^{*2}] &= O(\gamma_{t+m}^{2-1/(2\eta)}) && (a)\\ E[w_t w_{t+m} a_{t-1}^{*2} a_{t+m-1}^{*2}] &= O(\gamma_t^{2-1/(2\eta)}\gamma_{t+m}) + O(g_{t,t+m}^2(\alpha(m) + \gamma_t^{1-1/(2\eta)}))\\ &\quad + o(\gamma_t \gamma_{t+m}^{2-1/(2\eta)}) && (b)\end{aligned}$$

Remark A.4 *It is apparent from the proof of* (b) *that*

$$E[w_{t+m}a_{t-1}^{*2}a_{t+m-1}^{*2}] = O(E[w_t w_{t+m} a_{t-1}^{*2} a_{t+m-1}^{*2}]).$$

Proof of lemma A.3-(a): By equation (A.4) and assumption A-A1, for all $t \in \mathbb{N}_1$

$$E[w_t a_{t-1}^{*2}] = E[a_0^* w_t \vartheta_{t-1}^2] + E[w_t \xi_{t-1}^2]. \tag{A.80}$$

By lemma A.2-(b) and Hölder's inequality, $|\kappa_a^{(1)}|(\|x_1\|_\infty^2 + \kappa_x^{(2)})\exp\{-\lambda t^b\}$ for all $t \ge t_0$. Turning to the second summand in (A.80), recall for $n \le m+1$ from equation (A.6) and lemma A.2 the notations $g_{n,m} = \gamma_n \Phi_{m,n+1}$ and $V_{m,n} = \vartheta_{m,n}/\Phi_{m,n}$, respectively. Hence, $E[\xi_{t-1}^2(x_t^2 - \kappa_x^{(2)})]$ can be decomposed as

$$\begin{aligned} E[w_t \xi_{t-1}^2] &= (\tau^2/\kappa_x^{(2)}) \sum_{k=1}^{t-1} \gamma_k^2 E[w_t x_k^2 \vartheta_{t-1,k+1}^2]\\ &= (\tau^2/\kappa_x^{(2)}) \sum_{k=1}^{t-1} g_{k,t-1}^2 E[w_t x_k^2 V_{t-1,k+1}^2] =: (\tau^2/\kappa_x^{(2)})(A_t + B_t),\end{aligned}$$

say, where

$$A_t \coloneqq \sum_{k=1}^{t-1} g_{k,t-1}^2 \text{cov}[x_1^2, x_{t-k}^2] \tag{A.81}$$

$$B_t \coloneqq \sum_{k=1}^{t-1} g_{k,t-1}^2 E[w_t x_k^2 (V_{t-1,k+1}^2 - 1)]. \tag{A.82}$$

By Hölder's inequality, property (mix-ii) and lemma A.1-(d), one gets

$$A_t \le 6\,\|x_1\|_\infty^4\, C(\sum_{k=1}^{t-1} g_{k,t-1}^4)^{1/2} = O(\gamma_t^{3/2}), \tag{A.83}$$

where $C^2 \coloneqq \sum_{k=1}^{\infty} \alpha(k)^2$ is finite by condition (ii) of assumption (B2). Since $\eta > 1/2$,

$A_t = o(\gamma_t^{2-1/(2\eta)})$. Next, Hölder's inequality yields

$$B_t \le \|x_1\|_\infty^2 (\|x_1\|_\infty^2 + \kappa_x^{(2)}) \sum_{k=1}^{t-1} g_{k,t-1}^2 E[|V_{t-1,k+1}^2 - 1|]. \tag{A.84}$$

Now, by lemma A.2-(a), $E[|V_{t,k+1}^2 - 1|] = O(\gamma_k^{1-1/2\eta})$, so that, by lemma A.1-(d), $B_t = O(\gamma_t^{2-1/2\eta})$. Hence, $E[(x_t^2 - \kappa_x^{(2)}) a_{t-1}^{*2}] = O(\gamma_t^{2-1/(2\eta)})$. □

Proof of lemma A.3-(b): By equation (A.4),

$$a_t^{*2} a_{t+m}^{*2} = ((a_0^* \vartheta_t)^2 + \xi_t^2 + 2a_0^* \vartheta_t \xi_t)((a_0^* \vartheta_{t+m})^2 + \xi_{t+m}^2 + 2a_0^* \vartheta_{t+m} \xi_{t+m}). \tag{A.85}$$

Hence, by assumption B-A1,

$$\begin{aligned} E[w_{t+1} w_{t+m+1} a_t^{*2} a_{t+m}^{*2}] &= \kappa_a^{(4)} E[w_{t+1} w_{t+m+1} \vartheta_t^2 \vartheta_{t+m}^2] + \kappa_a^{(2)} E[w_{t+1} w_{t+m+1} \vartheta_t^2 \xi_{t+m}^2] \\ &\quad + \kappa_a^{(2)} E[w_{t+1} w_{t+m+1} \vartheta_{t+m}^2 \xi_t^2] + 2\kappa_a^{(1)} E[w_{t+1} w_{t+m+1} \vartheta_{t+m} \xi_{t+m} \xi_t^2] \\ &\quad + 2\kappa_a^{(1)} E[w_{t+1} w_{t+m+1} \xi_{t+m}^2 \vartheta_t \xi_t] + 4\kappa_a^{(2)} E[w_{t+1} w_{t+m+1} \vartheta_{t+m} \vartheta_t \xi_t \xi_{t+m}] \\ &\quad + E[w_{t+1} w_{t+m+1} \xi_{t+m}^2 \xi_t^2] \\ &=: \kappa_a^{(4)} A_{t,m} + \kappa_a^{(2)} B_{t,m} + \kappa_a^{(2)} C_{t,m} + 2\kappa_a^{(1)} D_{t,m} + 2\kappa_a^{(1)} E_{t,m} + 4\kappa_a^{(2)} F_{t,m} \\ &\quad + G_{t,m}. \end{aligned} \tag{A.86}$$

Set $C := (\|x_1\|_\infty^2 + \kappa_x^{(2)})^2$ and note that lemma A.2-(b) ensures the existence of constants $(\lambda_i,\ i = 1, \ldots, 7)$ such that

$$\begin{aligned} A_{t,m} &\le C\, \|\vartheta_t\|_4^2 \, \|\vartheta_{t+m}\|_4^2 = O(\exp\{-\lambda_0 t^b\} \exp\{-\lambda_1 (t+m)^b\}) \\ B_{t,m} &\le C\, \|\vartheta_t\|_4^2 \, \|\xi_{t+m}\|_4^2 = O(\exp\{-\lambda_2 t^b\} \gamma_{t+m}) \\ C_{t,m} &\le C\, \|\vartheta_{t+m}\|_4^2 \, \|\xi_t\|_4^2 = O(\exp\{-\lambda_3 (t+m)^b\} \gamma_t) \\ D_{t,m} &\le C\, \|\vartheta_{t+m}\|_4 \, \|\xi_{t+m}\|_4 \, \|\xi_t\|_4^2 = O(\exp\{-\lambda_4 (t+m)^b\} \gamma_{t+m}^{1/2} \gamma_t) \\ E_{t,m} &\le C\, \|\vartheta_t\|_4 \, \|\xi_t\|_4 \, \|\xi_{t+m}\|_4^2 = O(\exp\{-\lambda_5 t^b\} \gamma_t^{1/2} \gamma_{t+m}) \\ F_{t,m} &\le C\, \|\xi_t\|_4 \, \|\xi_{t+m}\|_4 \, \|\vartheta_t\|_4 \, \|\vartheta_{t+m}\|_4 = O(\exp\{-\lambda_6 t^b\} \exp\{-\lambda_7 (t+m)^b\} \gamma_t^{1/2} \gamma_{t+m}^{1/2}), \end{aligned}$$

where the first inequalities are due to Cauchy-Schwarz's and Hölder's inequality while the final approximations (involving orders of magnitude) are due to lemma A.4. Hence, the rates of decay of $A_{t,m}, B_{t,m}, \ldots, F_{t,m}$ are faster than the rates listed on the right-hand side of lemma A.2-(b). Furthermore, recall from (A.4) that $\xi_t := \sum_{k=1}^t \gamma_k \vartheta_{t,k+1} u_k$. Hence,

$$G_{t,m} = \kappa_x^{(2)-4} \sum_{i,j=1}^{t} \sum_{k,s=1}^{t+m} E[w_{t+1} w_{t+m+1} (x_i \varepsilon_i \vartheta_{t,i+1})(x_j \varepsilon_j \vartheta_{t,j+1})(x_k \varepsilon_k \vartheta_{t+m,k+1})(x_s \varepsilon_s \vartheta_{t+m,s+1})].$$

Using arguments similar to those which lead to (A.114) reveals

$$G_{t,m} = (\kappa_\varepsilon^{(4)} / \kappa_x^{(2)4}) H_{t,m} + (\sigma / \kappa_x^{(2)})^4 (6 I_{t,m} + J_{t,m}), \tag{A.87}$$

with

$$H_{t,m} := \sum_{k=1}^{t} \gamma_k^4 E[w_{t+1} w_{t+m+1} x_k^4 \vartheta_{t,k+1}^2 \vartheta_{t+m,k+1}^2]$$

$$I_{t,m} := \sum_{k=2}^{t} \sum_{s=1}^{k-1} \gamma_k^2 \gamma_s^2 E[w_{t+1} w_{t+m+1} x_k^2 x_s^2 \vartheta_{t,k+1}^2 \vartheta_{t+m,s+1}^2]$$

$$J_{t,m} := \sum_{k=1}^{t} \sum_{s=t+1}^{t+m} \gamma_k^2 \gamma_s^2 E[w_{t+1} w_{t+m+1} x_k^2 x_s^2 \vartheta_{t,k+1}^2 \vartheta_{t+m,s+1}^2]. \tag{A.88}$$

It will be shown that

$$H_{t,m} = O(g_{t,t+m}^2(\gamma_t \alpha(m) + \gamma_t^{2-1/2\eta})) \tag{A.89}$$

$$I_{t,m} = O(g_{t,t+m}^2(\alpha(m) + \gamma_t^{1-1/2\eta})) \tag{A.90}$$

$$J_{t,m} = O(\gamma_t^{2-1/(2\eta)} \gamma_{t+m}) + o(\gamma_t \gamma_{t+m}^{2-1/(2\eta)}) \tag{A.91}$$

Proof of (A.89)**:** Since $\Phi_{t,k+1}^2 \Phi_{t+m,k+1}^2 = \Phi_{t,k+1}^4 \Phi_{t+m,t+1}^2$, $A_{t,m}$ can be rewritten as

$$H_{t,m} = \Phi_{t+m,t+1}^2 (H_{t,m}^{(1)} + H_{t,m}^{(2)}), \tag{A.92}$$

with

$$H_{t,m}^{(1)} := \sum_{k=1}^{t} g_{k,t}^4 E[w_{t+1} w_{t+m+1} x_k^4] \tag{A.93}$$

$$H_{t,m}^{(2)} := \sum_{k=1}^{t} g_{k,t}^4 E[w_{t+1} w_{t+m+1} x_k^4 (V_{t,k+1}^2 V_{t+m,k+1}^2 - 1)]. \tag{A.94}$$

Recall from lemma A.2 equation (A.6) that $V_{m,n} = \vartheta_{m,n}/\Phi_{m,n}$ and $g_{n,m} = \gamma_n \Phi_{m,n}$. Now, by Hölder's inequality, property (mix-ii) and lemma A.1-(d)

$$H_{t,m}^{(1)} \leq \|x_1\|_\infty^4 (\|x_1\|_\infty^2 + \kappa_x^{(2)}) E[|E[x_{t+m+1}^2 \mid \mathcal{F}_t] - \kappa_x^{(2)}|] \sum_{k=1}^{t} g_{k,t}^4 = O(\gamma_t^3 \alpha(m)), \tag{A.95}$$

while, by Hölder's inequality,

$$H_{t,m}^{(2)} \leq \|x_1\|_\infty^4 (\|x_1\|_\infty^2 + \kappa_x^{(2)})^2 \sum_{k=1}^{t} g_{k,t}^4 E[|V_{t,k+1}^2 V_{t+m,k+1}^2 - 1|]. \tag{A.96}$$

Then, by arguments similar to the those which lead to lemma A.2-(a),

$$E[|V_{t,k+1}^2 V_{t+m,k+1}^2 - 1|] = O(\gamma_k^{1-1/(2\eta)}). \tag{A.97}$$

Therefore, by lemma A.1-(d), $H_{t,m}^{(2)} = O(\gamma_t^{4-1/2\eta})$. This, in turn, implies

$$H_{t,m} = O(g_{t,t+m}^2(\gamma_t \alpha(m) + \gamma_t^{2-1/2\eta})). \tag{A.98}$$

Proof of (A.90)**:** Similar to $H_{t,m}$, $I_{t,m}$ can be rewritten as $I_{t,m} = \Phi^2_{t+m,t+1}(I^{(1)}_{t,m} + I^{(2)}_{t,m})$, where

$$I^{(1)}_{t,m} := \sum_{k=2}^{t} \sum_{s=1}^{k-1} g^2_{s,t} g^2_{k,t} E[w_{t+1} w_{t+m+1} x^2_k x^2_s] \tag{A.99}$$

$$I^{(2)}_{t,m} := \sum_{k=2}^{t} \sum_{s=1}^{k-1} g^2_{s,t} g^2_{k,t} E[w_{t+1} w_{t+m+1} x^2_k x^2_s (V^2_{t,k+1} V^2_{t+m,s+1} - 1)]. \tag{A.100}$$

Now, using, respectively, Hölder's inequality and property (mix-ii), yields

$$\begin{aligned} I^{(1)}_{t,m} &\leq \|x_1\|^4_\infty (\|x_1\|^2_\infty + \kappa^{(2)}_x) E[|E[x^2_{t+m+1} \mid \mathcal{F}_t] - \kappa^{(2)}_x|] \sum_{k=2}^{t} \sum_{s=1}^{k-1} g^2_{s,t} g^2_{k,t} \\ &\leq C\alpha(m) \sum_{k=2}^{t} \sum_{s=1}^{k-1} g^2_{s,t} g^2_{k,t}, \end{aligned} \tag{A.101}$$

with $C := \|x_1\|^6_\infty (\|x_1\|^2_\infty + \kappa^{(2)}_x)(8^{1/2} + 2)$. But, as $g_{s,t} g_{k,t} = g_{s,k} g_{k,t} \Phi_{t,k+1}$ $(s \leq k \leq t)$, one gets

$$\sum_{k=2}^{t} \sum_{s=1}^{k-1} g^2_{s,t} g^2_{k,t} = \sum_{k=2}^{t} g^2_{k,t} \Phi^2_{t,k+1} \sum_{s=1}^{k-1} g^2_{s,k} = \sum_{k=2}^{t} O(g^3_{k,t}) \Phi_{t,k+1} = O(\gamma^2_t), \tag{A.102}$$

using repeatedly lemma A.4-(d). Hence, $I^{(1)}_{t,m} = O(\gamma^2_t \alpha(m))$. Next, by Hölder's inequality,

$$I^{(2)}_{t,m} \leq \|x_1\|^4_\infty (\|x_1\|^2_\infty + \kappa^{(2)}_x)^2 \sum_{k=2}^{t} \sum_{s=1}^{k-1} g^2_{s,t} g^2_{k,t} E[|(V_{t,k+1} V_{t+m,s+1})^2 - 1|]. \tag{A.103}$$

Again, using the same line of reasoning which lead to lemma A.2-(a), one gets

$$E[|(V_{t,k+1} V_{t+m,s+1})^2 - 1|] = O(\gamma_s^{1-1/(2\eta)}), \tag{A.104}$$

for $s \leq k \leq t$. Hence, by lemma A.1-(d) and equation (A.102), it is readily seen that

$$\begin{aligned} I^{(2)}_{t,m} &= \sum_{k=2}^{t} g^2_{k,t} \Phi^2_{t,k+1} \sum_{s=1}^{k-1} g^2_{s,k} O(\gamma_s^{1-1/(2\eta)}) \\ &= \sum_{k=2}^{t} O(g^3_{k,t} \gamma_k^{1-1/(2\eta)}) \Phi_{t,k+1} = O(\gamma_t^{3-1/(2\eta)}), \end{aligned} \tag{A.105}$$

so that $I_{t,m} = O(g^2_{t,t+m}(\alpha(m) + \gamma_t^{1-1/2\eta}))$. ***Proof of*** (A.91)**:** Rewrite $J_{t,m} = J^{(1)}_{t,m} + J^{(2)}_{t,m}$,

where

$$J^{(1)}_{t,m} := \sum_{k=1}^{t} g^2_{k,t} \sum_{s=t+1}^{t+m} g^2_{s,t+m} E[w_{t+1}w_{t+m+1}x_k^2x_s^2] \tag{A.106}$$

$$J^{(2)}_{t,m} := \sum_{k=1}^{t} g^2_{k,t} \sum_{s=t+1}^{t+m} g^2_{s,t+m} E[w_{t+1}w_{t+m+1}x_k^2x_s^2(V^2_{t,k+1}V^2_{t+m,s+1} - 1)]. \tag{A.107}$$

Now, by Hölder's inequality, property (mix-ii) and lemma A.1-(d),

$$\begin{aligned} J^{(1)}_{t,m} &\leq \|x_1\|^4 (\|x_1\|^2_\infty + \kappa^{(2)}_x) \sum_{k=1}^{t} g^2_{k,t} \sum_{s=t+1}^{t+m} g^2_{s,t+m} E[|E[x^2_{t+m+1}|\mathcal{F}_s] - \kappa^{(2)}_x|] \\ &\leq C \sum_{k=1}^{t} g^2_{k,t} \sum_{s=t+1}^{t+m} g^2_{s,t+m}\alpha(t+m-s) \\ &\leq C \sum_{k=1}^{t} g^2_{k,t} (\sum_{s=1}^{t+m} g^4_{s,t+m})^{1/2} (\sum_{s=1}^{\infty} \alpha(s)^2)^{1/2} = O(\gamma_t\gamma^{3/2}_{t+m}). \end{aligned} \tag{A.108}$$

with $C := (8^{1/2}+2)(\|x_1\|^2_\infty + \kappa^{(2)}_x)\|x_1\|^6$. Since $\eta > 1/2$, $C^{(1)}_{t,m} = o(\gamma_t\gamma^{2-1/2\eta}_{t+m})$. Furthermore, by Hölder's inequality,

$$J^{(2)}_{t,m} \leq \|x_1\|^4 (\|x_1\|^2 + \kappa^{(2)}_x)^2 \sum_{k=1}^{t} g^2_{k,t} \sum_{s=t+1}^{t+m} g^2_{s,t+m} E[|V^2_{t,k+1}V^2_{t+m,s+1} - 1|]. \tag{A.109}$$

But, using again lemma A.2-(a), $E[|V^2_{t,k+1}V^2_{t+m,s+1} - 1|] = O(\gamma_k^{1-1/(2\eta)})$ ($k \leq t \leq s$), so that

$$J^{(2)}_{t,m} \leq C\,(\sum_{k=1}^{t} g^2_{k,t}\gamma_k^{1-1/(2\eta)})(\sum_{s=t+1}^{t+m} g^2_{s,t+m}) = O(\gamma_t^{2-1/(2\eta)}\gamma_{t+m}), \tag{A.110}$$

for some $C > 0$; using repeatedly lemma A.4-(d). This proves the claim. □

A.3 Lemma A.4

Note. For the sake of better clarity and readability, lemma A.4 will be proven under the assumption of strongly mixing regressors (part (2) of assumption (B2)). For the corresponding proof under the assumption that $(\{x_t - E[x_1], \mathcal{F}_t\},\ t \geq 1)$ forms a martingale difference sequence (part (1) of assumption (B2)), please refer to section S.2.2 of the supplementary material.

Lemma A.4 *There exists some* $\lambda > 0$ *such that*

$$|E[a^*_t]| = O(\exp\{-\lambda t^b\}), \tag{a}$$

where $b = 1 - \eta$. Furthermore, recall that $c = 1 - \beta$ and $\tau^2 = \sigma^2/\kappa_x^{(2)}$. Then,

$$E[a_t^{*2}] = \frac{\tau^2 \gamma_t}{2c} + o(\gamma_t) \tag{b}$$
$$E[a_t^{*4}] = O(\gamma_t^2). \tag{c}$$

Remark A.5 *Note that* (b) *implies that*

$$\lim_{T\to\infty} T^{-b} \sum_{t=1}^{T} E[a_t^{*2}] = \frac{\tau^2\gamma}{2cb}. \tag{d}$$

Furthermore, by assumption (B2)*,* (b) *and* (c) *imply $E[z_t^2] = O(\gamma_t)$ and $E[z_t^4] = O(\gamma_t^2)$, where $z_t \coloneqq x_t a_{t-1}^*$.*

Proof of part (a)**:** From equation (A.4) and assumption B, one gets

$$\begin{aligned} E[a_t] &= \alpha + E[\xi_t] + E[a_0^* \vartheta_t] \\ &= \alpha + \kappa_x^{(2)-2} \sum_{k=1}^{t} \gamma_k E[\vartheta_{t,k+1} x_k E[\varepsilon_k | \mathcal{G}_{k-1}]] + \kappa_a^{(1)} E[\vartheta_t] = \alpha + \kappa_a^{(1)} E[\vartheta_t], \end{aligned} \tag{A.111}$$

where the final equality uses that, by assumption (B1), $E[\varepsilon_k|\mathcal{G}_{k-1}] = 0$. The claim follows from lemma A.2-(b) and the almost sure boundedness of the sequence $(x_k,\, k \geq 1)$. □

Proof of part (b)**:** Note that $\text{cov}[\vartheta_t, \xi_t] = 0$, which implies

$$\begin{aligned} E[a_t^{*2}] &= E[\xi_t^2] + \kappa_a^{(2)} E[\vartheta_t^2] \\ &= \tau^2 \phi_t^i + \kappa_a^{(2)} E[\vartheta_t^2] + \tau^2 \sum_{k=1}^{t} \gamma_k^2 E[x_k^{*2}(\vartheta_{t,k+1}^2 - \Phi_{t,k+1}^2)]. \end{aligned} \tag{A.112}$$

By part (b) of lemma A.1, the first term equals $\tau^2(2^{-1}c^{-1}\gamma_t + o(\gamma_t))$. It thus remains to be shown that the remaining summands are $o(\gamma_t)$. By lemma A.2-(b) and the boundedness of $(x_k,\, k \geq 1)$, there exists some $\lambda > 0$ so that $E[\vartheta_t^2] = O(\exp\{-\lambda t\})$. The third term is, by Hölder's inequality, bounded by

$$C \sum_{k=1}^{t} \gamma_{k,t}^2 E[|V_{t,k+1}^2 - 1|] = C \sum_{k=1}^{t} \gamma_{k,t}^2 O(\gamma_k^{1-1/(2\eta)}) = O(\gamma_t^{2-1/2\eta}) = o(\gamma_t), \tag{A.113}$$

with $C \coloneqq \tau^2 \kappa_x^{(2)-1} \|x_1\|_\infty^2$; using lemma A.2-($a$), lemma A.1-($d$) and assumption A. □

Proof of part (c)**:** By Minkowski's inequality, $\|a_t^*\|_4 \leq \kappa_a^{(4)1/4} \|\vartheta_t\|_4 + \|\xi_t\|_4$. By lemma A.2-($b$) and the boundedness of $(x_k,\, k \geq 1)$, there exists some $\lambda > 0$ so that $E[\vartheta_t^4] = O(\exp\{-\lambda t\})$. Therefore, it remains to be shown that $\|\xi_t\|_4^4 = O(\gamma_t^2)$. Begin with the following observation: for positive integers $(k_i, 1 \leq i \leq 4)$, it is seen, using the martingale difference property of $(\varepsilon_k,\, k \geq 1)$ (cf. assumption (B1)), that the expectation of the quadruple $\varepsilon_{k_1}\varepsilon_{k_2}\varepsilon_{k_3}\varepsilon_{k_4}$ is nonzero if either all indices are equal, or if two of the indices are equal within pairs but unequal across pairs (3 occurences). Thus using, in addition, the strict-exogeneity of $(x_k,\, k \geq 1)$ (cf.

assumption (B1)), one gets from (A.4)

$$
\begin{aligned}
E[\xi_t^4] &= (1/\kappa_x^{(2)4}) \sum_{k_1,k_2,k_3,k_4=1}^{t} E[(\varepsilon_{k_1} x_{k_1} \vartheta_{t,k_1+1})(\varepsilon_{k_2} x_{k_2} \vartheta_{t,k_2+1})(\varepsilon_{k_3} x_{k_3} \vartheta_{t,k_3+1})(\varepsilon_{k_4} x_{k_4} \vartheta_{t,k_4+1})] \\
&= (\kappa_\varepsilon^{(4)}/\kappa_x^{(2)4}) \sum_{k=1}^{t} \gamma_k^4 E[x_k^4 \vartheta_{t,k+1}^4] + (\sigma/\kappa_x^{(2)})^4 6 \sum_{k=2}^{t} \sum_{s=1}^{k-1} \gamma_s^2 \gamma_k^2 E[\vartheta_{t,k+1}^2 \vartheta_{t,s+1}^2 x_k^2 x_s^2] \\
&=: (\kappa_\varepsilon^{(4)}/\kappa_x^{(2)4}) A_t + (\sigma/\kappa_x^{(2)})^4 6 B_t, \qquad \text{(A.114)}
\end{aligned}
$$

say. Note that A_t can be rewritten as

$$
A_t = \sum_{k=1}^{t} g_{k,t}^4 E[x_k^4 (V_{t,k+1}^4 - 1)] + \kappa_x^{(4)} \sum_{k=1}^{t} g_{k,t}^4 =: A_t^{(1)} + \kappa_x^{(4)} A_t^{(2)}, \qquad \text{(A.115)}
$$

say. By part (d) of lemma A.1, $A_t^{(2)} = O(\gamma_t^3)$, while, by Hölder's inequality,

$$
A_t^{(1)} \leq \|x_1\|_\infty^4 \sum_{k=1}^{t} g_{k,t}^4 E[|V_{t,k+1}^4 - 1|]. \qquad \text{(A.116)}
$$

Since, by lemma A.2-(a), $E[|V_{t,k+1}^4 - 1|] = \gamma_k^{1-1/(2\eta)}$, it follows that $A_t^{(1)} = O(\gamma_t^{4-1/(2\eta)})$. Because $1/(2\eta) < 1$, $A_t = O(\gamma_t^3)$. Next, due to the stationarity of x_k one can rewrite B_t as

$$
\begin{aligned}
B_t = \pm \kappa_x^{(2)2} \sum_{k=2}^{t} \sum_{s=1}^{k-1} g_{k,t}^2 g_{s,t}^2 + B_t = \kappa_x^{(2)2} & \sum_{k=2}^{t} \sum_{s=1}^{k-1} g_{k,t}^2 g_{s,t}^2 \\
&+ \sum_{k=2}^{t} \sum_{s=1}^{k-1} g_{k,t}^2 g_{s,t}^2 \operatorname{cov}[x_1^2, x_{k-s}^2] \\
&+ \sum_{k=2}^{t} \sum_{s=1}^{k-1} g_{s,t}^2 g_{k,t}^2 E[(V_{t,k+1}^2 V_{t,s+1}^2 - 1) x_k^2 x_s^2] \\
=:& \, \kappa_x^{(2)2} B_t^{(1)} + B_t^{(2)} + B_t^{(3)}, \qquad \text{(A.117)}
\end{aligned}
$$

say. First, by lemma A.1-(d),

$$
2B_t^{(1)} = (\sum_{k=1}^{t} g_{k,t}^2)^2 + \sum_{k=1}^{t} g_{k,t}^4 = O(\gamma_t^2). \qquad \text{(A.118)}
$$

Next, by property (mix-i), Hölder's inequality and lemma A.1-(d),

$$\begin{aligned} B_t^{(2)} &\leq 6\,\|x_1\|_\infty^4 \sum_{k=2}^{t}\sum_{s=1}^{k} g_{k,t}^2 g_{s,t}^2 \alpha(k-s) \\ &= 6\,\|x_1\|_\infty^4 \sum_{k=2}^{t} g_{k,t}^2 \Phi_{t,k+1}^2 \sum_{s=1}^{k-1} g_{s,k}^2 \alpha(k-s) \\ &\leq 6\,\|x_1\|_\infty^4 \sum_{k=2}^{t} g_{k,t}^2 \Phi_{t,k+1}^2 (\sum_{s=1}^{k-1} g_{s,k}^4)^{1/2} (\sum_{s=1}^{\infty} \alpha(s)^2)^{1/2} \\ &\leq C6\,\|x_1\|_\infty^4 \sum_{k=2}^{t} \gamma_k^{7/2} \Phi_{t,k+1}^4 = O(\gamma_t^{5/2}), \end{aligned} \tag{A.119}$$

for some finite constant $C > 0$. Furthermore, as $g_{s,t}^2 g_{k,t}^2 = g_{k,t}^2 g_{s,k}^2 \Phi_{t,k+1}^2$ for $s < k \leq t$,

$$\begin{aligned} B_t^{(3)} &\leq \|x_1\|_\infty^4 \sum_{k=2}^{t}\sum_{s=1}^{k-1} g_{s,t}^2 g_{k,t}^2 E[|V_{t,k+1}^2 V_{t,s+1}^2 - 1|] \\ &\leq C\,\|x_1\|_\infty^4 (\sum_{k=1}^{t} g_{k,t}^2 \Phi_{t,k+1}^2)(\sum_{s=1}^{k} g_{s,k}^2 \gamma_k^{1-1/(2\eta)}) \\ &= C\,\|x_1\|_\infty^4 \sum_{k=1}^{t} g_{k,t}^2 \Phi_{t,k+1}^2 O(\gamma_k^{2-1/(2\eta)}) = O(\gamma_t^{3-1/(2\eta)}), \end{aligned} \tag{A.120}$$

using Hölder's inequality, lemma A.1-(d) and

$$E[|V_{t,k+1}^2 V_{t,s+1}^2 - 1|] = O(\gamma_k^{1-1/(2/\eta)}), \tag{A.121}$$

for $s \leq k \leq t$, which follows from the same arguments which lead to lemma A.2-(a). Hence, $B_t = O(\gamma_t^2)$, thereby proving the claim. □

A.4 Lemma A.5

Note. For the sake of better clarity and readability, lemma A.5 will be proven under the assumption of strongly mixing regressors (part (2) of assumption (B2)). For the corresponding proof under the assumption that $(\{x_t - E[x_1], \mathcal{F}_t\},\, t \geq 1)$ forms a martingale difference sequence (part (1) of assumption (B2)), please refer to section S.2.3 of the supplementary material.

Lemma A.5 *Recall that* $z_t = x_t a^*_{t-1}$, *with* $a^*_t = a_t - \alpha$. *Then,*

$$T^{-b}\sum_{t=1}^{T} z_t^2 \xrightarrow{p} \sigma^2 \frac{\gamma}{2cb} \qquad (a)$$

$$T^{-1/2}\sum_{t=1}^{T} a^*_t = O_p(1) \qquad (b)$$

$$T^{-1/2}\sum_{t=1}^{T}(x_t^2 - \kappa_x^{(2)})a^*_{t-1} = o_p(1) \qquad (c)$$

$$T^{-b/2}\sum_{t=1}^{T} z_t\varepsilon_t \xrightarrow{d} \mathcal{N}(0, \sigma^4\gamma/(2cb)). \qquad (d)$$

Remark A.6 *Equations* (*b*) *and* (*c*) *imply*

$$T^{-1/2}\sum_{t=1}^{T} x_t z_t = O_p(1) \;\Leftrightarrow\; T^{-(1+b)/2}\sum_{t=1}^{T} x_t z_t = O_p(T^{-b/2}), \qquad (e)$$

a result that is frequently used below.

Proof of lemma A.5 (*a*)**:** Set

$$T^{-b}\sum_{t=1}^{T} z_t^2 = \kappa_x^{(2)} T^{-b}\sum_{t=1}^{T} a^{*2}_{t-1} + T^{-b}\sum_{t=1}^{T}(x_t^2 - \kappa_x^{(2)})a^{*2}_{t-1} =: \kappa_x^{(2)} I_T + J_T, \qquad \text{(A.122)}$$

say. It will be shown that

$$(i)\ I_T = \sigma^2\frac{\gamma}{2cb} + o_p(1), \ \text{ and } \ (ii)\ J_T = o_p(1). \qquad \text{(A.123)}$$

Step (A.123)-(i). First, it will be shown that I_T converges in expectation to $\sigma^2\gamma/(2cb)$. From the remark accompanying lemma A.4, one gets $E[I_T] = \tau^2\gamma/(2cb) + o(1)$. Hence, by Chebychev's inequality, step (A.123)-(i) follows if $\mathrm{var}[I_T] = o(1)$, i.e. if

$$\mathrm{var}[\sum_{t=1}^{T} a^{*2}_t] = \sum_{t=1}^{T}\mathrm{var}[a^{*2}_t] + 2\sum_{t=1}^{T-1}\sum_{m=1}^{T-t}\mathrm{cov}[a^{*2}_t, a^{*2}_{t+m}] = o(T^{2b}). \qquad \text{(A.124)}$$

By lemma A.3 and assumption A, the first term on the right-hand side is bounded in T as it behaves like $\sum_{t\geq 1} O(\gamma_t^2)$. Now, one gets from equation (A.1) for $m \geq 0$

$$\begin{aligned} a^{*2}_{t+m} = & a^{*2}_{t+m-1} + \gamma^2_{t+m}(u^2_{t+m} + c^2 x^{*4}_{t+m}a^{*2}_{t+m-1} - 2c x^{*2}_{t+m}a_{t+m-1}u_{t+m}) \\ & + 2\gamma_{t+m}a^*_{t+m-1}(u_{t+m} - c x^{*2}_{t+m}a^*_{t+m-1}). \end{aligned} \qquad \text{(A.125)}$$

This yields

$$\begin{aligned} E[a^{*2}_t a^{*2}_{t+m}] = & E[a^{*2}_t a^{*2}_{t+m-1}] + (c\gamma_{t+m})^2 E[a^{*2}_t x^{*4}_{t+m}a^{*2}_{t+m-1}] \\ & - 2c\gamma_{t+m}E[a^{*2}_t x^{*2}_{t+m}a^{*2}_{t+m-1}] + \tau^2\gamma_{t+m}E[a^{*2}_t x^{*2}_{t+m}] \end{aligned} \qquad \text{(A.126)}$$

and

$$
\begin{aligned}
E[a_t^{*2}]E[a_{t+m}^{*2}] =& E[a_t^{*2}]E[a_{t+m-1}^{*2}] + (c\gamma_{t+m})^2 E[a_t^{*2}]E[x_{t+m}^{*4} a_{t+m-1}^{*2}] \\
&- 2c\gamma_{t+m} E[a_t^{*2}]E[x_{t+m}^{*2} a_{t+m-1}^{*2}] + \tau^2 \gamma_{t+m} E[a_t^{*2}]. \qquad \text{(A.127)}
\end{aligned}
$$

Putting the previous two equations together gets

$$
\operatorname{cov}[a_t^{*2}, a_{t+m}^{*2}] = \operatorname{cov}[a_t^{*2}, a_{t+m-1}^{*2}]\psi_{t+m} + \theta_{t+m}, \qquad \text{(A.128)}
$$

where $\psi_i := 1 - c\gamma_i \left(2 - c\gamma_i \kappa_x^{(4)} / \kappa_x^{(2)2}\right)$ and

$$
\theta_{t+m} := (\tau^2/\kappa_x^{(2)})\gamma_{t+m}^2 \theta_{t+m}^{(1)} + (c\gamma_{t+m}/\kappa_x^{(2)})((c\gamma_{t+m}/\kappa_x^{(2)})\theta_{t+m}^{(2)} - 2\theta_{t+m}^{(3)}), \qquad \text{(A.129)}
$$

with[13]

$$
\begin{aligned}
\theta_{t+m}^{(1)} &:= \operatorname{cov}[a_t^{*2}, x_{t+m}^2 - \kappa_x^{(2)}] \qquad &\text{(A.130)}\\
\theta_{t+m}^{(2)} &:= \operatorname{cov}[a_t^{*2}, a_{t+m-1}^{*2}(x_{t+m}^4 - \kappa_x^{(4)})] \qquad &\text{(A.131)}\\
\theta_{t+m}^{(3)} &:= \operatorname{cov}[a_t^{*2}, a_{t+m-1}^{*2}(x_{t+m}^2 - \kappa_x^{(2)})]. \qquad &\text{(A.132)}
\end{aligned}
$$

Solving equation (A.128) recursively, yields

$$
\operatorname{cov}[a_t^{*2}, a_{t+m}^{*2}] = \operatorname{var}[a_t^{*2}]\Psi_{t+m,t+1} + \sum_{j=1}^{m} \theta_{t+j}\Psi_{t+m,t+j+1}, \qquad \text{(A.133)}
$$

with

$$
\Psi_{m,n} := \prod_{i=n}^{m} \psi_i, \;\; n \le m+1. \qquad \text{(A.134)}
$$

It will be frequently be used that for $n \le m+1$ sufficiently large, $\Psi_{m,n} \le \Phi_{m,n} = \prod_{i=n}^{m}(1 - c\gamma_i)$; cf. equation (A.5). To see this, recall that $\psi_t = 1 - c\gamma_t(2 - c\gamma_t \kappa_x^{(4)}/\kappa_x^{(2)2})$ and define

$$
t_0 := \min\{t : \gamma_t \le \kappa_x^{(2)2}/(c\kappa_x^{(4)})\}. \qquad \text{(A.135)}
$$

Then, $\psi_t \le 1 - c\gamma_t$ for all $t \ge t_0$. Since, by lemma A.3, $\operatorname{var}[a_t^{*2}] = O(\gamma_t^2)$, one has

$$
\begin{aligned}
\sum_{t=1}^{T-1}\sum_{m=1}^{T-t} \operatorname{var}[a_t^{*2}]\Psi_{t+m,t+1} &= \sum_{t=1}^{T-1}\sum_{m=1}^{T-t} O(\gamma_t^2)\Psi_{t+m,t+1} \\
&= \sum_{t=2}^{T}\sum_{m=1}^{t-1} O(\gamma_m^2)\Psi_{t,m+1} \\
&\le \sum_{t=2}^{T}\sum_{m=1}^{t-1} O(\gamma_m g_{m,t}) = \sum_{t=2}^{T} O(\gamma_t) = O(T^b). \qquad \text{(A.136)}
\end{aligned}
$$

In deriving the above it has been used that for $n \le m+1$ sufficiently large, $\Psi_{m,n} \le \Phi_{m,n}$;

[13]Clearly, under part (1) of assumption (B2), $\theta_t = 0$ for all t.

see the discussion below equation (A.135). The final line of the preceding display uses lemma A.1-(d) and equation (A.9); recalling from equation (A.6) that $g_{m,t} = \gamma_m \Phi_{t,m+1}$. Hence, $\sum_{t=1}^{T-1}\sum_{m=1}^{T-t} \mathrm{var}[a_t^{*2}]\Psi_{t+m,t+1} = o(T^{2b})$. Taking equation (A.124) into account, the claim thus follows if

$$\sum_{t=1}^{T-1}\sum_{m=1}^{T-t}\sum_{j=1}^{m} \theta_{t+j}\Psi_{t+m,t+j+1} = o(T^{2b}).$$

Or, by equation (A.129), if

$$R_{1,T} \coloneqq \sum_{t=1}^{T-1}\sum_{m=1}^{T-t}\sum_{j=1}^{m} \gamma_{t+j}^2 \theta_{t+j}^{(1)} \Psi_{t+m,t+j+1} = o(T^{2b}), \tag{R-1}$$

$$R_{2,T} \coloneqq \sum_{t=1}^{T-1}\sum_{m=1}^{T-t}\sum_{j=1}^{m} \gamma_{t+j}^2 \theta_{t+j}^{(2)} \Psi_{t+m,t+j+1} = o(T^{2b}), \tag{R-2}$$

$$R_{3,T} \coloneqq \sum_{t=1}^{T-1}\sum_{m=1}^{T-t}\sum_{j=1}^{m} \gamma_{t+j} \theta_{t+j}^{(3)} \Psi_{t+m,t+j+1} = o(T^{2b}). \tag{R-3}$$

Proof of equation (R-1)**:** Consider $\theta_{t+m}^{(1)}$ introduced by equation (A.130) above. Define the sigma algebra

$$\mathcal{A}_t \coloneqq \sigma(\{(x_s,\, \varepsilon_s),\, s \le t\} \cup a_0), \tag{A.137}$$

and recall (from the discussion of assumption B) $\mathcal{V}_t = \sigma(\{(x_s,\, \varepsilon_s),\, s \le t\})$ and $\mathcal{F}_t = \sigma(\{x_s,\, s \le t\})$, so that $\mathcal{F}_t \subseteq \mathcal{V}_t \subseteq \mathcal{A}_t$. Then, by assumption (B3) and condition (iii) of assumption (B2),

$$E[x_{t+m}^2 \mid \mathcal{A}_t] = E[x_{t+m}^2 \mid \mathcal{V}_t] = E[x_{t+m}^2 \mid \mathcal{F}_t] \quad a.s.. \tag{A.138}$$

Therefore, by Cauchy-Schwarz's inequality, lemma A.3-(b) and property (mix-ii),

$$|\theta_{t+m}^{(1)}| = |E[a_t^{*2}(E[x_{t+m}^2 \mid \mathcal{F}_t] - \kappa_x^{(2)})]| \le \left\|a_t^{*2}\right\|_2 \left\|E[x_{t+m}^2 \mid \mathcal{F}_t] - \kappa_x^{(2)}\right\|_2 = O(\gamma_t \alpha(m)^{1/2}).$$

Hence, for some finite constant $C > 0$

$$\begin{aligned} R_{1,T} \le C \sum_{t=1}^{T-1} \gamma_t \sum_{m=1}^{T-t}\sum_{j=1}^{m} \gamma_{t+j}^2 \alpha(j)^{1/2} \Psi_{t+m,t+j+1} &= C \sum_{t=2}^{T}\sum_{m=1}^{t-1} \gamma_m \sum_{j=1}^{t-m} \gamma_{m+j}^2 \alpha(j)^{1/2} \Psi_{t,m+j+1} \\ &\le C \sum_{t=2}^{T}\sum_{m=1}^{t-1} \gamma_m \sum_{j=1}^{t-m} \gamma_{m+j}^2 \alpha(j)^{1/2} \Phi_{t,m+j+1}, \end{aligned}$$

where the final inequality uses, again, that for $n \le m+1$ sufficiently large, $\Psi_{m,n} \le \Phi_{m,n}$. Next, by Hölder's inequality, and lemma A.1-(d),

$$\sum_{j=1}^{t-m} \gamma_{m+j}^2 \alpha(j)^{1/2} \Phi_{t,m+j+1} \le C(\sum_{j=1}^{t-m} \gamma_{m+j}^4 \Phi_{t,m+j+1}^2)^{1/2} = O(\gamma_{t-m}^{3/2}), \tag{A.139}$$

where $C^2 \coloneqq \sum_{j=1}^{\infty} \alpha(j)$ is, by condition (ii) of assumption (B2), finite. Thus, as $3/2 >$

$2 - 1/(2\eta)$,

$$R_{1,T} \leq C \sum_{t=2}^{T} \sum_{m=1}^{t-1} \gamma_m \gamma_{t-m}^{3/2} \leq C \sum_{t=2}^{T} \sum_{m=1}^{t-1} \gamma_m \gamma_{t-m}^{2-1/(2\eta)}. \tag{A.140}$$

But, as shown below

$$T^{-2b} \sum_{t=2}^{T} \sum_{m=1}^{t-1} \gamma_m \gamma_{t-m}^{2-1/(2\eta)} \sim \begin{cases} T^{-(\eta-1/2)} & \text{if } \eta < 3/4 \\ T^{-b} \ln(T) & \text{if } \eta = 3/4 \\ T^{-b} & \text{if } \eta > 3/4. \end{cases} \tag{A.141}$$

Hence, $R_{1,t} = o(T^{2b})$.

Proof of equation (A.141)**:** The aim is to establish for

$$A_t := \int_1^{t-1} \gamma_k \gamma_{t-k}^{2-1/(2\eta)} dk, \tag{A.142}$$

that:

$$A_t \sim \begin{cases} \gamma_t^{3(1-1/(2\eta))} & \text{if } \eta < 3/4 \\ \gamma_t \ln(t) & \text{if } \eta = 3/4 \\ \gamma_t & \text{if } \eta > 3/4. \end{cases} \tag{A.143}$$

Since there exists some constant $C > 0$ so that

$$\sum_{k=1}^{t-1} \gamma_k \gamma_{t-k}^{2-1/(2\eta)} \leq C A_t, \tag{A.144}$$

equation (A.141) follows by (A.143) and equation (A.9). For the case of $\eta = 3/4$, note that

$$\int_1^T 1/t^\eta \ln(t) dt = O(T^b \ln(T)). \tag{A.145}$$

In order to verify (A.143), note that $A_t := (B_{t-1,t} - B_{1,t})$, with

$$B_{k,t} := k^b (t-k)^{3/2-2\eta} {}_2F_1(1, 5/2 - 3\eta; 2 - \eta; k/t)/(tb), \tag{A.146}$$

i.e.

$$B_{t-1,t} = (t-1)^b {}_2F_1(1, 5/2 - 3\eta; 2 - \eta; 1 - 1/t)/(tb) \tag{A.147}$$

$$B_{1,t} = (t-1)^{3/2-2\eta} {}_2F_1(1, 5/2 - 3\eta; 2 - \eta; 1/t)/(tb). \tag{A.148}$$

Here, ${}_2F_1(a, b; c; z)$ denotes Gauss's hypergeometric function

$${}_2F_1(a, b; c; z) := \nu \int_0^1 \frac{t^{b-1}(t-1)^{c-b-1}}{(1-zt)^a} dt, \text{ with } \nu := \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}; \tag{A.149}$$

see, e.g., Temme (2015, ch. 12). Let us consider the three cases: (*i*) $\eta > 3/4$, (*ii*) $\eta = 3/4$, (*iii*)

$\eta < 3/4$. Begin with (i) and note that

$$\begin{aligned} t^{\eta} B_{t-1,t} &= (1-1/t)^b {}_2F_1(1, 5/2-3\eta; 2-\eta; 1-1/t)/b \\ &\to {}_2F_1(1, 5/2-3\eta; 2-\eta; 1)/b = 2/(4\eta-3), \end{aligned} \tag{A.150}$$

where it has been used that $2-\eta > 1+5/2-3\eta$ for $\eta > 3/4$, and thus, by Gauss's theorem, ${}_2F_1(1, 5/2-3\eta; 2-\eta; 1) = 2b/(4\eta-3)$; see, e.g., Temme (2015, eq. (12.0.2)). Turning to case (ii), note that

$$t^{\eta}/\ln(t) B_{t-1,1} = 4(1-1/t)^{1/4} {}_2F_1(1, 1/4; 1+1/4; 1-1/t)/\ln(t). \tag{A.151}$$

But, ν in (A.149) is 1/4 while

$$\ln(t)^{-1} \int_0^1 \frac{u^{b-1}(u-1)^{c-b-1}}{(1-(1-1/t)u)^a} du = \ln(t)^{-1} \int_0^1 \frac{u^{1/4-1}}{1-(1-1/t)u} du \to 1. \tag{A.152}$$

Similarly, assume (iii) holds true and consider

$$\begin{aligned} t^{3\eta-3/2} B_{t-1,t} &= (1-1/t)^b \frac{{}_2F_1(1, 5/2-3\eta; 2-\eta; 1-1/t)}{bt^{3/2-2\eta}} \\ &= (1-1/t)^b \frac{{}_2F_1(1, a; 1+a-x; 1-1/t)}{bt^x}, \end{aligned} \tag{A.153}$$

with $a := 5/2-3\eta \in (1/4, 1)$ and $x := 3/2-2\eta \in (0, 1/2)$. But,

$$\frac{{}_2F_1(1, a; 1+a-x; 1-1/t)}{t^x} \to \frac{\Gamma(x)\Gamma(1+a-x)}{\Gamma(a)}. \tag{A.154}$$

Finally, for any a, b, c: ${}_2F_1(a, b; c; 0) = 1$, which proves the claim. □

Proof of equation (R-2)**:** Consider $\theta^{(2)}_{t+m}$ defined in equation (A.131):

$$\begin{aligned} |\theta^{(2)}_{t+m}| &\le |E[a^{*2}_t a^{*2}_{t+m-1}(x^4_{t+m} - \kappa^{(4)}_x)]| + |E[a^{*2}_t] E[a^{*2}_{t+m-1}(x^4_{t+m} - \kappa^{(4)}_x)]| \\ &\le (\|x_1\|^4_\infty + \kappa^{(4)}_x)(\|a^*_t\|^2_4 \left\|a^*_{t+m-1}\right\|^2_4 + E[a^{*2}_{t-1}]E[a^{*2}_{t+m-1}]) = O(\gamma_t \gamma_{t+m}), \end{aligned} \tag{A.155}$$

using Cauchy-Schwarz's and Hölder's inequality; the order of magnitude is due to lemma A.3. Therefore, by similar arguments which lead to equation (A.140), there exists some constant $C > 0$ such that

$$R_{2,T} \le C \sum_{t=2}^{T} \sum_{m=1}^{t-1} \gamma_m \sum_{j=1}^{t-m} \gamma^3_{m+j} \Phi_{t,m+j+1} = C \sum_{t=2}^{T} \sum_{m=1}^{t-1} \gamma_m O(\gamma^2_{t-m}) \le O(R_{1,t}). \tag{A.156}$$

Proof of equation (R-3)**:** Finally, in order to show $R_{3,T} = o(T^{2b})$, note that

$$|\theta^{(3)}_{t+m}| \le |E[a^{*2}_t a^{*2}_{t+m-1}(x^2_{t+m} - \kappa^{(2)}_x)]| + E[a^{*2}_t] |E[a^{*2}_{t+m-1}(x^2_{t+m} - \kappa^{(2)}_x)]|. \tag{A.157}$$

Furthermore, by lemma A.3, for $m \geq 0$

$$E[(x_{t+m}^2 - \kappa_x^{(2)})a_{t+m-1}^{*2}] = O(\gamma_{t+m}^{2-1/(2\eta)}) \tag{A.158}$$

$$\begin{aligned} E[(x_{t+m}^2 - \kappa_x^{(2)})a_{t-1}^{*2}a_{t+m-1}^{*2}] &= O(\gamma_t^{2-1/(2\eta)}\gamma_{t+m}) + O(g_{t,t+m}^2(\alpha(m) + \gamma_t^{1-1/(2\eta)})) \\ &\quad + o(\gamma_t\gamma_{t+m}^{2-1/(2\eta)}). \end{aligned} \tag{A.159}$$

Therefore, $R_{3,T} = o(T^{2b})$ follows if

$$\sum_{t=1}^{T-1} \gamma_t^2 \sum_{m=1}^{T-t} \sum_{j=1}^{m} \gamma_{t+j}\alpha(j)\Phi_{t+j,t+1}^2 \Psi_{t+m,t+j+1} = o(T^{2b}), \tag{R-3a}$$

$$\sum_{t=1}^{T-1} \gamma_t^{3-1/(2\eta)} \sum_{m=1}^{T-t} \sum_{j=1}^{m} \gamma_{t+j}\Phi_{t+j,t+1}^2 \Psi_{t+m,t+j+1} = o(T^{2b}), \tag{R-3b}$$

$$\sum_{t=1}^{T-1} \gamma_t^{2-1/(2\eta)} \sum_{m=1}^{T-t} \sum_{j=1}^{m} \gamma_{t+j}^2 \Psi_{t+m,t+j+1} = o(T^{2b}), \tag{R-3c}$$

$$\sum_{t=1}^{T-1} \gamma_t \sum_{m=1}^{T-t} \sum_{j=1}^{m} \gamma_{t+j}^{3-1/(2\eta)} \Psi_{t+m,t+j+1} = o(T^{2b}). \tag{R-3d}$$

Proof of equation (R-3a): Since $\Phi_{t+j,t+1}^2 \leq \Phi_{t+j,t+1}$ and $\Psi_{t+m,t+j+1} \leq \Phi_{t+m,t+j+1}$ for sufficiently large t, one has

$$\Phi_{t+j,t+1}^2 \Psi_{t+m,t+j+1} \leq C\Phi_{t+j,t+1}\Phi_{t+m,t+j+1} = C\Phi_{t+m,t+1}, \tag{A.160}$$

for some finite $C > 0$. Hence, for some finite $C > 0$, (R-3a) can be bounded from above by

$$\begin{aligned} C\sum_{t=1}^{T-1} \gamma_t^2 \sum_{m=1}^{T-t} \Phi_{t+m,t+1} \sum_{j=1}^{m} \gamma_{t+j}\alpha(j) &= C\sum_{t=2}^{T}\sum_{m=1}^{t-1} g_{m,t}\gamma_m \sum_{j=1}^{t-m} \gamma_{m+j}\alpha(j) \\ &\leq C(\sum_{j=1}^{\infty} \gamma_j\alpha(j)) \sum_{t=2}^{T}\sum_{m=1}^{t-1} g_{m,t}\gamma_m \\ &\leq C(\sum_{j=1}^{\infty} \gamma_j\alpha(j)) \sum_{t=1}^{T} O(\gamma_t) = O(T^b), \end{aligned} \tag{A.161}$$

using lemma A.1-(d) and equation (A.9). *Proof of equation* (R-3b): Similarly to equation (R-3a), there is some finite $C > 0$ such that equation (R-3b) can be bounded from above by

$$\begin{aligned} C\sum_{t=1}^{T-1} \gamma_t^{3-1/(2\eta)} \sum_{m=1}^{T-t} \Phi_{t+m,t+1} \sum_{j=1}^{m} \gamma_{t+j} &= C\sum_{t=2}^{T}\sum_{m=1}^{t-1} g_{m,t}\gamma_m^{2-1/(2\eta)} \sum_{j=1}^{t-m} \gamma_{m+j} \\ &= C\sum_{t=2}^{T}(\sum_{j=1}^{t} \gamma_j)(\sum_{m=1}^{t-1} g_{m,t}\gamma_m^{2-1/(2\eta)}) \\ &= C\sum_{t=1}^{T} O(t^b\gamma_t^{2-1/(2\eta)}). \end{aligned} \tag{A.162}$$

using $\sum_{j=1}^{t-m} \gamma_{m+j} \leq \sum_{j=1}^{t} \gamma_j = O(t^b)$ and lemma A.1-(d). Furthermore,

$$T^{-2b} \sum_{t=1}^{T} O(t^b \gamma_t^{2-1/(2\eta)}) = T^{-2b} \sum_{t=1}^{T} O(t^{3(1/2-\eta)}) \sim \begin{cases} T^{-(\eta-1/2)} & \text{if } \eta < 5/6 \\ T^{-2b} \ln(T) & \text{if } \eta = 5/6 \\ T^{-2b} & \text{if } \eta > 5/6, \end{cases} \tag{A.163}$$

building on the discussion surrounding equation (A.9). *Proof of equation* (R-3c): The partial sum (R-3c) is bounded from above by

$$\begin{aligned} C \sum_{t=1}^{T-1} \gamma_t^{2-1/(2\eta)} \sum_{m=1}^{T-t} \sum_{j=1}^{m} \gamma_{t+j}^2 \Phi_{t+m,t+j+1} &= C \sum_{t=2}^{T} \sum_{m=1}^{t-1} \gamma_m^{2-1/(2\eta)} \sum_{j=1}^{t-m} \gamma_{m+j} g_{t,m+j} \\ &= C \sum_{t=2}^{T} \sum_{m=1}^{t-1} O(\gamma_m^{2-1/(2\eta)} \gamma_{t-m}), \end{aligned} \tag{A.164}$$

where $C \in (0, \infty)$ and the order of magnitude is due to lemma A.1-(d). Since

$$\sum_{t=2}^{T} \sum_{m=1}^{t-1} \gamma_m \gamma_{t-m}^{2-1/(2\eta)} = \sum_{t=2}^{T} \sum_{m=1}^{t-1} \gamma_m^{2-1/(2\eta)} \gamma_{t-m}, \tag{A.165}$$

the claim follows by equation (A.141). *Proof of equation* (R-3d): Equation (R-3d) can be bounded by

$$\begin{aligned} C \sum_{t=1}^{T-1} \gamma_t \sum_{m=1}^{T-t} \sum_{j=1}^{m} \gamma_{t+j}^{3-1/(2\eta)} \Phi_{t+m,t+j+1} &= C \sum_{t=2}^{T} \sum_{m=1}^{t-1} \gamma_m \sum_{j=1}^{t-m} \gamma_{m+j}^{2-1/(2\eta)} g_{t,m+j} \\ &= C \sum_{t=2}^{T} \sum_{m=1}^{t-1} O(\gamma_m \gamma_{t-m}^{2-1/(2\eta)}), \end{aligned} \tag{A.166}$$

where $C \in (0, \infty)$ and, again, the order of magnitude is due to lemma A.1-part (d); the claim follows from (A.141). Putting equations (R-3a)-(R-3d) together verifies equation (R-3). This, in turn, proves (A.124), thereby establishing (A.123)-(i). □

Step (A.123)-(ii). Recall that $E[(x_t^2 - \kappa_x^{(2)}) a_{t-1}^{*2}] = O(\gamma_t^{2-1/(2\eta)})$, cf. equation ($a$). Hence,

$$E[J_T] = T^{-b} \sum_{t=1}^{T} E[(x_t^2 - \kappa_x^{(2)}) a_{t-1}^{*2}] = T^{-b} \sum_{t=1}^{T} O(\gamma_t^{2-1/(2\eta)}) = O(T^{-(\eta-1/2)}), \tag{A.167}$$

which is, by assumption A, $o(1)$. Next, define

$$\rho_{t,t+m} := \operatorname{cov}[(x_t^2 - \kappa_x^{(2)}) a_{t-1}^{*2}, (x_{t+m}^2 - \kappa_x^{(2)}) a_{t+m-1}^{*2}], \tag{A.168}$$

where it is understood that $\rho_t := \rho_{t,t} = \operatorname{var}[(x_t^2 - \kappa_x^{(2)}) a_{t-1}^{*2}]$. By Chebychev's inequality, the desired result thus follows if

$$\operatorname{var}[\sum_{t=1}^{T} (x_t^2 - \kappa_x^{(2)}) a_{t-1}^{*2}] = \sum_{t=1}^{T} \rho_t + 2 \sum_{t=1}^{T-1} \sum_{m=1}^{T-t} \rho_{t,t+m} = o(T^{2b}). \tag{A.169}$$

Due to the almost sure boundedness of the regressor and lemma A.3, $\rho_t = O(\gamma_t^2)$, thereby showing that the first term is bounded in T. Turning to the sum of covariances, consider

$$\begin{aligned}|\rho_{t,t+m}| \leq & |E[(x_t^2 - \kappa_x^{(2)})(x_{t+m}^2 - \kappa_x^{(2)})a_{t-1}^{*2}a_{t+m-1}^{*2}]| \\ & + |E[(x_t^2 - \kappa_x^{(2)})a_{t-1}^{*2}]||E[(x_{t+m}^2 - \kappa_x^{(2)})a_{t+m-1}^{*2}]| \\ = & |E[(x_t^2 - \kappa_x^{(2)})(x_{t+m}^2 - \kappa_x^{(2)})a_{t-1}^{*2}a_{t+m-1}^{*2}]| + O(\gamma_t^{2-1/(2\eta)}\gamma_{t+m}^{2-1/(2\eta)}),\end{aligned} \tag{A.170}$$

where the final line uses that $E[(x_{t+m}^2 - \kappa_x^{(2)})a_{t+m-1}^{*2}] = O(\gamma_{t+m}^{2-1/(2\eta)})$ for all $m \geq 0$, cf. equation (a). The first term on the majorant side of (A.170) is of size

$$O(\gamma_t^{2-1/(2\eta)}\gamma_{t+m}) + O(g_{t,t+m}^2(\alpha(m) + \gamma_t^{1-1/(2\eta)})) + o(\gamma_t\gamma_{t+m}^{2-1/(2\eta)}), \tag{A.171}$$

see remark **??** accompanying the proof of equation (b). Hence, $\sum_{t=1}^{T-1}\sum_{m=1}^{T-t}\rho_{t,t+m} = o(T^{2b})$ if

$$R_T^{(j)} := \sum_{t=1}^{T-1}\sum_{m=1}^{T-t} r_{t,m}^{(j)} = o(T^{2b}), \tag{A.172}$$

with $r_{t,m}^{(1)} := \gamma_t^{2-1/(2\eta)}\gamma_{t+m}$, $r_{t,m}^{(2)} := \gamma_t\gamma_{t+m}^{2-1/(2\eta)}$, $r_{t,m}^{(3)} := g_{t,t+m}^2\alpha(m)$ and $r_{t,m}^{(4)} := g_{t,t+m}^2\gamma_t^{1-1/(2\eta)}$.
Proof of $R_T^{(1)} = o(T^{2b})$: Note that

$$R_T^{(1)} = \sum_{t=2}^{T}\gamma_t\sum_{m=1}^{t-1}\gamma_m^{2-1/(2\eta)}. \tag{A.173}$$

From the discussion of equation (A.9),

$$\sum_{m=1}^{t-1}\gamma_m^{2-1/(2\eta)} \sim \begin{cases} t^{2\eta-3/2} & \text{if } \eta < 3/4 \\ \ln(t) & \text{if } \eta = 3/4 \\ C & \text{if } \eta > 3/4, \end{cases} \tag{A.174}$$

for some finite $C > 0$. Hence,[14]

$$T^{-2b}\sum_{t=2}^{T}\gamma_t\sum_{m=1}^{t-1}\gamma_m^{2-1/(2\eta)} \sim \begin{cases} T^{-(\eta-1/2)} & \text{if } \eta < 3/4 \\ T^{-b}\ln(T) & \text{if } \eta = 3/4 \\ T^{-b} & \text{if } \eta > 3/4, \end{cases} \tag{A.176}$$

which proves $R_T^{(1)} = o(T^{2b})$. *Proof of* $R_T^{(2)} = o(T^{2b})$: $R_T^{(2)}$ can be rewritten as

$$\sum_{t=2}^{T}\gamma_t^{2-1/(2\eta)}\sum_{m=1}^{t-1}\gamma_m, \tag{A.177}$$

[14] For the case of $\eta = 3/4$, note that

$$\int_1^T 1/t^\eta \ln(t)dt = O(T^b\ln(T)). \tag{A.175}$$

which, as $\sum_{m=1}^{t-1}\gamma_m = O(t^b)$ (cf. equation (A.9)), implies $R_T^{(2)} = o(T^{2b})$ (cf. equation (A.163)). *Proof of* $R_T^{(3)} = o(T^{2b})$: $R_T^{(3)}$ can be rewritten as

$$R_T^{(3)} = \sum_{t=2}^{T}\sum_{m=1}^{t-1} g_{m,t}^2 \alpha(t-m) \leq C\sum_{t=1}^{T}(\sum_{m=1}^{t-1} g_{m,t}^4)^{1/2} = C\sum_{t=1}^{T} O(\gamma_t^{3/2}) = o(T^{2b}), \tag{A.178}$$

using Hölder's inequality, lemma A.1-(d) and equation (A.9); with $C^2 := \sum_{m=1}^{\infty}\alpha(m)^2$ being finite by condition (ii) of assumption (B2). *Proof of* $R_T^{(4)} = o(T^{2b})$: By lemma A.1-(d), $R_T^{(4)}$ can be rewritten as

$$R_T^{(4)} = \sum_{t=2}^{T}\sum_{m=1}^{t-1} g_{m,t}^2 \gamma_m^{1-1/(2\eta)} = \sum_{t=1}^{T} O(\gamma_t^{2-1/(2\eta)}) = o(T^{2b}), \tag{A.179}$$

using equation (A.174) for the final conclusion. This proves (A.169) and, therefore, the claim that $J_T = o_p(1)$. This completes the proof of lemma A.5-(a). □

Proof of lemma A.5 (b)**:** From equation (A.4) one deduces that

$$T^{-1/2}\sum_{t=1}^{T} a_t^* = T^{-1/2} a_0^* \sum_{t=1}^{T}\vartheta_t + T^{-1/2}\sum_{t=1}^{T}\xi_t =: A_T + B_T, \tag{A.180}$$

say. It follows from lemma A.2-(b) and Hölder's inequality that for some $\lambda_p > 0$ and $p > 0$, $\|\vartheta_t\|_p = O(\exp\{-\lambda_p t^b\})$. Furthermore, by equation (G-d) and the integral comparison test, $\sum_{t=1}^{T} O(\exp\{-\lambda_p t^b\}) = O(1)$. It will be shown that

$$(i)\ A_T = O_{a.s.}(T^{-1/2}) \text{ and } (ii)\ B_T = O_p(1). \tag{A.181}$$

Begin with (i). The above and Cauchy-Schwarz's inequality ensure that there exists some $\lambda > 0$ such that

$$\sum_{t=1}^{\infty} E[|a_0^*\vartheta_t|] \leq \kappa_a^{(2)\,1/2}\sum_{t=1}^{\infty}\|\vartheta_t\|_2 = \sum_{t=1}^{\infty} O(\exp\{-\lambda t^b\}) < \infty, \tag{A.182}$$

which, in turn, implies that $\sum_{t=1}^{\infty} a_0^*\vartheta_t$ converges $a.s.$. Turning to (ii), observe that $E[B_T] = 0$, while

$$E[B_T^2] = T^{-1}\sum_{t=1}^{T} E[\xi_t^2] + 2T^{-1}\sum_{t=2}^{T}\sum_{s=1}^{t-1} E[\xi_t\xi_s]. \tag{A.183}$$

Suppose $1 \leq k \leq s \leq t$: Since $\Phi_{t,s+1}\Phi^2_{s,k+1} = \Phi_{t,k+1}\Phi_{s,k+1}$, one has

$$\begin{aligned} E[\xi_t\xi_s] = \tau^2 \sum_{k=1}^{s} \gamma_k^2 E[\vartheta_{t,k+1}\vartheta_{s,k+1}x_k^{*2}] &= \tau^2 \sum_{k=1}^{s} \gamma_k^2 E[\vartheta_{t,s+1}\vartheta^2_{s,k+1}x_k^{*2}] \\ &= \tau^2 \sum_{k=1}^{s} g_{k,t}g_{k,s} \\ &\quad + \tau^2\Phi_{t,s+1}\sum_{k=1}^{s} g^2_{k,s}E[x_k^{*2}(V_{t,s+1}V^2_{s,k+1}-1)], \end{aligned} \tag{A.184}$$

where the notation $V_{m,n} = \vartheta_{m,n}/\Phi_{m,n}$ has been introduced in lemma A.2. Using Hölder's inequality, the second term on the right-hand side can be bounded by

$$(\tau^2/\kappa_x^{(2)})\, \|x_1\|^2_\infty\, \Phi_{t,s+1} \sum_{k=1}^{s} g^2_{k,s} E[|V_{t,s+1}V^2_{s,k+1} - 1|]. \tag{A.185}$$

By lemma A.2-(a), $E[|V_{t,s+1}V^2_{s,k+1}-1|] = O(\gamma_k^{1-1/2\eta})$ for $k \leq s \leq t$. Thus, by lemma A.1-(d), for $s \leq t$

$$E[\xi_t\xi_s] = \tau^2 \sum_{k=1}^{s} g_{k,t}g_{k,s} + O(\gamma_s^{1-1/(2\eta)}g_{s,t}). \tag{A.186}$$

Specifically, this yields for $s = t$

$$E[\xi_t^2] = \tau^2\phi_t^{ii} + O(\gamma_t^{2-1/(2\eta)}), \tag{A.187}$$

where $\phi_t^{ii} = \sum_{k=1}^{t} g^2_{k,t}$ has been defined in lemma A.1-(b). Hence,

$$E[B_T^2] = T^{-1}\Big(\sum_{t=1}^{T} O(\phi_t^{ii} + \gamma_t^{2-1/(2\eta)}) + \sum_{t=2}^{T}\sum_{s=1}^{t-1} O(g_{k,t}g_{k,s} + \gamma_s^{1-1/(2\eta)}g_{s,t})\Big). \tag{A.188}$$

By lemma A.1-(b), $T^{-1}\sum_{t=1}^{T}\phi_t^{ii} = o(1)$, while, by lemma A.1-($c$),

$$T^{-1}\sum_{t=2}^{T}\sum_{s=1}^{t-1} g_{k,t}g_{k,s} = T^{-1}\sum_{t=2}^{T}\phi_t^{iii} = O(1). \tag{A.189}$$

The claim thus follows from

$$T^{-1}\sum_{t=1}^{T}\gamma_t^{2-1/(2\eta)} = O(T^{-2\eta-1/2}) = o(1) \tag{A.190}$$

$$T^{-1}\sum_{t=2}^{T}\sum_{s=1}^{t-1} g_{s,t}\gamma_s^{1-1/(2\eta)} = T^{-1}\sum_{t=1}^{T} O(\gamma_t^{1-1/(2\eta)}) = O(T^{-(\eta-1/2)}) = o(1), \tag{A.191}$$

using assumption A and lemma A.1-(d). This shows that $B_T = O_p(1)$, thereby proving the claim. □

Proof of lemma A.5 (c)**:** Similar to the proof of part (b) above, define

$$A_T := T^{-1/2}\sum_{t=1}^{T}(x_t^2-\kappa_x^{(2)})a_0^*\vartheta_{t-1} \quad\text{and}\quad B_T := T^{-1/2}\sum_{t=1}^{T}(x_t^2-\kappa_x^{(2)})\xi_{t-1}, \tag{A.192}$$

say. From equation (A.4) one gets $T^{-1/2}\sum_{t=1}^{T}(x_t^2-\kappa_x^{(2)})a_{t-1}^* = A_T+B_T$. It will be shown that both, A_T and B_T are $o_p(1)$. From the proof of part (b) above, it follows readily that $A_T = o_{a.s.}(1)$. Turning to B_T, note that $E[(x_t^2-\kappa_x^{(2)})\xi_{t-1}] = 0$. The claim thus follows, by Chebychev's inequality, if $\text{var}[B_T] = o(1)$. Note that,

$$\begin{aligned}\text{var}[B_T] =& T^{-1}\sum_{t=1}^{T}E[(x_t^2-\kappa_x^{(2)})^2\xi_{t-1}^2] \qquad\qquad (A.193)\\ &+2T^{-1}\sum_{t=2}^{T}\sum_{s=1}^{t-1}E[(x_t^2-\kappa_x^{(2)})(x_s^2-\kappa_x^{(2)})\xi_{t-1}\xi_{s-1}] =: \Gamma_{1,T}+2\Gamma_{2,T},\end{aligned}$$

say. The proof is complete if $\Gamma_{1,T} = o(1)$ and $\Gamma_{2,T} = o(1)$. Since, by Hölder's inequality, $E[(x_t^2-\kappa_x^{(2)})^2\xi_{t-1}^2] \le (\|x_1\|_\infty^2+\kappa_x^{(2)})^2E[\xi_{t-1}^2]$, the discussion following (A.187) yields

$$\Gamma_{1,T} = T^{-1}\sum_{t=1}^{T}O(\phi_t^{ii}+\gamma_t^{2-1/(2\eta)}) = o(1). \tag{A.194}$$

Using the strict exogeneity of $(x_k,\, k\ge 1)$ and arguments similar to those which lead to equation (A.184), it is seen that $E[(x_t^2-\kappa_x^{(2)})(x_s^2-\kappa_x^{(2)})\xi_{t-1}\xi_{s-1}] = \tau^2(A_{s-1,t-1}+B_{s-1,t-1})$, with

$$A_{s,t} := \sum_{k=1}^{s}g_{k,t}g_{k,s}E[(x_{s+1}^2-\kappa_x^{(2)})(x_{t+1}^2-\kappa_x^{(2)})] \tag{A.195}$$

$$B_{s,t} := \Phi_{t,s+1}\sum_{k=1}^{s}g_{k,s}^2E[(x_{s+1}^2-\kappa_x^{(2)})(x_{t+1}^2-\kappa_x^{(2)})x_k^{*2}(V_{t,s+1}V_{s,k+1}^2-1)]. \tag{A.196}$$

Hence,

$$\Gamma_{2,T} = T^{-1}\tau^2\Big(\sum_{t=2}^{T}\sum_{s=1}^{t-1}A_{s-1,t-1}+\sum_{t=2}^{T}\sum_{s=1}^{t-1}B_{s-1,t-1}\Big). \tag{A.197}$$

Using respectively Hölder's inequality and property (mix-ii) reveals,

$$E[(x_s^2-\kappa_x^{(2)})(x_t^2-\kappa_x^{(2)})] \le (\|x_1\|_\infty^2+\kappa_x^{(2)})E[|E[x_t^2 \mid \mathcal{F}_s]-\kappa_x^{(2)}|] \le C\alpha(t-s), \tag{A.198}$$

for $s\le t$ and $C := \|x_1\|_\infty^2(\|x_1\|_\infty^2+\kappa_x^{(2)})(8^{1/2}+2)$. Hence, taking $g_{k,t}g_{k,s} = \Phi_{t,s+1}g_{k,s}^2$

$(k \leq s \leq t)$ into account, one gets

$$\sum_{s=1}^{t-1} A_{s,t} = \sum_{s=1}^{t-1}\sum_{k=1}^{s} g_{k,t} g_{k,s} E[(x_{s+1}^2 - \kappa_x^{(2)})(x_{t+1}^2 - \kappa_x^{(2)})] = \sum_{s=1}^{t-1} O(\alpha(t-s)) \sum_{k=1}^{s} g_{k,t} g_{k,s}$$
$$= \sum_{s=1}^{t-1} O(\alpha(t-s))\Phi_{t,s+1} \sum_{k=1}^{s} g_{k,s}^2$$
$$= \sum_{s=1}^{t-1} O(\alpha(t-s) g_{s,t}) \qquad \text{(A.199)}$$

using repeatedly lemma A.1-(d). Therefore, there exists some finite $C > 0$ such that

$$\sum_{s=1}^{t-1} A_{s,t} \leq C \sum_{s=1}^{t-1} \alpha(t-s) g_{s,t} \leq C(\sum_{s=1}^{\infty} \alpha(s)^2)^{1/2} (\sum_{s=1}^{t} g_{s,t}^2)^{1/2} = O(\gamma_t^{1/2}), \qquad \text{(A.200)}$$

where the second inequality is due to Hölder's inequality and the order of magnitude is due to lemma A.4-(d). Hence, $T^{-1} \sum_{t=2}^{T} \sum_{s=1}^{t-1} A_{s-1,t-1} = O(T^{-\eta/2})$. Turning to the second term, by Hölder's inequality,

$$E[(x_{t+1}^2 - \kappa_x^{(2)})(x_{s+1}^2 - \kappa_x^{(2)}) x_k^{*2} (V_{t,s+1} V_{s,k+1}^2 - 1)] \leq C E[|V_{t,s+1} V_{s,k+1}^2 - 1|]. \qquad \text{(A.201)}$$

with $C \coloneqq (\|x_1\|_\infty^2 / \kappa_x^{(2)})(\|x_1\|_\infty^2 + \kappa_x^{(2)})^2$. For $k \leq s \leq t$, lemma A.2-(a) implies

$$E[|V_{t,s+1} V_{s,k+1}^2 - 1|] = O(\gamma_k^{1-1/(2\eta)}), \qquad \text{(A.202)}$$

such that, by lemma A.4-(d),

$$B_{s,t} \leq C\Phi_{t,s+1} \sum_{k=1}^{s} g_{k,s}^2 O(\gamma_k^{1-1/(2\eta)}) = O(g_{s,t} \gamma_s^{1-1/(2\eta)}). \qquad \text{(A.203)}$$

Therefore, the second term in (A.197) behaves like

$$T^{-1} \sum_{t=2}^{T} \sum_{s=1}^{t-1} O(g_{s,t} \gamma_s^{1-1/(2\eta)}) = T^{-1} \sum_{t=1}^{T} O(\gamma_t^{1-1/(2\eta)}) = O(T^{-(\eta-1/2)}), \qquad \text{(A.204)}$$

using again lemma A.4-(d). This shows that $\text{var}[B_T] = o(1)$, which, implies $B_T = o_p(1)$; thereby proving part (c) of this lemma. □

Proof of lemma A.5 (d)**:** The following proof is based on corollary 24.14 which accompanies Central Limit Theorem (CLT) 24.6 in Davidson (1994). The corollary allows for asymptotically (as $t \to \infty$) degenerate variances of the underlying stochastic process. Here, the CLT is applied to scaled partial sums of $z_t \varepsilon_t$. By assumption B, $(\{z_k \varepsilon_k, \mathcal{A}_k\}, k \geq 1)$ forms a martingale difference sequence as

$$E[z_t \varepsilon_t \mid \mathcal{A}_{t-1}] = a_{t-1}^* E[x_t \varepsilon_t \mid \mathcal{A}_{t-1}] = a_{t-1}^* E[x_t \varepsilon_t \mid \mathcal{V}_{t-1}]$$
$$= a_{t-1}^* E[x_t E[\varepsilon_t \mid \mathcal{G}_{t-1}] \mid \mathcal{V}_{t-1}] = 0, \qquad \text{(A.205)}$$

with probability one. Recall that $\mathcal{A}_t = \sigma(\{(x_s, \varepsilon_s), s \le t\} \cup a_0)$, $\mathcal{G}_t = \sigma(\{x_k, k \ge 1\} \cup \{\varepsilon_s, s \le t\})$ and $\mathcal{V}_t = \sigma(\{(x_s, \varepsilon_s), s \le t\})$; while assumption (B3) ensures that a_0 is independent of $\mathcal{V}_t$. Furthermore, the unconditional variances $\sigma_t^2 := E[(z_t\varepsilon_t)^2] = \sigma^2 E[z_t^2]$ behave approximately like γ_t; see the remark accompanying lemma A.4. In order to apply the corollary in Davidson (1994), it is helpful to introduce the following array notation:

$$Z_{tT} := z_t\varepsilon_t / s_T \quad \text{with} \quad s_T^2 := \sum_{t=1}^{T} \sigma_t^2. \tag{A.206}$$

Note that Z_{tT} inherits the martingale difference property from z_t and, taking account of part (c) of lemma A.4, $s_T^2 = O(T^b)$. Now, according to Davidson (1994, corollary 24.14) if

i. there exists a positive constant array c_{tT} so that $\|Z_{tT}/c_{tT}\|_r < \infty$ uniformly for some $r > 2$;

ii. $M_T = o(1)$, where $M_T := \max\limits_{1 \le t \le T} c_{tT}$;

iii. $\sum_{i=1}^{T} M_{iT}^2 = O(1)$, where $M_{iT} := \max\limits_{\substack{(i-1) \le t \le i \\ i=1,\ldots,T}} c_{tT}$;

then

$$\sum_{t=1}^{T} Z_{tT} \xrightarrow{d} \mathcal{N}(0,1). \tag{A.207}$$

If one lets $c_{tT} := \sigma_t / s_T$, then, following the argumentation in Davidson (1994) surrounding the corollary, it can be seen that condition *ii.* and *iii.* are satisfied. Specifically, note that $c_{tT}^2 \sim t^{-\eta}T^{\eta-1}$ and thus $M_{iT}^2 \sim (i-1)^{-\eta}T^{\eta-1}$; see also the proof of lemma 24.12 in Davidson (1994). Hence, it suffices to verify that $\sup_{t,T} \|Z_{tT}/c_{tT}\|_4 < \infty$. But, $E[(Z_{tT}/c_{tT})^4] \sim \kappa_\varepsilon^{(4)} \gamma_t^{-2} E[z_t^4]$, which is, by lemma A.4, $O(1)$ uniformly in t. Furthermore,

$$\sum_{t=1}^{T} Z_{tT} = (T^{-b} \sum_{t=1}^{T} \sigma_t^2)^{-1/2} \sum_{t=1}^{T} z_t. \tag{A.208}$$

The claim thus follows from

$$\lim_{T \to \infty} T^{-b} \sum_{t=1}^{T} \sigma_t^2 = \frac{\gamma\sigma^4}{2cb}; \tag{A.209}$$

see the the remark below lemma A.4. This completes the proof. □

A.5 Lemma A.6

Lemma A.6 *Suppose* $\|r_0\|_2 < \infty$. *Then,*

$$r_t \xrightarrow{a.s.} \kappa_x^{(2)}.$$

Proof of lemma A.6: In view of equation (A.5) and (A.6), define

$$\bar{\Phi}_{t,k} \coloneqq \prod_{i=k}^{t} (1-\gamma_i),$$

with $\bar{\Phi}_{t,t+1} \coloneqq 1$ and $\bar{g}_{k,t} \coloneqq \bar{\Phi}_{t,k+1}\gamma_k$. Then,

$$r_t^* - 1 = (r_0^* - 1)\bar{\Phi}_t + \sum_{k=1}^{t} \bar{g}_{k,t}(x_k^{*2} - 1), \tag{A.210}$$

where $r_t^* \coloneqq r_t/\kappa_x^{(2)}$ and it has been used that $\sum_{k=1}^{t} \bar{g}_{k,t} + \bar{\Phi}_t = 1$. For any $\delta > 0$, it follows from Markov's inequality, equations (A.13) and (G-d) that

$$\sum_{t=1}^{\infty} \mathrm{P}(|r_0 - \kappa_x^{(2)}|\bar{\Phi}_t > \delta) \leq E[(r_0 - \kappa_x^{(2)})^2] \sum_{t=1}^{\infty} \bar{\Phi}_t^2 \delta^{-2} < \infty. \tag{A.211}$$

Hence, by the Borel-Cantelli lemma, $(r_0^* - 1)\bar{\Phi}_t \to 0$ *a.s.*. It thus remains to show that

$$\sum_{k=1}^{t} \bar{g}_{k,t}(x_k^2 - \kappa_x^{(2)}) \overset{a.s.}{\to} 0. \tag{A.212}$$

To begin with, when t is sufficiently large, there exists some k_0 such that $(\bar{g}_{k,t},\, k > k_0)$ is an increasing sequence. To see this, note that $\bar{g}_{k,t} \leq \bar{g}_{k+1,t} \Leftrightarrow \gamma_k/\gamma_{k+1} - \gamma_k \leq 1$. But this follows, for k sufficiently large, from assumption A, which implies $\gamma_k/\gamma_{k+1} = 1 + o(\gamma_k)$; see, e.g., Kushner and Yan (1993). Furthermore, it follows from the definition of $\bar{g}_{k,t}$ that

$$\max_{k_0 \leq k \leq t} \bar{g}_{k,t} = \gamma_t. \tag{A.213}$$

Hence, by theorem 6 in Miao and Xu (2013)[15], a sufficient condition for (A.212) is the following Marcinkiewicz-Zygmund type SLLN:

$$t^{-\eta} \sum_{k=1}^{t} (x_k^2 - \kappa_x^{(2)}) \overset{a.s.}{\to} 0. \tag{A.214}$$

Depending on whether assumption part (1) or (2) of assumption (B2) is true, the preceding display follows from Stoica (2011, theorem 1) or Rio (1995, theorem 1, comment 1), respectively. □

B Proofs of the main results

To begin with, let $m_{\mathcal{T}} \coloneqq T^{-1}\sum_{t=1}^{T} x_t^2$ and $u_{\mathcal{T}} \coloneqq T^{-1/2}\sum_{t=1}^{T} u_t$. Then,

$$(a)\ m_{\mathcal{T}} \overset{a.s.}{\longrightarrow} \kappa_x^{(2)},\ (b)\ u_{\mathcal{T}} \overset{d}{\longrightarrow} \mathcal{N}(0,\tau^2). \tag{B.1}$$

[15]See also the corresponding remark by da Silva (2015)

Proof of part (a): If part (1) of assumption (B2) holds, $(\{x_k^2 - \kappa_x^{(2)}, \mathcal{F}_k\},\, k \geq 1)$ forms a martingale difference sequence with constant conditional variance. The claim follows, for example, from Davidson (1994, theorem 20.10). If part (2) of assumption (B2) holds, the claim follows because x_t^2 is mixing and strictly stationary and therefore ergodic. **Proof of part (b):** Note that, by assumption (B1), $(\{u_k, \mathcal{G}_k\},\, k \geq 1)$ forms a martingale difference sequence with variance τ^2. The claim follows, for example, from theorem 24.3 in Davidson (1994).

B.1 Proof of proposition 2.1

Proof of proposition 2.1: To begin with, let us restate here for convenience the definition of the regressor (sample) second moment matrix $M_T = \sum_{t=1}^T w_t w_t'$ with $w_t = (x_t a_{t-1},\, x_t)'$; cf. equation (2.8). The scaled deviation of the joint OLS estimator (2.7) in deviation from λ can then be written as

$$T^{b/2}(\widehat{\lambda} - \lambda) = T^b M_T^{-1} \frac{1}{T^{b/2}} \sum_{t=1}^T w_t \varepsilon_t = \frac{T^{1+b}}{\det M_T} \left(\frac{Q_T/T}{T^{b/2}} \sum_{t=1}^T w_t \varepsilon_t \right), \tag{B.2}$$

using that $M_T^{-1} = Q_T / \det M_T$, with

$$Q_T := \sum_{t=1}^T \begin{bmatrix} x_t^2 & -x_t^2 a_{t-1} \\ -x_t^2 a_{t-1} & (x_t a_{t-1})^2 \end{bmatrix} \tag{B.3}$$

and

$$\det M_T := \sum_{t=1}^T x_t^2 \sum_{t=1}^T (x_t a_{t-1})^2 - \left(\sum_{t=1}^T x_t^2 a_{t-1} \right)^2 . \tag{B.4}$$

The proof is based on the following three steps

$$\operatorname*{plim}_{T\to\infty} \frac{\det M_T}{T^{1+b}} = \sigma^2 \kappa_x^{(2)} \gamma / (2cb) \tag{a}$$

$$\operatorname*{plim}_{T\to\infty} \frac{Q_T}{T} = \kappa_x^{(2)} \begin{bmatrix} 1 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix} \tag{b}$$

$$\frac{Q_T/T}{T^{b/2}} \sum_{t=1}^T w_t \varepsilon_t \xrightarrow{d} \mathcal{N}\left(0, \frac{\kappa_x^{(2)2} \gamma \sigma^4}{2cb} \begin{bmatrix} 1 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix}\right). \tag{c}$$

Step (a). Recall the definitions of $z_t = x_t a_{t-1}^*$ and $m_{\mathcal{T}} := \frac{1}{T} \sum_{t=1}^T x_t^2$, where the latter expression approaches $\kappa_x^{(2)}$ *a.s.* by (B.1). Now, taking

$$\sum_{t=1}^T x_t^2 a_{t-1} = \alpha \sum_{t=1}^T x_t^2 + \sum_{t=1}^T x_t z_t \tag{B.5}$$

$$\sum_{t=1}^T (x_t a_{t-1})^2 = \alpha^2 \sum_{t=1}^T x_t^2 + 2\alpha \sum_{t=1}^T x_t z_t + \sum_{t=1}^T z_t^2 \tag{B.6}$$

into account, one gets, after some rearrangement

$$\frac{\det M_T}{T^{1+b}} = m_T(T^{-b}\sum_{t=1}^{T} z_t^2 - m_T^{-1}(T^{-(1+b)/2}\sum_{t=1}^{T} x_t z_t)^2). \tag{B.7}$$

Now, by lemma A.5-(d), $T^{-b}\sum_{t=1}^{T} z_t^2 = \sigma^2\gamma/(2cb) + o_p(1)$. Consequently, taking equation (e) of lemma A.5, equation (B.7) and $m_T \to \kappa_x^{(2)}$ *a.s.* into account, yields

$$\frac{\det M_T}{T^{1+b}} = \frac{\kappa_x^{(2)}}{T^b}\sum_{t=1}^{T} z_t^2 + O_p(T^{-b}) = \frac{\kappa_x^{(2)}\sigma^2\gamma}{2cb} + o_p(1). \tag{B.8}$$

This completes the proof of step (a).

Step (b). Again, using equations (B.5) and (B.6) yields

$$Q_T/T = m_T\begin{bmatrix} 1 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix} + \begin{bmatrix} 0 & -\frac{1}{T}\sum_{t=1}^{T} x_t z_t \\ -\frac{1}{T}\sum_{t=1}^{T} x_t z_t & \frac{2\alpha}{T}\sum_{t=1}^{T} x_t z_t + \frac{1}{T}\sum_{t=1}^{T} z_t^2 \end{bmatrix}.$$

In view of $m_T \to \kappa_x^{(2)}$ *a.s.*, the claim follows because the elements of the second matrix are $o_p(1)$ by the same arguments, which have been used in step (1). Since the entries of M_T/T are found in rearranged order in Q_T/T, the convergence in probability of M_T/T mentioned in the main text is readily deduced from the above.

Step (c). The entries of the 2×1 vector

$$\frac{Q_T/T}{T^{b/2}}\sum_{t=1}^{T} w_t\varepsilon_t =: (\zeta_T^{(1)}, \zeta_T^{(2)})', \tag{B.9}$$

say, are respectively given by

$$\zeta_T^{(1)} = \frac{1}{T}\sum_{t=1}^{T} x_t^2 \frac{1}{T^{b/2}}\sum_{t=1}^{T} a_{t-1}x_t\varepsilon_t - \frac{1}{T}\sum_{t=1}^{T} x_t^2 a_{t-1}\frac{1}{T^{b/2}}\sum_{t=1}^{T} x_t\varepsilon_t \tag{B.10}$$

$$\zeta_T^{(2)} = \frac{1}{T}\sum_{t=1}^{T} (x_t a_{t-1})^2 \frac{1}{T^{b/2}}\sum_{t=1}^{T} x_t\varepsilon_t - \frac{1}{T}\sum_{t=1}^{T} x_t^2 a_{t-1}\frac{1}{T^{b/2}}\sum_{t=1}^{T} a_{t-1}x_t\varepsilon_t. \tag{B.11}$$

Making again use of equations (B.5) and (B.6), it is seen that equations (B.10) and (B.11) can respectively be rewritten as

$$\zeta_T^{(1)} = \frac{m_T}{T^{b/2}} \sum_{t=1}^{T} z_t \varepsilon_t - u_{\mathcal{T}} \left(\frac{\kappa_x^{(2)}}{T^{(1+b)/2}} \sum_{t=1}^{T} x_t z_t \right) \tag{B.12}$$

$$\begin{aligned} \zeta_T^{(2)} = & - \frac{\alpha m_T}{T^{b/2}} \sum_{t=1}^{T} z_t \varepsilon_t + u_{\mathcal{T}} \left(\frac{\alpha \kappa_x^{(2)}}{T^{(1+b)/2}} \sum_{t=1}^{T} x_t z_t + \frac{\kappa_x^{(2)}}{T^{(1+b)/2}} \sum_{t=1}^{T} z_t^2 \right) \\ & - \frac{1}{T} \sum_{t=1}^{T} x_t z_t \frac{1}{T^{b/2}} \sum_{t=1}^{T} z_t \varepsilon_t, \end{aligned} \tag{B.13}$$

where $u_{\mathcal{T}}$ has been defined in (B.1). Therefore,

$$\begin{aligned} \frac{Q_T/T}{T^{b/2}} \sum_{t=1}^{T} w_t \varepsilon_t = & \begin{bmatrix} 1 \\ -\alpha \end{bmatrix} \frac{m_T}{T^{b/2}} \sum_{t=1}^{T} z_t \varepsilon_t \\ & - \begin{bmatrix} 1 \\ -\alpha \end{bmatrix} \left(\frac{\kappa_x^{(2)}}{T^{(1+b)/2}} \sum_{t=1}^{T} x_t z_t \right) u_{\mathcal{T}} \\ & + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \left\{ \left(\frac{\kappa_x^{(2)}}{T^{(1+b)/2}} \sum_{t=1}^{T} z_t^2 \right) u_{\mathcal{T}} - \left(\frac{1}{T} \sum_{t=1}^{T} x_t z_t \right) \frac{1}{T^{b/2}} \sum_{t=1}^{T} z_t \varepsilon_t \right\} \\ = & \begin{bmatrix} 1 \\ -\alpha \end{bmatrix} \frac{m_T}{T^{b/2}} \sum_{t=1}^{T} z_t \varepsilon_t + O_p\left(\frac{1}{T^{b/2}} \right) + O_p\left(\frac{1}{T^{\eta/2}} \right) + O_p\left(\frac{1}{T^{1/2}} \right), \end{aligned} \tag{B.14}$$

where again equation (e) of lemma A.5 and step (1) of appendix B.4 have been used together with $u_{\mathcal{T}} = O_p(1)$ (by (B.1)) in order to obtain the size of the remainder terms. Hence, it follows that (B.9) is asymptotically equal to

$$\begin{bmatrix} 1 \\ -\alpha \end{bmatrix} \frac{\kappa_x^{(2)}}{T^{b/2}} \sum_{t=1}^{T} z_t \varepsilon_t \xrightarrow{d} \mathcal{N}\left(0, \frac{\kappa_x^{(2)2} \gamma \sigma^4}{2cb} \begin{bmatrix} 1 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix} \right). \tag{B.15}$$

The limiting distribution is a direct consequence of lemma A.5. Using step (a) of this proof in conjunction with Slutzky's theorem gives the stated result. □

Remark B.1 *Recall*

$$\widehat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} \widehat{\varepsilon}_t^2 \quad \textit{and} \quad \widehat{\varepsilon}_t = y_t - \widehat{\lambda}' w_t,$$

and observe that

$$\widehat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2 - \frac{1}{T} \left(\frac{1}{T^{b/2}} \sum_{t=1}^{T} w_t \varepsilon_t \right)' (T^b M_T^{-1}) \left(\frac{1}{T^{b/2}} \sum_{t=1}^{T} w_t \varepsilon_t \right) = \sigma^2 + o_p(1),$$

using the previous results and applying the LLN to the first term. The consistency of $T^b V_T$ for V mentioned in proposition 2.1 follows therefore immediately from step (a) and (b) above. Similarly, the asymptotic normality of the t statistics for β and δ mentioned in section 2.3 can *be easily established with the help of the previous results be recognizing that*

$$T^b m^{11}\,\widehat{\sigma}^{\,2} = \frac{m_T \widehat{\sigma}^{\,2}}{\det M_T / T^{1+b}} = \frac{2cb}{\gamma} + o_p(1) \tag{B.16}$$

$$T^b m^{22}\,\widehat{\sigma}^{\,2} = \frac{1}{T}\sum_{t=1}^{T}(x_t a_{t-1})^2 \frac{\widehat{\sigma}^{\,2}}{\det M_T / T^{1+b}} = \alpha^2 \frac{2cb}{\gamma} + o_p(1). \tag{B.17}$$

B.2 Proofs of equation (2.12), corollary 2.1 and equation (2.17)

Proof of equation (2.12)**:** Equation (2.12) is restated here for convenience:

$$D_T^{-1/2} G' M_T G D_T^{-1/2} \xrightarrow{p} \begin{bmatrix} \sigma^2\gamma/(2cb) & 0 \\ 0 & \kappa_x^{(2)} \end{bmatrix}. \tag{B.18}$$

To see that this statement is true, observe that

$$D_T^{-1/2} G' M_T G D_T^{-1/2} = \begin{bmatrix} T^{-b}\sum_{t=1} z_t^2 & T^{-(1+b)/2}\sum_{t=1} z_t x_t \\ T^{-(1+b)/2}\sum_{t=1} z_t x_t & m_T \end{bmatrix}. \tag{B.19}$$

By parts (a) and (e) of lemma A.4, together with $m_T \to \kappa_x^{(2)}$ ($a.s.$), equation (2.12) follows. □

Proof of corollary 2.1: Note that

$$D_T^{1/2} G^{-1}(\widehat{\lambda} - \lambda) = D_T^{1/2} G^{-1} M_T^{-1} \sum_{t=1}^{T} w_t \varepsilon_t \tag{B.20}$$

$$= (D_T^{-1/2} G' M_T G D_T^{-1/2})^{-1} D_T^{-1/2} \sum_{t=1}^{T} (z_t\varepsilon_t,\, x_t\varepsilon_t)', \tag{B.21}$$

using $G' \sum_{t=1}^{T} w_t \varepsilon_t = \sum_{t=1}^{T}(z_t\varepsilon_t,\, x_t\varepsilon_t)'$. Hence,

$$D_T^{1/2} G^{-1}(\widehat{\lambda} - \lambda) = D_T^{-1/2} \sum_{t=1}^{T} \left(z_t\varepsilon_t/\omega_1,\, x_t\varepsilon_t/\omega_2\right)' + o_p(1), \tag{B.22}$$

where $\omega_1 = \sigma^2\gamma/(2cb)$ and $\omega_2 = \kappa_x^{(2)}$. Applying part ($d$) of lemma A.4 and the CLT for *i.i.d.* random variables yields the result. □

Proof of equation (2.17)**:** The second entry of the two-dimensional vector $D_T^{1/2} G^{-1}(\widehat{\lambda} - \lambda)$ is given by $T^{1/2} V^{\perp}(\widehat{\lambda} - \lambda)$, with $V^{\perp} = (\alpha,\, 1)$. Since $V^{\perp}\lambda = \alpha$ and $\delta - \alpha = -\alpha\beta$, it follows that

$$V^{\perp}(\widehat{\lambda} - \lambda) = \alpha\widehat{\lambda}_\beta + \widehat{\lambda}_\delta - \alpha \tag{B.23}$$

$$= \alpha\widehat{\lambda}_\beta - \alpha + \delta + \widehat{\lambda}_\delta - \delta \tag{B.24}$$

$$= \alpha(\widehat{\lambda}_\beta - \beta) + \widehat{\lambda}_\delta - \delta. \tag{B.25}$$

Similarly, one gets

$$\widehat{\lambda}_\alpha - \alpha = \frac{\alpha(\widehat{\lambda}_\beta - \beta) + \widehat{\lambda}_\delta - \delta}{1 - \widehat{\lambda}_\beta}. \tag{B.26}$$

Hence,

$$T^{1/2}(\widehat{\lambda}_\alpha - \alpha) = \frac{T^{1/2} V^\perp(\widehat{\lambda} - \lambda)}{1 - \widehat{\lambda}_\beta}, \tag{B.27}$$

so that the limiting result of equation (2.17) follows as a by-product from the preceding arguments. □

Now, taking (B.3) and (B.4) into account, it is seen that

$$\widehat{\lambda}_\beta - \beta = \frac{1}{\det M_T}\left(\sum_{t=1}^T x_t^2 \sum_{t=1}^T a_{t-1} x_t \varepsilon_t - \sum_{t=1}^T x_t^2 a_{t-1} \sum_{t=1}^T x_t \varepsilon_t\right) \tag{B.28}$$

$$\widehat{\lambda}_\delta - \delta = \frac{1}{\det M_T}\left(\sum_{t=1}^T (x_t a_{t-1})^2 \sum_{t=1}^T x_t \varepsilon_t - \sum_{t=1}^T x_t^2 a_{t-1} \sum_{t=1}^T a_{t-1} x_t \varepsilon_t\right). \tag{B.29}$$

Consequently, one gets from equations (B.5), (B.6) and a little rearrangement

$$\begin{aligned} T^{1/2}(\widehat{\lambda}_\alpha - \alpha) &= \frac{(\det M_T/T^{1+b})^{-1}}{1 - \widehat{\lambda}_\beta} \frac{1}{T^{1/2}} \sum_{t=1}^T a_{t-1} x_t \varepsilon_t \left(\frac{\alpha}{T^b} \sum_{t=1}^T x_t^2 - \frac{1}{T^b} \sum_{t=1}^T x_t^2 a_{t-1}\right) \\ &\quad + \frac{(\det M_T/T^{1+b})^{-1}}{1 - \widehat{\lambda}_\beta} \frac{1}{T^{1/2}} \sum_{t=1}^T x_t \varepsilon_t \left(\frac{1}{T^b} \sum_{t=1}^T (x_t a_{t-1})^2 - \frac{\alpha}{T^b} \sum_{t=1}^T x_t^2 a_{t-1}\right) \\ &= \frac{u_{\mathcal{T}}}{1 - \widehat{\lambda}_\beta}\left(\frac{\kappa_x^{(2)} (\det M_T/T^{1+b})^{-1}}{T^b} \sum_{t=1}^T z_t^2\right) \\ &\quad - \frac{(\det M_T/T^{1+b})^{-1}}{1 - \widehat{\lambda}_\beta}\left(\frac{1}{T^{(1+b)/2}} \sum_{t=1}^T x_t z_t\right)\left(\frac{1}{T^{b/2}} \sum_{t=1}^T z_t \varepsilon_t\right). \end{aligned} \tag{B.30}$$

Hence, by (B.27),

$$\begin{aligned} T^{-1/2} V^\perp(\widehat{\lambda} - \lambda) &= u_{\mathcal{T}} + u_{\mathcal{T}}\left(\frac{\kappa_x^{(2)} (\det M_T/T^{1+b})^{-1}}{T^b} \sum_{t=1}^T z_t^2 - 1\right) \\ &\quad - (\det M_T/T^{1+b})^{-1}\left(\frac{1}{T^{(1+b)/2}} \sum_{t=1}^T x_t z_t\right)\left(\frac{1}{T^{b/2}} \sum_{t=1}^T z_t \varepsilon_t\right) \\ &=: u_{\mathcal{T}} + A_T + B_T, \end{aligned} \tag{B.31}$$

say. By (B.1), $u_{\mathcal{T}}$ is asymptotically normal with mean zero and variance τ^2. Turning to A_T, note that the definition of $\det M_T$ in equation (B.7) implies

$$\frac{\kappa_x^{(2)}(\det M_T/T^{1+b})^{-1}}{T^b}\sum_{t=1}^T z_t^2 = \frac{\kappa_x^{(2)}/m_T}{1-\nu_T} \tag{B.32}$$

$$\nu_T = m_T^{-1}\left(T^{-(1+b)/2}\sum_{t=1}^T x_t z_t\right)^2\left(T^{-b}\sum_{t=1}^T z_t^2\right)^{-1}. \tag{B.33}$$

But, by lemma A.4, as $\nu_T = O_p(T^{-b})$. Thus, a Taylor-expansion yields

$$(1-\nu_T)^{-1} = 1+\nu_T + O_p(\nu_T^2) = 1 + O_p(T^{-b}). \tag{B.34}$$

Hence, as $m_T/\kappa_x^{(2)} = 1 + O_p(T^{-1/2})$ and $u_{\mathcal{T}} = O_p(1)$, it follows that $A_T = O_p(T^{-\min(1/2,b)})$. Using lemma A.4 again, it is seen that $B_T = O_p(T^{-b/2})$. Therefore,

$$T^{-1/2}V^{\perp}(\widehat{\lambda}-\lambda) = u_{\mathcal{T}} + O_p(T^{-b/2}) \xrightarrow{d} \mathcal{N}(0,\tau^2) \tag{B.35}$$

$$T^{-1/2}(\widehat{\lambda}_\alpha - \alpha) = c^{-1}u_{\mathcal{T}} + O_p(T^{-b/2}) \xrightarrow{d} \mathcal{N}(0,(\tau/c)^2), \tag{B.36}$$

where the final equation uses that $\widehat{\lambda}_\beta = \beta + O_p(T^{-b/2})$ (cf. proposition 2.1) and (B.27).

B.3 Proof of corollary 2.2

Proof of corollary (2.2)**:** For the sequence of local alternatives (2.32), one has

$$\begin{aligned} D_T^{1/2}G^{-1}(\widehat{\lambda}-\lambda_0) =& (D_T^{-1/2}G'M_TGD_T^{-1/2})^{-1}D_T^{-1/2}\sum_{t=1}^T(z_t\varepsilon_t,\, x_t\varepsilon_t)' \\ &+ T^{-1/2}D_T^{1/2}G^{-1}\mu. \end{aligned} \tag{B.37}$$

As $T^{-1/2}D_T^{1/2}G^{-1}\mu = (T^{-\eta/2},\, \alpha\mu_1+\mu_2)'$, it follows from corollary 2.1 that

$$D_T^{1/2}G^{-1}(\widehat{\lambda}-\lambda_0) \xrightarrow{d} \mathcal{N}\big((0,\,\alpha\mu_1+\mu_2)',\, V_1\big), \tag{B.38}$$

with V_1 defined in corollary 2.1. The consistency of $\widehat{\sigma}^2$ in conjunction with equations (2.12) and (2.30) yields

$$\mathcal{W} = (\widehat{\lambda}-\lambda_0)'G^{-1\prime}D_T^{1/2}V_1^{-1}D_T^{1/2}G^{-1}(\widehat{\lambda}-\lambda_0) + o_p(1), \tag{B.39}$$

and the result follows immediately. □

B.4 Proof of proposition 2.2

Proof of proposition 2.2: Recall that $z_t = x_t a_{t-1}^*$ and consider

$$T^{b/2}(\widehat{\beta}_0 - \beta) = \left(T^{-b}\sum_{t=1}^T z_t^2\right)^{-1} T^{-b/2}\sum_{t=1}^T z_t\varepsilon_t. \tag{B.40}$$

Equations (a) and (d) of lemma A.5 yield

$$\operatorname*{plim}_{T\to\infty} T^{-b}\sum_{t=1}^{T} z_t^2 = \lim_{T\to\infty} T^{-b}\sum_{t=1}^{T} E[z_t^2] = \sigma^2\gamma/(2cb) \tag{B.41}$$

$$T^{-b/2}\sum_{t=1}^{T} z_t\varepsilon_t \xrightarrow{d} \mathcal{N}\left(0, \sigma^4\gamma/(2cb)\right). \tag{B.42}$$

Combining equations (B.41) and (B.42), completes the proof by Slutzky's theorem. □

B.5 Proof of proposition 2.3

Proof of proposition 2.3: By equation (A.1),

$$a_t^* = a_{t-1}^* + \gamma_t(h(a_{t-1}^*) + u_t), \text{ with } h(a) \coloneqq -c(a-\alpha). \tag{B.43}$$

Hence, the actual law of motion given by equation (1.5) can be rewritten as

$$y_t = \beta a_{t-1}x_t + \delta x_t + \varepsilon_t = \alpha x_t + \omega\, x_t h(a_{t-1}) + \varepsilon_t, \tag{B.44}$$

where $\omega \coloneqq \beta/(\beta-1)$, and it has been used that $\delta = \alpha(1-\beta)$. In order to proceed further, it is helpful to introduce the following partial sum

$$H_{\mathcal{T}} \coloneqq \frac{1}{\sqrt{T}}\sum_{t=1}^{T} H(a_{t-1}, u_t), \tag{B.45}$$

where the map $a \mapsto H(a, u_t)$ is defined as

$$H(a, u_t) \coloneqq h(a)x_t^{*2} + u_t. \tag{B.46}$$

Making use of $\widehat{\alpha} = \sum_{t=1}^{T} x_t y_t / (\sum_{t=1}^{T} x_t^2)$ (cf. equation (2.42)) and the representation of y_t in equation (B.44), it follows that

$$\begin{aligned} T^{1/2}(\widehat{\alpha}-\alpha) &= (\kappa_x^{(2)}/m_{\mathcal{T}})\, T^{-1/2}\sum_{t=1}^{T}\kappa_x^{(2)-1}(y_t x_t - \alpha x_t^2) \\ &= (\kappa_x^{(2)}/m_{\mathcal{T}})\, T^{-1/2}\sum_{t=1}^{T}(\omega\, x_t^{*2}h(a_{t-1}) + u_t) = (\kappa_x^{(2)}/m_T)\left(\omega H_{\mathcal{T}} + c^{-1}u_{\mathcal{T}}\right), \end{aligned} \tag{B.47}$$

where $m_{\mathcal{T}} = T^{-1}\sum_{t=1}^{T} x_t$, $u_{\mathcal{T}} = T^{-1}\sum_{t=1}^{T} u_t$ (cf. equation (B.1)). By equation (B.1), the scaled average $u_{\mathcal{T}}$ is seen to converge in distribution to a mean-zero Gaussian random variable with variance $\tau^2 = \sigma^2/\kappa_x^{(2)}$ while $(\kappa_x^{(2)}/m_{\mathcal{T}}) = 1 + o_{a.s}(1)$. It thus remains to be shown that $H_{\mathcal{T}} = o_p(1)$. Before doing so, note that the individual summands of the scaled partial sum $H_{\mathcal{T}}$ can be re-written as

$$H(a_{t-1}, u_t) = (u_t - ca_{t-1}^*) - (c/\kappa_x^{(2)})v_t, \text{ with } v_t \coloneqq (x_t^2 - \kappa_x^{(2)})a_{t-1}^*. \tag{B.48}$$

It will be shown that $H_{\mathcal{T}} = o_p(1)$ by establishing that

(1) $T^{-1/2}\sum_{t=1}^{T}(u_t - ca^*_{t-1}) = o_p(1)$,

(2) $T^{-1/2}\sum_{t=1}^{T} v_t = o_p(1)$.

The latter has already been established in lemma A.3-(c). Hence, consider part (1). From the recursive representation of a^*_t given by equation (A.4), one gets

$$\sum_{t=1}^{T} a^*_{t-1} = \sum_{t=1}^{T} \xi_t + a^*_0(1 + \sum_{t=1}^{T} \vartheta_t) - a^*_T. \tag{B.49}$$

Hence, with $\xi_{\mathcal{T}} := T^{-1/2}\sum_{t=1}^{T}\xi_t$, it follows that

$$T^{-1/2}\sum_{t=1}^{T}(u_t - ca^*_{t-1}) = u_{\mathcal{T}} - c\xi_{\mathcal{T}} + T^{-1/2}a^*_0(1 + \sum_{t=1}^{T}\vartheta_t) - T^{-1/2}a^*_T. \tag{B.50}$$

It is readily verified that both, $T^{-1/2}a^*_0(1+\sum_{t=1}^{T}\vartheta_t)$ and $T^{-1/2}a^*_T$ are $o_p(1)$. In order to see that $u_{\mathcal{T}} - c\xi_{\mathcal{T}} = o_p(1)$, note first that $E[u_{\mathcal{T}}] = E[\xi_{\mathcal{T}}] = 0$. By Chebychev's inequality, the claim follows because of

$$\mathrm{var}[u_{\mathcal{T}} - c\xi_{\mathcal{T}}] = E[u^2_{\mathcal{T}}] + c^2E[\xi^2_{\mathcal{T}}] - 2cE[u_{\mathcal{T}}\xi_{\mathcal{T}}] = o(1). \tag{B.51}$$

To see that the preceding conclusion is true, suppose first that part (1) of assumption (B2) holds true. Equations (S.103) and (S.104) of the supplementary material imply

$$E[\xi_t\xi_s] = \begin{cases} \tau^2\sum_{k=1}^{s} g_{k,t}g_{k,s} + O(\gamma_s^{2-1/\eta}g_{s,t}) & \text{if } s < t \\ \tau^2\phi^{ii}_t + O(\gamma_t^{2-1/\eta}) & \text{if } s = t, \end{cases} \tag{B.52}$$

where $g_{k,t} = \gamma_k\Phi_{k,t+1}$ (cf. equation (A.6)) and $\phi^{ii}_t = \sum_{k=1}^{t} g^2_{k,t}$ (cf. lemma A.1-(a)). Hence,

$$T^{-1}\sum_{t=1}^{T} E[\xi^2_t] = \tau^2\bar{\phi}^{ii} + T^{-1}\sum_{t=1}^{T} O(\gamma_t^{2-1/\eta}) = o(1) \tag{B.53}$$

where $\bar{\phi}^{ii} = T^{-1}\sum_{t=1}^{T}\phi^{ii}_t = o(1)$ (cf. remark A.2), and, by equation (A.9) and assumption A, $T^{-1}\sum_{t=1}^{T} O(\gamma_t^{2-1/\eta}) = O(T^{1-2\eta}) = o(1)$. Furthermore,

$$T^{-1}\sum_{t=2}^{T}\sum_{s=1}^{t-1} E[\xi_t\xi_s] = \tau^2\bar{\phi}^{iii} + T^{-1}\sum_{t=2}^{T}\sum_{s=1}^{t-1} O(\gamma_s^{2-1/\eta}g_{s,t}), \tag{B.54}$$

where $\bar{\phi}^{iii} = T^{-1}\sum_{t=1}^{T}\phi^{iii}_t = 1/(2c^2) + o(1)$, with $\phi^{iii}_t = \sum_{s=1}^{t-1}\sum_{k=1}^{s} g_{k,t}g_{k,s}$. By lemma A.1-($d$), the second term of equation (B.54) can be rewritten as

$$T^{-1}\sum_{t=2}^{T}\sum_{s=1}^{t-1} O(\gamma_s^{2-1/\eta}g_{s,t}) = T^{-1}\sum_{t=2}^{T} O(\gamma_t^{2-1/\eta}) = O(T^{1-2\eta}) = o(1), \tag{B.55}$$

using assumption A. Putting the above together yields

$$E[\xi_{\mathcal{T}}^2] = T^{-1}(\sum_{t=1}^{T} E[\xi_t^2] + 2\sum_{t=2}^{T}\sum_{s=1}^{t-1} E[\xi_t \xi_s]) = 2\tau^2 \bar{\phi}^{iii} + o(1). \tag{B.56}$$

Next, by lemma A.1-(a),

$$E[u_{\mathcal{T}}\xi_{\mathcal{T}}] = T^{-1}\sum_{s,t=1}^{T} E[\xi_t u_s] = T^{-1}\tau^2 \sum_{t=1}^{T}\sum_{s=1}^{t} g_{s,t} = \tau^2 \bar{\phi}^i, \tag{B.57}$$

where $\bar{\phi}^i = T^{-1}\sum_{t=1}^{T}\phi_t^i = 1/c + o(1)$, with $\phi_t^i = \sum_{k=1}^{t} g_{k,t}$; cf. remark A.2. To see this, note that by assumption B, for $1 \le s \le t$,

$$\begin{aligned} E[\xi_t u_s] =& \sum_{k=1}^{t} \gamma_k E[\vartheta_{t,k+1} u_k u_s] = \kappa_x^{(2)-1} \sum_{k=1}^{t} \gamma_k E[\vartheta_{t,k+1} x_k^* x_s^*] E[\varepsilon_k \varepsilon_s] \\ =& \tau^2 \gamma_s E[\vartheta_{t,s+1} x_s^{*2}] = \tau^2 g_{s,t}, \end{aligned} \tag{B.58}$$

otherwise $E[\xi_t u_s] = 0$. Hence, by combining the results and using lemma A.1,

$$\text{var}[u_{\mathcal{T}} - c\xi_{\mathcal{T}}] = \tau^2(1 + c^2 2\bar{\phi}^{iii} - 2c\bar{\phi}^i) + o(1) = o(1), \tag{B.59}$$

thereby showing that $T^{-1/2}\sum_{t=1}^{T}(u_t - ca_{t-1}^*) = o_p(1)$, which (1) (and therefore the proposition) given that part (1) of assumption (B2) holds true. Similarly, if part (2) of assumption (B2) is satisfied, it follows from the discussion following equation (A.186) and lemma A.1 that

$$E[\xi_{\mathcal{T}}^2] = T^{-1}(\sum_{t=1}^{T} E[\xi_t^2] + 2\sum_{t=2}^{T}\sum_{s=1}^{t-1} E[\xi_t \xi_s]) = \tau^2(1 + c^2 2\bar{\phi}^{iii} - 2c\bar{\phi}^i) + o(1), \tag{B.60}$$

which is, as discussed above, $o(1)$. This completes the proof. □

Remark B.2 *Proof of equation* (2.44)*. Note*

$$\begin{aligned} \tilde{x}_t =& x_t(b_{t-1} - \widehat{\alpha}) = z_t - (\widehat{\alpha} - \alpha)x_t \\ \tilde{y}_t =& y_t - \widehat{\alpha}x_t = \tilde{y}_{0t} - (\widehat{\alpha} - \alpha)x_t, \end{aligned} \tag{B.61, B.62}$$

with $z_t = x_t a^*_{t-1}$ *and* $\tilde{y}_{0t} = y_t - \alpha x_t$*, so that*

$$\begin{aligned}\frac{1}{T^b}\sum_{t=1}^{T}\tilde{x}_t^2 &= \frac{1}{T^b}\sum_{t=1}^{T} z_t^2 + (\widehat{\alpha}-\alpha)^2\frac{1}{T^b}\sum_{t=1}^{T} x_t^2 - 2(\widehat{\alpha}-\alpha)\frac{1}{T^b}\sum_{t=1}^{T} x_t^2 a^*_{t-1} \\ &= \frac{1}{T^b}\sum_{t=1}^{T} z_t^2 + O_p\left(\frac{1}{T^b}\right) \qquad \text{(B.63)}\\ \frac{1}{T^{b/2}}\sum_{t=1}^{T}\tilde{y}_t\tilde{x}_t &= \frac{1}{T^{b/2}}\sum_{t=1}^{T}\tilde{y}_{0t}z_t - \frac{(\widehat{\alpha}-\alpha)(\beta+1)}{T^{b/2}}\sum_{t=1}^{T} x_t^2 a^*_{t-1} \\ &\quad - \frac{\widehat{\alpha}-\alpha}{T^{b/2}}\sum_{t=1}^{T} x_t\varepsilon_t + \frac{(\widehat{\alpha}-\alpha)^2}{T^{b/2}}\sum_{t=1}^{T} x_t^2 \\ &= \frac{1}{T^{b/2}}\sum_{t=1}^{T}\tilde{y}_{0t}z_t + O_p\left(\frac{1}{T^{b/2}}\right). \qquad \text{(B.64)}\end{aligned}$$

using (*e*) *and proposition 2.3. Therefore, by Slutzky's theorem*

$$T^{b/2}(\widehat{\beta}-\beta) = T^{b/2}(\widehat{\beta}_0-\beta) + O_p\left(\frac{1}{T^{b/2}}\right). \qquad \text{(B.65)}$$

Proof of equation (2.46). *Using proposition 2.2 and 2.3, it follows*

$$\begin{aligned}T^{b/2}(\widehat{\delta}-\delta) =& T^{b/2}(\alpha(1-\widehat{\beta})-\delta) + T^{b/2}(\widehat{\alpha}-\alpha)(1-\widehat{\beta}) \\ =& -\alpha T^{b/2}(\widehat{\beta}-\beta) + O_p\left(T^{-\frac{\eta}{2}}\right). \qquad \text{(B.66)}\end{aligned}$$

S Supplementary material

S.1 Multiple regressors

Note. This appendix extends the results collected in section 2.2 to the multivariate case where $\boldsymbol{x}_t$ represents a k-dimensional column vector of regressors. As mentioned in the main text, the agent's multivariate recursion (2.21) for the k-dimensional RE equilibrium vector α is given by

$$\boldsymbol{a}_t = \boldsymbol{a}_{t-1} + \gamma_t E[\boldsymbol{x}_t\boldsymbol{x}_t']^{-1}\boldsymbol{x}_t(y_t - \boldsymbol{a}_{t-1}'\boldsymbol{x}_t), \qquad \text{(S.1)}$$

where, similar to assumption B-B1, the assumption $\mathbf{R}_t = E[\boldsymbol{x}_t\boldsymbol{x}_t']$ (a positive definite $k \times k$ matrix) has been made. Hence, $y_t^e = \boldsymbol{a}_{t-1}'\boldsymbol{x}_t$ represents the agent's forecast of y_t.

Assumption M.1 *The random vector* $\boldsymbol{v}_t := (\boldsymbol{x}_t', \varepsilon_t)'$ *is independent, identically distributed and* $\boldsymbol{x}_t'$ *is independent of* ε_t*. Furthermore,* $\kappa_\varepsilon^{(4)} < \infty$, $E\left[(\boldsymbol{a}_0'\boldsymbol{a}_0)^2\right] < \infty$ *and* $\boldsymbol{a}_0$ *is independent of* $\boldsymbol{v}_t$*, for all* t*.*

Assumption M.2 $E[\boldsymbol{x}_1\boldsymbol{x}_1']$ *is a positive definite matrix. The eigenvalues of* $E[(\boldsymbol{x}_1\boldsymbol{x}_1')^i]$ ($i = 2, 3, 4$) *and* $E[(\boldsymbol{x}_1\boldsymbol{x}_1' \otimes \boldsymbol{x}_1\boldsymbol{x}_1')^i]$ ($i = 1, 2$) *take their values in* $[0, K]$ *for some* $K < \infty$.[16]

[16] The notation $\mathbf{A}^i$ is used to denote the i^{th} matrix product of any symmetric matrix $\mathbf{A}$ for any integer i. For negative i, $\mathbf{A}^i$ is the i^{th} matrix product of the inverted matrix $\mathbf{A}$, assuming this inverse exists.

The specification of the gain (assumption (A)) remains the same. It will be shown that the $(k+1)$-dimensional OLS estimator $\widehat{\boldsymbol{\lambda}}$ based on

$$y_t = \beta \boldsymbol{a}_{t-1}' \boldsymbol{x}_t + \boldsymbol{\delta}' \boldsymbol{x}_t + \varepsilon_t, \quad \text{for} \quad t = 1, 2, \ldots, T, \tag{S.2}$$

where $|\beta| < 1$ and $\boldsymbol{\delta} \in \mathbb{R}^k$, has the following properties:

Proposition 3.4 *Suppose assumptions* A, M.1 *and* M.2 *hold. Then,*

$$T^{b/2}(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) \xrightarrow{d} \mathcal{N}(\mathbf{0}_{k+1}, \mathbf{V}), \quad \textit{with } \mathbf{V} := \frac{2cb}{k\gamma} \begin{bmatrix} 1 & -\boldsymbol{\alpha}' \\ -\boldsymbol{\alpha} & \boldsymbol{\alpha}\boldsymbol{\alpha}' \end{bmatrix} \tag{a}$$

$$\mathbf{D}_T^{1/2} \mathbf{G}^{-1}(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}_{k+1}, \mathbf{V}_1\right), \quad \textit{with } \mathbf{V}_1 := \begin{bmatrix} 2cb/(k\gamma) & \mathbf{0}_k' \\ \mathbf{0}_k & \sigma^2 \mathbf{Q}^{-1} \end{bmatrix}. \tag{b}$$

Proof of proposition 3.4: To begin with, multivariate equivalents of lemma (A.3) and A.4 are stated. The proofs of lemma S.1 and S.2 are delegated to the end of this section.

Lemma S.1 *Suppose assumptions 2.1, M.1 and M.2 hold and define $\boldsymbol{a}_t^* := \boldsymbol{a}_t - \boldsymbol{\alpha}$, $\boldsymbol{Q} := E[\boldsymbol{x}_1 \boldsymbol{x}_1']$. Then,*

$$E[\boldsymbol{a}_t^{*\prime} \mathbf{Q} \boldsymbol{a}_t^*] = O(\gamma_t) \tag{a}$$

$$E[(\boldsymbol{a}_t^{*\prime} \mathbf{Q} \boldsymbol{a}_t^*)^2] = O(\gamma_t^2), \tag{b}$$

and

$$\lim_{T \to \infty} T^{-b} \sum_{t=1}^{T} E[\boldsymbol{a}_t^{*\prime} \mathbf{Q} \boldsymbol{a}_t^*] = k\sigma^2 \gamma/(2cb). \tag{c}$$

Lemma S.2 *Let $z_t := \boldsymbol{a}_{t-1}^{*}{}' \boldsymbol{x}_t$. Suppose assumptions* 2.1, M.1 *and* M.2 *hold. Then,*

$$T^{-b} \sum_{t=1}^{T} z_t^2 \xrightarrow{p} k\sigma^2 \frac{\gamma}{2cb} \tag{a}$$

$$T^{-(1+b)/2} \sum_{t=1}^{T} \boldsymbol{x}_t z_t = O_p(T^{-b/2}) \tag{b}$$

$$T^{-b/2} \sum_{t=1}^{T} z_t \varepsilon_t \xrightarrow{d} \mathcal{N}(0, k\sigma^4 \frac{\gamma}{2cb}). \tag{c}$$

In order to prove proposition 3.4-(a), recall that $\widehat{\boldsymbol{\lambda}}$ denotes the $(k+1)$-dimensional OLS estimator of $\boldsymbol{\lambda} = (\beta, \boldsymbol{\delta}')'$ based on

$$y_t = \boldsymbol{\lambda}' \boldsymbol{w}_t + \varepsilon_t, \text{ for } t = 1, 2, \ldots, T, \tag{S.3}$$

where $\boldsymbol{w}_t := (y_t^e, \boldsymbol{x}_t')'$ and $y_t^e = \boldsymbol{a}_{t-1}' \boldsymbol{x}_t$. Specifically, upon defining the second moment matrix

$$\mathbf{M}_T := \sum_{t=1}^{T} \boldsymbol{w}_t \boldsymbol{w}_t' = \mathbf{W}'\mathbf{W}, \tag{S.4}$$

with $\mathbf{W} := (\boldsymbol{w}_1', \ldots, \boldsymbol{w}_T')'$ (a $T \times (k+1)$ matrix), the OLS estimator can be rewritten as

$$\widehat{\boldsymbol{\lambda}} = \mathbf{M}_T^{-1}\mathbf{W}'\boldsymbol{y}, \tag{S.5}$$

where $\boldsymbol{y} := (y_1, \ldots, y_T)'$. In a first step, it will be shown that

$$T^b\mathbf{M}_T^{-1} = \frac{2cb}{k\gamma\sigma^2}\begin{bmatrix} 1 & -\boldsymbol{\alpha}' \\ -\boldsymbol{\alpha} & \boldsymbol{\alpha}\boldsymbol{\alpha}' \end{bmatrix} + o_p(1). \tag{S.6}$$

Note, that

$$\mathbf{M}_T^{-1} = \frac{1}{\zeta_T}\begin{bmatrix} 1 & -\boldsymbol{b}_T' \\ -\boldsymbol{b}_T & \mathbf{B}_T \end{bmatrix}, \tag{S.7}$$

where

$$\boldsymbol{b}_T := \mathbf{Q}_T^{-1}\boldsymbol{q}_T,\ \ \boldsymbol{q}_T := \mathbf{X}'\boldsymbol{y}^e,\ \ \mathbf{Q}_T := \mathbf{X}'\mathbf{X}, \tag{S.8}$$

with $\mathbf{X} := (\boldsymbol{x}_1', \ldots, \boldsymbol{x}_T')'$ (a $T \times k$ matrix), $\boldsymbol{y}^e := (y_1^e, \ldots, y_T^e)'$ (a T dimensional vector); and

$$\mathbf{B}_T := \mathbf{Q}_T^{-1}(\mathbf{I}_k\zeta_T + \boldsymbol{q}_T\boldsymbol{q}_T'\mathbf{Q}_T^{-1}),\ \text{with}\ \zeta_T := \boldsymbol{y}^{e\prime}\boldsymbol{y}^e - \boldsymbol{q}_T'\mathbf{Q}_T^{-1}\boldsymbol{q}_T'; \tag{S.9}$$

see, e.g., Abadir and Magnus (2006, exercise 5.15). It follows readily from the (multivariate) weak LLN for *i.i.d.* random vectors applied to $T^{-1}\mathbf{Q}_T$ and lemma S.2 that

$$T^{-1}\mathbf{Q}_T = \mathbf{Q} + o_p(1) \tag{S.10}$$

$$T^{-1}\boldsymbol{q}_T = T^{-1}\mathbf{Q}_T\boldsymbol{\alpha} + T^{-1}\boldsymbol{\ell}_T = \mathbf{Q}\boldsymbol{\alpha} + o_p(1), \tag{S.11}$$

where $\boldsymbol{\ell}_T := \mathbf{X}'\boldsymbol{z}$, $\boldsymbol{z} := (z_1, \ldots, z_T)'$ (a T dimensional vector). Similarly,

$$T^{-b}\zeta_T = T^{-b}\boldsymbol{z}'\boldsymbol{z} - T^{-(1+b)/2}\boldsymbol{\ell}_T'(T^{-1}\mathbf{Q}_T)^{-1}\boldsymbol{\ell}_T T^{-(1+b)/2} = k\sigma^2\frac{\gamma}{2cb} + o_p(1) \tag{S.12}$$

$$\mathbf{B}_T = (T^{-1}\mathbf{Q}_T)^{-1}(\mathbf{I}_k(T^{-1}\zeta_T) + T^{-2}\boldsymbol{q}_T\boldsymbol{q}_T'(T^{-1}\mathbf{Q}_T)^{-1}) = \boldsymbol{\alpha}\boldsymbol{\alpha}' + o_p(1); \tag{S.13}$$

it has repeatedly been used that $\mathbf{X}'\boldsymbol{y}^e = \boldsymbol{\ell}_T + \mathbf{Q}_T\boldsymbol{\alpha}$. The result in equation (S.6) follows directly. Now, taking (S.7) into account, one gets

$$\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda} = \mathbf{M}_T^{-1}\mathbf{W}'\boldsymbol{\varepsilon} = \zeta_T^{-1}((1, -\boldsymbol{b}_T')'\boldsymbol{y}^{e\prime}\varepsilon + (-\boldsymbol{b}_T, \mathbf{B}_T)'\boldsymbol{u}_T), \tag{S.14}$$

where $(-\boldsymbol{b}_T, \mathbf{B}_T)$ is a $k \times (k+1)$ matrix, $\boldsymbol{u}_T := \mathbf{X}'\boldsymbol{\varepsilon}$ with $\boldsymbol{\varepsilon} := (\varepsilon_1, \ldots, \varepsilon_T)'$. Next, use $\boldsymbol{y}^{e\prime}\boldsymbol{\varepsilon} = \boldsymbol{z}'\boldsymbol{\varepsilon} + \boldsymbol{\alpha}'\mathbf{X}'\boldsymbol{\varepsilon}$ to rewrite the preceding display as

$$\mathbf{M}_T^{-1}\mathbf{W}'\boldsymbol{\varepsilon} = \zeta_T^{-1}((1, -\boldsymbol{b}_T')'\tilde{u}_T + (1, -\boldsymbol{b}_T')'\boldsymbol{\alpha}'\boldsymbol{u}_T + (-\boldsymbol{b}_T, \mathbf{B}_T)'\boldsymbol{u}_T) =: \zeta_T^{-1}(A_1 + A_2 + A_3),$$

with $\tilde{u}_T := \boldsymbol{z}'\boldsymbol{\varepsilon}$, say. Using that

$$\boldsymbol{b}_T = \boldsymbol{\alpha} + \mathbf{Q}_T^{-1}\boldsymbol{\ell}_T \tag{S.15}$$

$$\boldsymbol{q}_T = \boldsymbol{\ell}_T + \mathbf{Q}_T\boldsymbol{\alpha} \tag{S.16}$$

$$\mathbf{B}_T = \mathbf{Q}_T^{-1}\zeta_T + \boldsymbol{\alpha}\boldsymbol{\alpha}' + \boldsymbol{\alpha}\boldsymbol{\ell}_T'\mathbf{Q}_T^{-1} + \mathbf{Q}_T^{-1}\boldsymbol{\ell}_T\boldsymbol{\alpha}' + \mathbf{Q}_T^{-1}\boldsymbol{\ell}_T\boldsymbol{\ell}_T'\mathbf{Q}_T^{-1}, \tag{S.17}$$

A_1, A_2 and A_3 can be rewritten as

$$\begin{aligned} A_1 &= (1, -\boldsymbol{\alpha}')'\tilde{u}_T - (\mathbf{0}_k, \mathbf{I}_k)'\mathbf{Q}_T^{-1}\boldsymbol{\ell}_T\tilde{u}_T \\ A_2 &= (1, -\boldsymbol{\alpha}')'\boldsymbol{\alpha}'\boldsymbol{u}_T - (\mathbf{0}_k, \mathbf{I}_k)'\mathbf{Q}_T^{-1}\boldsymbol{\ell}_T\boldsymbol{\alpha}'\boldsymbol{u}_T \\ A_3 &= (-\boldsymbol{b}_T, \mathbf{O})'\boldsymbol{u}_T + (\mathbf{0}_k, \mathbf{I}_k)'(\mathbf{Q}_T^{-1}\zeta_T + \mathbf{Q}_T^{-1}\boldsymbol{\ell}_T\boldsymbol{\ell}_T'\mathbf{Q}_T^{-1})\boldsymbol{u}_T \\ &\qquad + (\mathbf{0}_k, \mathbf{I}_k)'(\boldsymbol{\alpha}\boldsymbol{\alpha}' + \boldsymbol{\alpha}\boldsymbol{\ell}_T'\mathbf{Q}_T^{-1} + \mathbf{Q}_T^{-1}\boldsymbol{\ell}_T\boldsymbol{\alpha}')\boldsymbol{u}_T, \end{aligned} \tag{S.18}$$

where $\mathbf{O}$ is a $k \times k$ matrix of zeros. Furthermore,

$$\begin{aligned} A_2 - &(\boldsymbol{b}_T, \mathbf{O})'\boldsymbol{u}_T + (\mathbf{0}_k, \mathbf{I}_k)' \\ &\times (\boldsymbol{\alpha}\boldsymbol{\alpha}' + \boldsymbol{\alpha}\boldsymbol{\ell}_T'\mathbf{Q}_T^{-1} + \mathbf{Q}_T^{-1}\boldsymbol{\ell}_T\boldsymbol{\alpha}')\boldsymbol{u}_T = -(1, -\boldsymbol{\alpha}')'\boldsymbol{\ell}_T'\mathbf{Q}_T^{-1}\boldsymbol{u}_T. \end{aligned} \tag{S.19}$$

To see that this is true, note that the left-hand side of the preceding display equals

$$((1, -\boldsymbol{\alpha}')'\boldsymbol{\alpha}'\boldsymbol{u}_T - (\boldsymbol{b}_T, \mathbf{O})'\boldsymbol{u}_T) + (\mathbf{0}_k, \mathbf{I}_k)'(\boldsymbol{\alpha}\boldsymbol{\alpha}' + \boldsymbol{\alpha}\boldsymbol{\ell}_T'\mathbf{Q}_T^{-1})\boldsymbol{u}_T. \tag{S.20}$$

Note that the second summand is a $(k+1)$-dimensional vector with first entry zero and with remaining elements given by $(\boldsymbol{\alpha}\boldsymbol{\alpha}' + \boldsymbol{\alpha}\boldsymbol{\ell}_T'\mathbf{Q}_T^{-1})\boldsymbol{u}_T$. On the other hand, the first term can be rewritten as

$$(1, -\boldsymbol{\alpha}')'\boldsymbol{\alpha}'\boldsymbol{u}_T - (\boldsymbol{b}_T, \mathbf{O})'\boldsymbol{u}_T = -(\mathbf{Q}_T^{-1}\boldsymbol{\ell}_T, \boldsymbol{\alpha}\boldsymbol{\alpha}')'\boldsymbol{u}_T, \tag{S.21}$$

which is a $(k+1)$-dimensional vector whose first entry is $-\boldsymbol{\ell}_T'\mathbf{Q}_T^{-1}\boldsymbol{u}_T$ and the remaining k elements are given by $-\boldsymbol{\alpha}\boldsymbol{\alpha}'\boldsymbol{u}_T$; thereby proving equation (S.19) follows. Hence,

$$\begin{aligned} A_1 + A_2 + A_3 &= A_1 - (1, -\boldsymbol{\alpha}')'\boldsymbol{\ell}_T'\mathbf{Q}_T^{-1}\boldsymbol{u}_T \\ &\qquad + (\mathbf{0}_k, \mathbf{I}_k)'(\mathbf{Q}_T^{-1}\zeta_T + \mathbf{Q}_T^{-1}\boldsymbol{\ell}_T\boldsymbol{\ell}_T'\mathbf{Q}_T^{-1})\boldsymbol{u}_T. \end{aligned} \tag{S.22}$$

It follows from the above and lemma S.2 that

$$\begin{aligned} T^{b/2}(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) &= (T^{-b}\zeta_T)^{-1}(1, -\boldsymbol{\alpha}')'T^{-b/2}\tilde{u}_T \\ &\quad - (T^{-b}\zeta_T)^{-1}(1, -\boldsymbol{\alpha}')'T^{-(1+b)/2}\boldsymbol{\ell}_T'(T^{-1}\mathbf{Q}_T)^{-1}T^{-1/2}\boldsymbol{u}_T \\ &\quad + (T^{-b}\zeta_T)^{-1}(\mathbf{0}_k, \mathbf{I}_k)'\big[(T^{-1}\mathbf{Q}_T)^{-1}T^{(1+b)/2}\zeta_T \\ &\qquad\qquad + (T^{-1}\mathbf{Q}_T)^{-1}(T^{-(1+b)/2}\boldsymbol{\ell}_T)(T^{-1}\boldsymbol{\ell}_T)' \\ &\qquad\qquad \times (T^{-1}\mathbf{Q}_T)^{-1}\big]T^{1/2}\boldsymbol{u}_T \\ &\quad - (T^{-b}\zeta_T)^{-1}\mathbf{0}_k, \mathbf{I}_k)'(T^{-1}\mathbf{Q}_T)^{-1}T^{-(1+b)/2}\boldsymbol{\ell}_T T^{-1/2}\tilde{u}_T \\ &= (T^{-b}\zeta_T)^{-1}(1, -\boldsymbol{\alpha}')'T^{-b/2}\tilde{u}_T + o_p(1) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}_{k+1}, \mathbf{V}\right), \end{aligned} \tag{S.23}$$

thereby proving part (a) of proposition 3.4. Turning to part (b), consider

$$\mathbf{D}_T^{1/2}\mathbf{G}^{-1}(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) = (\mathbf{D}_T^{-1/2}\mathbf{G}'\mathbf{M}_T\mathbf{G}\mathbf{D}_T^{-1/2})^{-1}\mathbf{D}_T^{-1/2}\mathbf{G}'\mathbf{W}'\boldsymbol{\varepsilon}. \tag{S.24}$$

By lemma S.2, the LLN (applied to $T^{-1}\mathbf{Q}_T$) and the CLT for *i.i.d.* random vectors (applied to $T^{-1/2}\boldsymbol{u}_T$), one gets

$$\mathbf{D}_T^{-1/2}\mathbf{G}'\mathbf{M}_T\mathbf{G}\mathbf{D}_T^{-1/2} = \begin{bmatrix} T^{-b}\boldsymbol{z}'\boldsymbol{z} & T^{-(1+b)/2}\boldsymbol{\ell}_T' \\ T^{-(1+b)/2}\boldsymbol{\ell}_T & T^{-1}\mathbf{Q}_T \end{bmatrix} \xrightarrow{p} \begin{bmatrix} k\sigma^2\gamma/(2cb) & \mathbf{0}_k' \\ \mathbf{0}_k & \mathbf{Q} \end{bmatrix} \tag{S.25}$$

$$\mathbf{D}_T^{-1/2}\mathbf{G}'\mathbf{W}'\boldsymbol{\varepsilon} = \begin{bmatrix} T^{-b/2}\tilde{u}_T \\ T^{-1/2}\boldsymbol{u}_T \end{bmatrix} \xrightarrow{d} N\left(\mathbf{0}_{k+1}, \begin{bmatrix} k\sigma^4\gamma/(2cb) & \mathbf{0}_k' \\ \mathbf{0}_k & \sigma^2\mathbf{Q} \end{bmatrix}\right). \tag{S.26}$$

The claim now follows from Slutzky's lemma. □

Proof of lemma S.1 and S.2

Note. For notational ease, set $\boldsymbol{a}_t := \boldsymbol{a}_t^*$ during the proof. Throughout, $\otimes$ denotes the Kronecker product; $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$ represent respectively the largest and smallest eigenvalue of any symmetric, real matrix $\mathbf{A}$. The following will be frequently used: (i) Let $\mathbf{A}$ and $\mathbf{B}$ represent $m \times m$ matrices, then, for any conformable vector $\boldsymbol{x}$

$$(\boldsymbol{x}'\mathbf{A}\boldsymbol{x})(\boldsymbol{x}'\mathbf{B}\boldsymbol{x}) = (\boldsymbol{x}' \otimes \boldsymbol{x}')(\mathbf{A} \otimes \mathbf{B})(\boldsymbol{x} \otimes \boldsymbol{x}). \tag{S.27}$$

(ii) The eigenvalues of $\mathbf{A} \otimes \mathbf{B}$ ($\mathbf{A}$ and $\mathbf{B}$, not necessarily symmetric) are equal to the product of the respective eigenvalues; see, for example, Abadir and Magnus (2006, exercise 10.10). (iii) If $\mathbf{A}$ is symmetric, then

$$\lambda_{\min}(\mathbf{A}) \leq \frac{\boldsymbol{x}'\mathbf{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} \leq \lambda_{\max}(\mathbf{A}); \tag{S.28}$$

see, for example, Abadir and Magnus (2006, exercise 7.53).

Proof of lemma S.1-(a): Taking equations (S.1) and (S.2) into account, $\boldsymbol{a}_t$ can be rewritten as $\boldsymbol{a}_t = \boldsymbol{a}_{t-1} + \gamma_t\boldsymbol{h}_t$, with $\boldsymbol{h}_t := \mathbf{Q}^{-1}(\boldsymbol{x}_t\varepsilon_t - c\boldsymbol{x}_t\boldsymbol{x}_t'\boldsymbol{a}_{t-1})$. Therefore,

$$\boldsymbol{a}_t'\mathbf{Q}\boldsymbol{a}_t = \boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1} + 2\gamma_t\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{h}_t + \gamma_t^2\boldsymbol{h}_t'\mathbf{Q}\boldsymbol{h}_t. \tag{S.29}$$

Taking conditional expectations yields:

$$\begin{aligned} E[\boldsymbol{a}_t'\mathbf{Q}\boldsymbol{a}_t|\mathcal{A}_{t-1}] = &(1 - 2c\gamma_t)\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1} + \gamma_t^2\sigma^2 E[\boldsymbol{x}_1'\mathbf{Q}^{-1}\boldsymbol{x}_1] \\ &+ (c\gamma_t)^2\boldsymbol{a}_{t-1}'E[\boldsymbol{x}_1\boldsymbol{x}_1'\mathbf{Q}^{-1}\boldsymbol{x}\boldsymbol{x}_1']\boldsymbol{a}_{t-1}. \end{aligned} \tag{S.30}$$

Repeated application of equation (S.28) shows that for any conformable vector $\boldsymbol{a}$,

$$\boldsymbol{a}'E[\boldsymbol{x}_1\boldsymbol{x}_1'\mathbf{Q}^{-1}\boldsymbol{x}\boldsymbol{x}_1']\boldsymbol{a} \leq \mu_4\boldsymbol{a}'\mathbf{Q}\boldsymbol{a}, \tag{S.31}$$

where, by assumption M.2,

$$\mu_4 := \lambda_{\max}(E[(\boldsymbol{x}_1\boldsymbol{x}_1')^2])/\lambda_{\min}(\mathbf{Q})^2 \tag{S.32}$$

is a non-negative and finite-valued constant. Therefore,

$$E[\boldsymbol{a}_t'\mathbf{Q}\boldsymbol{a}_t|\mathcal{A}_{t-1}] \leq (1-2c\gamma_t)\boldsymbol{a}_{t-1}'\boldsymbol{a}_{t-1} + k\gamma_t^2\sigma^2 + (c\gamma_t)^2\mu_4\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1}. \tag{S.33}$$

Hence, there exists an integer t_0 such that $\gamma_t\mu_4 \leq 1/c$ for all $t \geq t_0$. This, in turn, implies for t sufficiently large:

$$E[\boldsymbol{a}_t'\mathbf{Q}\boldsymbol{a}_t] \leq (1-c\gamma_t)E[\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1}] + k\sigma^2\gamma_t^2 = \Phi_t E[\boldsymbol{a}_0'\mathbf{Q}\boldsymbol{a}_0] + k\sigma^2\sum_{k=1}^{t} g_{k,t}^2, \tag{S.34}$$

which, by equation (A.13) and lemma A.1-(d), is of order $O(\gamma_t)$. □

Proof of lemma S.1-(b): Consider

$$\begin{aligned}(\boldsymbol{a}_t'\mathbf{Q}\boldsymbol{a}_t)^2 - (\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1})^2 &= 4\gamma_t\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1}\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{h}_t \\ &\quad + 4\gamma_t^2(\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{h}_t)^2 \\ &\quad + \gamma_t^4(\boldsymbol{h}_t'\mathbf{Q}\boldsymbol{h}_t)^2 \\ &\quad + 2\gamma_t^2\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1}\boldsymbol{h}_t'\mathbf{Q}\boldsymbol{h}_t \\ &\quad + 4\gamma_t^3\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{h}_t\boldsymbol{h}_t'\mathbf{Q}\boldsymbol{h}_t \\ &=: B_1 + B_2 + B_3 + B_4 + B_5,\end{aligned} \tag{S.35}$$

say. First, note that $E[B_1] = -4c\gamma_t E\left[(\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1})^2\right]$. Now, suppose that

$$E[B_i] = O(\gamma_t^3) + O(\gamma_t^2)E[(\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1})^2], \text{ for } i = 2,3,4,5. \tag{S.36}$$

By hypothesis, one gets

$$E[(\boldsymbol{a}_t'\mathbf{Q}\boldsymbol{a}_t)^2] = (1-4c\gamma_t)E[(\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1})^2] + O(\gamma_t^2)E[(\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1})^2] + O(\gamma_t^3). \tag{S.37}$$

Hence, for t sufficiently large and some finite constant $C > 0$

$$\begin{aligned}E[(\boldsymbol{a}_t'\mathbf{Q}\boldsymbol{a}_t)^2] &\leq (1-c\gamma_t)E[(\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1})^2] + C\gamma_t^3 \\ &= \Phi_t E[(\boldsymbol{a}_0'\mathbf{Q}\boldsymbol{a}_0)^2] + C\sum_{k=1}^{t} g_{k,t}\gamma_k^2,\end{aligned} \tag{S.38}$$

which, by equation (A.13) and lemma A.1-(d), is of order $O(\gamma_t^2)$. The remainder of this proof thus aims at establishing (S.36). Using the properties mentioned at the beginning of this proof yields

$$E[B_2|\mathcal{A}_{t-1}] = 4\gamma_t^2\sigma^2 \boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1} + (2c\gamma_t)^2(\boldsymbol{a}_{t-1}' \otimes \boldsymbol{a}_{t-1}')E[\boldsymbol{x}_1\boldsymbol{x}_1' \otimes \boldsymbol{x}_1\boldsymbol{x}_1'](\boldsymbol{a}_{t-1} \otimes \boldsymbol{a}_{t-1}) \leq \boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1} + (2c\gamma_t)^2\mu_4^*(\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1})^2, \tag{S.39}$$

where

$$\mu_4^* := \lambda_{\max}(E[\boldsymbol{x}_1\boldsymbol{x}_1' \otimes \boldsymbol{x}_1\boldsymbol{x}_1'])/\lambda_{\min}(\mathbf{Q})^2, \tag{S.40}$$

which, by assumption M.2, takes its values in $(0, \infty)$. Hence, by part (a) of this lemma,

$$E[B_2] \leq 4\gamma_t^2\sigma^2 E[\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1}] + (2c\gamma_t)^2\mu_4^* E[(\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1})^2] = O(\gamma_t^3) + O(\gamma_t^2)E[(\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1})^2]. \tag{S.41}$$

Turning to B_3, observe that

$$(\boldsymbol{h}_t'\mathbf{Q}\boldsymbol{h}_t)^2 = \varepsilon_t^4(\boldsymbol{x}_t'\mathbf{Q}^{-1}\boldsymbol{x}_t)^2 + 6c^2\varepsilon_t^2(\boldsymbol{x}_t'\mathbf{Q}^{-1}\boldsymbol{x}_t\boldsymbol{x}_t'\boldsymbol{a}_{t-1})^2 + c^4(\boldsymbol{a}_{t-1}'\boldsymbol{x}_t\boldsymbol{x}_t'\mathbf{Q}^{-1}\boldsymbol{x}_t\boldsymbol{x}_t'\boldsymbol{a}_{t-1})^2 - 4c\varepsilon_t^3\boldsymbol{x}_t'\mathbf{Q}^{-1}\boldsymbol{x}_t\boldsymbol{x}_t'\mathbf{Q}^{-1}\boldsymbol{x}_t\boldsymbol{x}_t'\boldsymbol{a}_{t-1} - 4c^3\varepsilon_t\boldsymbol{x}_t'\mathbf{Q}^{-1}\boldsymbol{x}_t\boldsymbol{x}_t'\boldsymbol{a}_{t-1}\boldsymbol{a}_{t-1}'\boldsymbol{x}_t\boldsymbol{x}_t'\mathbf{Q}^{-1}\boldsymbol{x}_t\boldsymbol{x}_t'\boldsymbol{a}_{t-1} =: A_1 + 6c^2A_2 + c^4A_3 - 4cA_4 - 4c^3A_5, \tag{S.42}$$

say. Begin with A_1:

$$E[A_1] = \kappa_\varepsilon^{(4)}E[(\boldsymbol{x}_t' \otimes \boldsymbol{x}_t')(\mathbf{Q}^{-1} \otimes \mathbf{Q}^{-1})(\boldsymbol{x}_t \otimes \boldsymbol{x}_t)] \leq \kappa_\varepsilon^{(4)}\lambda_{\max}(\mathbf{Q}^{-1})^2\mathrm{tr}\{E[(\boldsymbol{x}_1\boldsymbol{x}_1')^2]\} \leq \mu_4 k\kappa_\varepsilon^{(4)}, \tag{S.43}$$

which is, by assumption M.2, finite. Next,

$$E[A_2|\mathcal{A}_{t-1}] = \kappa_\varepsilon^{(2)}\boldsymbol{a}_{t-1}'E[\boldsymbol{x}_1\boldsymbol{x}_1'\mathbf{Q}^{-2}(\boldsymbol{x}_1\boldsymbol{x}_1')^2]\boldsymbol{a}_{t-1} \leq \kappa_\varepsilon^{(2)}\mu_6\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1}, \tag{S.44}$$

where

$$\mu_6 := \lambda_{\max}(E[(\boldsymbol{x}_1\boldsymbol{x}_1')^3])/\lambda_{\min}(\mathbf{Q})^3, \tag{S.45}$$

which, by assumption M.2, takes its values in $(0, \infty)$. Hence, by part (a) of this lemma, $E[A_2] = O(\gamma_t)$. Turning to A_3, first note that $A_3 \leq \lambda_{\min}(\mathbf{Q})^{-2}(\boldsymbol{a}_{t-1}'(\boldsymbol{x}_t\boldsymbol{x}_t')^2\boldsymbol{a}_{t-1})^2$. Thus,

$$E[A_3|\mathcal{A}_{t-1}] \leq \lambda_{\min}(\mathbf{Q})^{-2}(\boldsymbol{a}_{t-1}' \otimes \boldsymbol{a}_{t-1}')E[(\boldsymbol{x}_t\boldsymbol{x}_t')^2 \otimes (\boldsymbol{x}_t\boldsymbol{x}_t')^2](\boldsymbol{a}_{t-1} \otimes \boldsymbol{a}_{t-1}) \leq \mu_8^*(\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1})^2, \tag{S.46}$$

where

$$\mu_8^* := \lambda_{\max}(E[(\boldsymbol{x}_t\boldsymbol{x}_t' \otimes \boldsymbol{x}_t\boldsymbol{x}_t')^2])/\lambda_{\min}(\mathbf{Q})^4, \tag{S.47}$$

which, by assumption M.2, takes its values in $(0,\infty)$. Hence, $E[A_3] \leq \mu_8^* E[(\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1})^2]$. Next, it is readily seen that $A_4 \leq \lambda_{\min}(\mathbf{Q})^{-2}\boldsymbol{a}_{t-1}'(\boldsymbol{x}_t\boldsymbol{x}_t')^2\boldsymbol{x}_t'\boldsymbol{a}_{t-1}$. Therefore, using Cauchy-Schwarz's inequality, equations (S.53) and (A.13) yields

$$\begin{aligned}|E[A_4]| &\leq \lambda_{\min}(\mathbf{Q})^{-2}|\kappa_\varepsilon^{(3)}|E[(\boldsymbol{x}_1'\boldsymbol{x}_1)^2\boldsymbol{x}_1'E[\boldsymbol{a}_0]]\Phi_{t-1} \\ &\leq \lambda_{\min}(\mathbf{Q})^{-2}|\kappa_\varepsilon^{(3)}|\mathrm{tr}\{E[(\boldsymbol{x}_1\boldsymbol{x}_1')^4]\}^{1/2}\mathrm{tr}\{\mathbf{Q}E[\boldsymbol{a}_0]E[\boldsymbol{a}_0]'\}^{1/2}\Phi_{t-1} \\ &\leq k|\kappa_\varepsilon^{(3)}|\mu_8^{1/2}E[\boldsymbol{a}_0]'E[\boldsymbol{a}_0]\Phi_{t-1} = O(\exp\{-at^b\}),\end{aligned} \tag{S.48}$$

where

$$\mu_8 := \lambda_{\max}(E[(\boldsymbol{x}_1\boldsymbol{x}_1')^4])/\lambda_{\min}(\mathbf{Q})^{-4}. \tag{S.49}$$

Finally, $E[A_5] = 0$. Putting the above together yields

$$E[B_3] = O(\gamma_t^4)E\big[(\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1})^2\big] + O(\gamma_t^4). \tag{S.50}$$

Next, $E[B_4|\mathcal{A}_{t-1}] = 2\gamma_t^2\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1}E[\boldsymbol{h}_t'\mathbf{Q}\boldsymbol{h}_t|\mathcal{A}_{t-1}]$, where

$$\begin{aligned}E[\boldsymbol{h}_t'\mathbf{Q}\boldsymbol{h}_t|\mathcal{A}_{t-1}] &= \sigma^2 + c^2\boldsymbol{a}_{t-1}'E[\boldsymbol{x}_1\boldsymbol{x}_1'\mathbf{Q}^{-1}\boldsymbol{x}_1\boldsymbol{x}_1']\boldsymbol{a}_{t-1} \\ &\leq \sigma^2 k + c^2\mu_4\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1},\end{aligned} \tag{S.51}$$

so that $E[B_4] = O(\gamma_t^3) + E[(\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1})^2]O(\gamma_t^2)$. Finally, Cauchy-Schwarz's inequality yields

$$E[B_5] \leq 4\gamma_t^3 E[(\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{h}_t)^2]^{1/2}E[(\boldsymbol{h}_{t-1}'\mathbf{Q}\boldsymbol{h}_t)^2]^{1/2} = o(\gamma_t^3); \tag{S.52}$$

where the order of magnitude can be deduced from the analysis of $E[B_2]$ and $E[B_3]$. □

Proof of lemma S.1-(c): Solving (S.1) recursively yields

$$\boldsymbol{a}_t = \boldsymbol{a}_{t-1}(1-c\gamma_t) + \gamma_t\boldsymbol{u}_t + \gamma_t\boldsymbol{e}_t = \boldsymbol{a}_0\Phi_t + \sum_{k=1}^{t} g_{k,t}\boldsymbol{u}_k + \sum_{k=1}^{t} g_{k,t}\boldsymbol{e}_k, \tag{S.53}$$

where $\boldsymbol{u}_t := \mathbf{Q}^{-1}\boldsymbol{x}_t\varepsilon_t$ and $\boldsymbol{e}_t := c(\mathbf{I}_k - \mathbf{Q}^{-1}\boldsymbol{x}_t\boldsymbol{x}_t')\boldsymbol{a}_{t-1}$. Because $E[\boldsymbol{u}_k'\boldsymbol{e}_s] = 0$ for all k and s,

$$E[\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1}] = \Phi_t^2E[\boldsymbol{a}_0'\mathbf{Q}\boldsymbol{a}_0] + \sigma^2\sum_{k=1}^{t} g_{k,t}^2E[\boldsymbol{x}_k'\mathbf{Q}^{-1}\boldsymbol{x}_k] + \sum_{k=1}^{t} g_{k,t}^2E[\boldsymbol{e}_k'\mathbf{Q}\boldsymbol{e}_k]. \tag{S.54}$$

It follows from assumption M.2 and equation (A.13) that the first term decays to zero exponentially. By linearity of the trace operator and lemma A.1-(b),

$$\sigma^2\sum_{k=1}^{t} g_{k,t}^2E[\boldsymbol{x}_k'\mathbf{Q}^{-1}\boldsymbol{x}_k] = \sigma^2\mathrm{tr}\{E[\mathbf{Q}^{-1}\boldsymbol{x}_1\boldsymbol{x}_1']\}\sum_{k=1}^{t} g_{k,t}^2 = \frac{k\gamma_t}{2c}(1+o(1)). \tag{S.55}$$

Turning to the third summand, note that

$$\begin{aligned}|E[\boldsymbol{e}_t'\mathbf{Q}\boldsymbol{e}_t|\mathcal{A}_{t-1}]| &\leq c^2|\boldsymbol{a}_{t-1}'(E[\boldsymbol{x}_1\boldsymbol{x}_1'\mathbf{Q}^{-1}\boldsymbol{x}_1\boldsymbol{x}_1'] - \mathbf{Q})\boldsymbol{a}_{t-1}| \\ &\leq c^2|\mu_4 - 1|\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1}; \end{aligned} \tag{S.56}$$

note that in the scalar case $\mu_4 \geq 1$ by Lyapunov's inequality. By lemma S.1-(a) $E[\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1}] = O(\gamma_t)$, so that, by assumption M.2,

$$\sum_{k=1}^{t} g_{k,t}^2 E[\boldsymbol{e}_k'\mathbf{Q}\boldsymbol{e}_k] = \sum_{k=1}^{t} g_{k,t}^2 O(\gamma_k) = O(\gamma_t), \tag{S.57}$$

using lemma A.1-(d) for the final equality. Therefore, $E[\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1}] = \sigma^2 k\gamma_t/(2c)(1 + o(1))$, which, by equation (A.9), proves the claim. □

Proof of lemma S.2-(a): By lemma S.1-(c),

$$T^{-b}\sum_{t=1}^{T} E[z_t^2] = T^{-b}\sum_{t=1}^{T} E[\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1}] \to k\sigma^2\gamma/(2cb). \tag{S.58}$$

The claim thus follows (by the weak LLN) if

$$\mathrm{var}[\sum_{t=1}^{T} z_t^2] = \sum_{t=1}^{T}\mathrm{var}[z_t^2] + 2\sum_{t=1}^{T-1}\sum_{m=1}^{T-t}\mathrm{cov}[z_t^2, z_{t+m}] = o(T^{2b}). \tag{S.59}$$

To begin with, observe that $\mathrm{var}[z_t^2] = O(\gamma_t^2)$. To see this, consider

$$\begin{aligned}E[z_t^4|\mathcal{A}_{t-1}] &= (\boldsymbol{a}_{t-1}' \otimes \boldsymbol{a}_{t-1}')E[(\boldsymbol{x}_1\boldsymbol{x}_1' \otimes \boldsymbol{x}_1\boldsymbol{x}_1')](\boldsymbol{a}_{t-1} \otimes \boldsymbol{a}_{t-1}) \\ &\leq \mu_4^*(\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1})^2. \end{aligned} \tag{S.60}$$

Hence, by lemma S.1-(a) and S.1-(b), $E[z_t^4] \leq \mu_4^* E[(\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1})^2] = O(\gamma_t^2)$, while $E[z_t^2] = E[\boldsymbol{a}_{t-1}'\mathbf{Q}\boldsymbol{a}_{t-1}] = O(\gamma_t)$. Hence, $\sum_{t=1}^{T}\mathrm{var}[z_t^2] = O(T^b)$. Turning to the covariance expression, it is readily derived that

$$\mathrm{cov}[z_t^2, z_{t+m}^2] = \mathrm{cov}[z_t^2, \boldsymbol{a}_{t+m-1}'\mathbf{Q}\boldsymbol{a}_{t+m-1}]. \tag{S.61}$$

In order to evaluate this term, consider first

$$E[z_t^2\boldsymbol{a}_{t+m}'\mathbf{Q}\boldsymbol{a}_{t+m}|\mathcal{A}_{t-1}] = \boldsymbol{a}_{t-1}'E[\boldsymbol{a}_{t+m}'\mathbf{Q}\boldsymbol{a}_{t+m}\boldsymbol{x}_t\boldsymbol{x}_t'|\mathcal{A}_{t-1}]\boldsymbol{a}_{t-1}. \tag{S.62}$$

Taking account of equation (S.29), it follows that

$$\begin{aligned}E[\boldsymbol{a}_{t+m}'\mathbf{Q}\boldsymbol{a}_{t+m}\boldsymbol{x}_t\boldsymbol{x}_t'|\mathcal{A}_{t-1}] &= (1 - 2c\gamma_{t+m})E[\boldsymbol{a}_{t+m-1}'\mathbf{Q}\boldsymbol{a}_{t+m-1}\boldsymbol{x}_t\boldsymbol{x}_t'|\mathcal{A}_{t-1}] \\ &\quad + \gamma_{t+m}^2\sigma^2 k\mathbf{Q} \\ &\quad + (c\gamma_{t+m})^2 E[(\boldsymbol{a}_{t+m-1}'\mathbf{Q}_4\boldsymbol{a}_{t+m-1})\boldsymbol{x}_t\boldsymbol{x}_t'|\mathcal{A}_{t-1}], \end{aligned} \tag{S.63}$$

with $\mathbf{Q}_4 := E[\boldsymbol{x}_1\boldsymbol{x}_1'\mathbf{Q}^{-1}\boldsymbol{x}_1\boldsymbol{x}_1']$. Similarly, it is seen that

$$\begin{aligned} E[\boldsymbol{a}_{t+m}'\mathbf{Q}\boldsymbol{a}_{t+m}] = &(1-2c\gamma_{t+m})E[\boldsymbol{a}_{t+m-1}'\mathbf{Q}\boldsymbol{a}_{t+m-1}] \\ &+ \gamma_{t+m}^2 k\sigma^2 \\ &+ (c\gamma_{t+m})^2 E[\boldsymbol{a}_{t+m-1}'\mathbf{Q}_4\boldsymbol{a}_{t+m-1}]. \end{aligned} \tag{S.64}$$

Therefore,

$$\begin{aligned} |\text{cov}[z_t^2, z_{t+m}^2]| &\le (1-2c\gamma_{t+m-1})|\text{cov}[z_t^2, z_{t+m-1}^2]| + \theta_{t+m} \\ &\le \text{var}[z_t^2]\Psi_{t+m,t+1} + \sum_{j=1}^{m}\theta_{t+j}\Psi_{t+m,t+j+1}, \end{aligned} \tag{S.65}$$

with $\theta_{t+m} := (c\gamma_{t+m-1})^2|\text{cov}[z_t^2, \boldsymbol{a}_{t+m-2}'\mathbf{Q}_4\boldsymbol{a}_{t+m-2}]|$. Since, as shown already in (A.136),

$$\sum_{t=1}^{T-1}\sum_{m=1}^{T-t}\text{var}[z_t^2]\Psi_{t+m,t+1} = o(T^{2b}), \tag{S.66}$$

the aim is to verify that $\theta_{t+m} = O(\gamma_t\gamma_{t+m}^3)$, because it follows then from the analysis of (R-2) that also

$$\sum_{t=1}^{T-1}\sum_{m=1}^{T-t}\sum_{j=1}^{m}\theta_{t+j}\Psi_{t+m,t+j+1} = o(T^{2b}). \tag{S.67}$$

As a starting point, observe that Cauchy-Schwarz's inequality yields

$$\theta_{t+m} \le (c\gamma_{t+m-1})^2\text{var}[z_t^2]^{1/2}\text{var}[\boldsymbol{a}_{t+m-2}'\mathbf{Q}_4\boldsymbol{a}_{t+m-2}]^{1/2}.$$

Next, it will be shown that $\text{var}[\boldsymbol{a}_{t+m-2}'\mathbf{Q}_4\boldsymbol{a}_{t+m-2}] = O(\gamma_{t+m}^2)$. By lemma S.1-($b$),

$$E[(\boldsymbol{a}_t'\mathbf{Q}_4\boldsymbol{a}_t)^2] = (\boldsymbol{a}_t'\otimes\boldsymbol{a}_t')(\mathbf{Q}_4\otimes\mathbf{Q}_4)(\boldsymbol{a}_t\otimes\boldsymbol{a}_t) \le \kappa_4^2 E[(\boldsymbol{a}_t'\mathbf{Q}\boldsymbol{a}_t)^2] = O(\gamma_t^2), \tag{S.68}$$

Similarly, $E[\boldsymbol{a}_t'\mathbf{Q}_4\boldsymbol{a}_t] \le \kappa_4 E[\boldsymbol{a}_t'\mathbf{Q}'\boldsymbol{a}_t] = O(\gamma_t)$; therby proving $\theta_{t+m} = O(\gamma_t\gamma_{t+m}^3)$. □

Proof of lemma S.2-(b): The proof is an immediate extension of the proof of A.4-(b) and A.4-(c). □

Proof of lemma S.2-(c): The proof is an immediate extension of the proof of A.4-(d). □

S.2 Martingale difference regressors

S.2.1 General remarks

Recall from (A.4) that

$$a_t - \alpha =: a_t^* = \xi_t + a_0^*\vartheta_t, \text{ with } \xi_t = \sum_{k=1}^{t}\gamma_k\vartheta_{t,k+1}u_k, \tag{S.69}$$

where $\vartheta_{m,n} = \prod_{i=n}^{m} f_i(x_i)$, $1 \le n \le m+1$, with $f_i(x_i) = 1 - c\gamma_i x_i^{*2}$; while $u_t = x_t^* \varepsilon_t^*$, with $x_t^* = x_t/\kappa_x^{(2)1/2}$ and $\varepsilon_t^* = \varepsilon_t/\kappa_x^{(2)1/2}$. The stochastic properties of ξ_t play a key role in developing the asymptotic theory; cf. section A.3. Specifically, as demonstrated in section (A.2), evaluation of moments of $\vartheta_{m,n}$ turns out to be non-trivial in case of weakly dependent regressors. However, under part (1) of assumption (B2), the conditional homoskedasticity of $(x_t,\, t \ge 1)$ in conjunction with the tower property of conditional expectations (see, e.g., Davidson (1994, theorem 10.26)) gets (to safe space, set $f_i := f_i(x_i)$)

$$
\begin{aligned}
E[\vartheta_{m,n}] &= E[f_n E[f_{n+1} E[f_{n+2} \dots E[f_{m-1} E[f_m|\mathcal{F}_{m-1}]|\mathcal{F}_{m-2}] \dots |\mathcal{F}_{n+1}]|\mathcal{F}_n]] \\
&= \prod_{i=n}^{m} E[f_i(x_1)] = \prod_{i=n}^{m} (1 - c\gamma_i) = \Phi_{m,n}, \qquad \text{(S.70)}
\end{aligned}
$$

where $\mathcal{F}_t = \sigma(\{x_s,\, s \le t\})$; while the conditional homokurtosis of $(x_t,\, t \ge 1)$ reveals (by similar arguments) that

$$
E[\vartheta_{m,n}^2] = \prod_{i=n}^{m} \psi_i = \Psi_{m,n}, \qquad \text{(S.71)}
$$

where $\psi_i = E[f_i(x_1)^2] = 1 - c\gamma_i \left(2 - c\gamma_i \kappa_x^{(4)}/\kappa_x^{(2)2}\right)$.

S.2.2 Proof of lemma A.4

Lemma S.3 *Assume that part* (1) *of assumption* (B2) *holds true and recall that* $\tau^2 = \sigma^2/\kappa_x^{(2)}$, *while* $b = 1 - \eta$ *and* $a = c\gamma/b$ *with* $c = 1 - \beta$. *Then,*

$$
\begin{aligned}
|E[a_t^*]| &= O(\exp\{-at^b\}), && (a) \\
E[a_t^{*2}] &= \frac{\tau^2 \gamma_t}{2c} + o(\gamma_t), && (b) \\
E[a_t^{*4}] &= O(\gamma_t^2). && (c)
\end{aligned}
$$

Proof of lemma S.3-(a)**:** By equation (A.111), $E[a_t^*] = \kappa_a^{(1)} E[\vartheta_t]$. Assumption (B3) implies $|\kappa_a^{(1)}| < \infty$. By (S.70) and (G-$d$), $E[\vartheta_t] = \Phi_t \le C \exp\{-at^b\}$ for some $C \in (0, \infty)$. □

Proof of lemma S.3-(b)**:** By equation (A.112),

$$
E[a_t^{*2}] = \tau^2 \gamma_t/(2c) + \kappa_a^{(2)} A_t + \tau^2 B_t + o(\gamma_t), \qquad \text{(S.72)}
$$

where $A_t := E[\vartheta_t^2]$ and $B_t := \sum_{k=1}^{t} \gamma_k^2 E[x_k^{*2}(\vartheta_{t,k+1}^2 - \Phi_{t,k+1}^2)]$. It thus remains to be shown that A_t and B_t are $o(\gamma_t)$. From equation (S.71), $E[\vartheta_t^2] = \Psi_t$. From the discussion of (A.135), one gets for t sufficiently large $\Psi_t \le \Phi_t = O(\exp\{-at^b\})$. Turning to B_t, use (S.71) to get $E[x_k^{*2} E[\vartheta_{t,k+1}^2 \mid \mathcal{F}_k]] = \Psi_{t,k+1}$, so that (for $0 \le k \le t$):

$$
E[x_k^{*2}(\vartheta_{t,k+1}^2 - \Phi_{t,k+1}^2)] = \Psi_{t,k+1} - \Phi_{t,k+1}^2 = \Phi_{t,k+1}^2 (V_{t,k+1} - 1), \qquad \text{(S.73)}
$$

where

$$V_{t,k+1} \coloneqq \Psi_{t,k+1}/\Phi^2_{t,k+1} = \prod_{i=k+1}^{t} \psi_i (1-\gamma_{ci})^{-2} = \prod_{i=k+1}^{t} (1+(\mu_4 - 1)(\gamma_{ci}^{-1} - 1)^{-2}), \quad \text{(S.74)}$$

with $\gamma_{ci} \coloneqq c\gamma_i$ and $\mu_4 \coloneqq \kappa_x^{(4)} \kappa_x^{(2)-2}$. Note, that $\mu_4 \geq 1$ (by Lyapunov's inequality) and, by definition, $V_{m,n} \geq 1$). A Laurent-series approximation yields,

$$(\gamma_{ci}^{-1} - 1)^{-2} = \gamma_{ci}^2 + O(\gamma_{ci}^3), \quad \text{(S.75)}$$

where $O(\gamma_{ci}^3)$ denotes a positive quantity. Because $\gamma_{ci} \to 0$, there exists some $i_0 \in \mathbb{N}_1$ such that $O(\gamma_{ci}^3) \leq \gamma_{ci}^2$ for all $i \geq i_0$. For simplicity, assume that $i_0 = 1$. In consequence, a Taylor-series approximation of $\ln(1+x)$ around 0 yields

$$\begin{aligned} \ln(1+(\mu_4-1)(\gamma_{ci}^{-1}-1)^{-2}) &\leq \ln(1+2(\mu_4-1)\gamma_{ci}^2) \\ &= 2(\mu_4-1)\gamma_{ci}^2 - (2(\mu_4-1)\gamma_{ci}^2/\xi_i)^2 \leq 2(\mu_4-1)\gamma_{ci}^2, \end{aligned} \quad \text{(S.76)}$$

where the Lagrangian remainder ξ_i lies on the line segment connecting $\sqrt{2}$ and $\sqrt{2}(1+2(\mu_4 - 1)\gamma_{ci}^2)$; the final inequality assumes that i is large enough to ensure $2(\mu_4 - 1)\gamma_{ci}^2 < 1$. Then, for $k \leq t$ sufficiently large,

$$1 \leq V_{t,k+1} \leq \exp\{2(\mu_4 - 1)c^2 \sum_{i=k+1}^{\infty} \gamma_i^2\} = \exp\{O(\gamma_k^{2-1/\eta})\}, \quad \text{(S.77)}$$

where the order of magnitude follows from equation (A.65); the term $O(\gamma_k^{2-1/\eta})$ is a positive quantity. Assume k is large enough to ensure $O(\gamma_k^{2-1/\eta}) < 1$. The elementary inequality $|\exp\{x\} - 1| \leq |x|(7/4)$ $(0 < |x| < 1)$ then reveals

$$V_{t,k+1} - 1 \leq O(\gamma_k^{2-1/\eta}). \quad \text{(S.78)}$$

Hence, by lemma A.1-(d),

$$B_t = \sum_{k=1}^{t} g_{k,t}^2 O(\gamma_k^{2-1/\eta}) = O(\gamma_t^{3-1/\eta}) = o(\gamma_t), \quad \text{(S.79)}$$

using that, by assumption A, $3 - 1/\eta \in (1, 2)$. This completes the proof. □

Proof of lemma S.3-(c): By Minkowski's inequality, $\|a_t^*\|_4 \leq \kappa_a^{(4)1/4} \|\vartheta_t\|_4 + \|\xi_t\|_4$. By assumption (B3), $\kappa_a^{(4)} < \infty$. Next, by condition (i) of assumption (B2),

$$\kappa_x^{(n)} = E[x_t^n \mid \mathcal{V}_{t-1}] = E[x_t^n \mid \mathcal{F}_{t-1}], \; n = 1, 2, \ldots, 8, \quad \text{(S.80)}$$

with probability one. The tower property of conditional expectation (see also the discussion surrounding equation (A.5)) yields thus for any $n \le m+1$:

$$E[\vartheta_{m,n}^4] = \prod_{i=n}^{m} E[f_i(x_1)^4], \tag{S.81}$$

$$E[f_i(x_1)^4] = 1 - (c\gamma_i)[4 - 6c\gamma_i\mu_4 + 4(c\gamma_i)^2\mu_6 - (c\gamma_i)^3\mu_8], \tag{S.82}$$

where $\mu_\ell := \kappa_x^{(\ell)}/\kappa_x^{(2)\ell}$. Since $\mu_\ell \in (1,\infty)$ for $\ell \in \{4,6,8\}$ (by Lyapunov's inequality and assumption (B2)) and $\gamma_n \to 0$ as $n \to \infty$, there exists some $n_0 \in \mathbb{N}_1$ such that $4 - 6c\gamma_n\mu_2 + 4(c\gamma_n)^2\mu_3 - (c\gamma_n)^3\mu_4 \ge 1$ and thus

$$E[\vartheta_{m,n}^4] \le \Phi_{m,n}, \tag{S.83}$$

for all $n \ge n_0$. Hence, $\|\vartheta_t\|_4$ decays at an exponential rate to zero. Turning to $\|\xi_t\|_4$, it has been shown in (A.114) that $E[\xi_t^4](\kappa_\varepsilon^{(4)}/\kappa_x^{(2)4})A_t + (\sigma/\kappa_x^{(2)})^4 6B_t$, where

$$A_t := \sum_{k=1}^{t} \gamma_k^4 E[x_k^4 \vartheta_{t,k+1}^4] \tag{S.84}$$

$$B_t := \sum_{k=2}^{t}\sum_{s=1}^{k-1} \gamma_s^2\gamma_k^2 E[\vartheta_{t,k+1}^2\vartheta_{t,s+1}^2 x_k^2 x_s^2]. \tag{S.85}$$

Equation (S.83) together with the tower property of conditional expectations and lemma A.1-(d) yields

$$A_t \le \sum_{k=1}^{t} \gamma_k^4 \Phi_{t,k+1} = O(\gamma_t^3). \tag{S.86}$$

Next, by equation (S.83), Cauchy-Schwarz's inequality and the the tower property of conditional expectations,

$$\begin{aligned} B_t \le \kappa_x^{(2)} \sum_{k=2}^{t}\sum_{s=1}^{k-1} \gamma_s^2\gamma_k^2 \Phi_{t,k+1}^{1/2}\Phi_{t,s+1}^{1/2} &= \kappa_x^{(2)} \sum_{k=2}^{t} \gamma_k^2 \Phi_{t,k+1} \sum_{s=1}^{k-1} \gamma_s^2 \Phi_{k,s+1}^{1/2} \\ &= \kappa_x^{(2)} \sum_{k=2}^{t} O(\gamma_k^3)\Phi_{t,k+1} = O(\gamma_t^2), \end{aligned} \tag{S.87}$$

using repeatedly lemma (A.1)-(d). Hence, $\|\xi_t\|_4^4 = O(\gamma_t^2)$, thereby completing the proof. □

S.2.3 Proof of lemma A.5

Lemma S.4 *Assume that part* (1) *of assumption* (B2) *holds true and recall that* $z_t = x_t a^*_{t-1}$, *with* $a^*_t = a_t - \alpha$. *Then,*

$$T^{-b}\sum_{t=1}^{T} z_t^2 \xrightarrow{p} \sigma^2 \frac{\gamma}{2cb} \qquad (a)$$

$$T^{-1/2}\sum_{t=1}^{T} a^*_t = O_p(1) \qquad (b)$$

$$T^{-1/2}\sum_{t=1}^{T}(x_t^2 - \kappa_x^{(2)})a^*_{t-1} = o_p(1) \qquad (c)$$

$$T^{-b/2}\sum_{t=1}^{T} z_t\varepsilon_t \xrightarrow{d} \mathcal{N}(0, \sigma^4\gamma/(2cb)). \qquad (d)$$

Proof of lemma S.4-(a): From the remark accompanying lemma A.4, one deduces that

$$\lim_{T\to\infty} T^{-b}\sum_{t=1}^{T} E[z_t^2] = \kappa_x^{(2)} \lim_{T\to\infty} T^{-b}\sum_{t=1}^{T} E[a^{*2}_{t-1}] = \sigma^2\gamma/(2cb). \qquad \text{(S.88)}$$

Hence, by Chebychev's inequality, a sufficient condition for lemma S.4-(a) to hold is

$$\text{var}[\sum_{t=1}^{T} z_t^2] = \sum_{t=1}^{T}\text{var}[z_t^2] + 2\sum_{t=1}^{T-1}\sum_{m=1}^{T-t}\text{cov}[z_t^2, z_{t+m}^2] = o(T^{2b}). \qquad \text{(S.89)}$$

By condition (i) of assumption (B2) and lemma A.3, one gets

$$\text{var}[z_t^2] = \kappa_x^{(4)} E[a^{*4}_{t-1}] - \kappa_x^{(2)2} E[a^{*2}_{t-1}]^2 = O(\gamma_t^2). \qquad \text{(S.90)}$$

Hence, $\sum_{t=1}^{\infty}\text{var}[z_t^2] = \sum_{t=1}^{\infty} O(\gamma_t^2) < \infty$, using assumption A. Next, it will be shown that

$$\text{cov}[z_t^2, z_{t+m}^2] = \kappa_x^{(2)2}\text{var}[a^{*2}_{t-1}]\Psi_{t+m-1,t}, \qquad \text{(S.91)}$$

where $\Psi_{m,n}$ has been defined in equation (A.134). The following makes heavily use of assumption (B2) (condition (i)), assumption (B3), and the tower property of conditional expectations (see, e.g., Davidson (1994, theorem 10.26)). Specifically, it will be used that for $1 \le n \le 4$

$$E[x_t^n|\mathcal{A}_{t-1}] = E[x_t^n|\mathcal{V}_{t-1}] = E[x_t^n|\mathcal{F}_{t-1}] = \kappa_x^{(n)} \ a.s., \qquad \text{(S.92)}$$

with the sigma algebra $\mathcal{A}_t$ being defined as $\mathcal{A}_t = \sigma(\{(x_s, \varepsilon_s), s \le t\} \cup a_0)$; for the definitions of $\mathcal{F}_t$ and $\mathcal{V}_t$ please refer to assumption (B2). Clearly, $\mathcal{F}_t \subseteq \mathcal{V}_t \subseteq \mathcal{A}_t$. The preceding yields $\text{cov}[z_t^2, z_{t+m}^2] = \kappa_x^{(2)}\text{cov}[z_t^2, a^{*2}_{t+m-1}]$, with $\text{cov}[z_t^{*2}, a^{*2}_{t+m-1}] = E[a^{*2}_{t-1}E[x_t^2 a^{*2}_{t+m-1}|\mathcal{A}_{t-1}]] - \kappa_x^{(2)}E[a^{*2}_{t-1}]E[a^{*2}_{t+m-1}]$. Furthermore, one gets from equation (A.125)

$$E[x_t^2 a^{*2}_{t+m-1}|\mathcal{A}_{t-1}] = E[x_t^2 a^{*2}_{t+m-2}|\mathcal{A}_{t-1}]\psi_{t+m-1} + \gamma^2_{t+m-1}\sigma^2 \qquad \text{(S.93)}$$

$$E[a^{*2}_{t+m-1}] = E[a^{*2}_{t+m-1}]\psi_{t+m} + \gamma^2_{t+m-1}\tau^2. \qquad \text{(S.94)}$$

Therefore,

$$\text{cov}[z_t^2, a_{t+m-1}^{*2}] = \text{cov}[z_t^2, a_{t+m-2}^{*2}]\psi_{t+m-1}. \tag{S.95}$$

Solving recursively thus yields (S.91). By equation (A.136), this yields (S.89); thereby proving the claim. □

Proof of lemma S.4-(b)**:** By equation (A.180),

$$T^{-1/2}\sum_{t=1}^{T} a_t^* = T^{-1/2}a_0^*\sum_{t=1}^{T}\vartheta_t + T^{-1/2}\sum_{t=1}^{T}\xi_t =: A_T + B_T, \tag{S.96}$$

say. It will be shown that

$$(i)\ A_T = O_{a.s.}(T^{-1/2}) \text{ and } (ii)\ B_T = O_p(1). \tag{S.97}$$

Begin with (i). Equations (S.71) and (A.13) in conjunction with Cauchy-Schwarz's inequality yield for some $C > 0$

$$\sum_{t=1}^{\infty} E[|a_0^*\vartheta_t|] \le \kappa_a^{(2)\,1/2}\sum_{t=1}^{\infty}\|\vartheta_t\|_2 \le C\kappa_a^{(2)\,1/2}\sum_{t=1}^{\infty}\Phi_t^{1/2} < \infty, \tag{S.98}$$

which, in turn, implies that $\sum_{t=1}^{\infty} a_0^*\vartheta_t$ converges $a.s.$. Turning to (ii), observe that $E[B_T] = 0$, while

$$E[B_T^2] = T^{-1}\sum_{t=1}^{T} E[\xi_t^2] + 2T^{-1}\sum_{t=2}^{T}\sum_{s=1}^{t-1} E[\xi_t\xi_s]. \tag{S.99}$$

Now, for $1 \le s \le t$:

$$\begin{aligned} E[\xi_t\xi_s] &= \sum_{k=1}^{t}\sum_{j=1}^{s}\gamma_k\gamma_j E[\vartheta_{t,k+1}\vartheta_{s,j+1}u_k u_j] \\ &= \tau^2\sum_{k=1}^{s}\gamma_k^2 E[\vartheta_{t,k+1}\vartheta_{s,k+1}x_k^{*2}] \\ &= \tau^2\sum_{k=1}^{s}\gamma_k^2 E[x_k^{*2}E[\vartheta_{t,k+1}\vartheta_{s,k+1} \mid \mathcal{F}_k]]. \end{aligned} \tag{S.100}$$

Taking equations (A.5) and (A.134) into account, it follows from the power property of conditional expectations that

$$E[\vartheta_{t,k+1}\vartheta_{s,k+1} \mid \mathcal{F}_k] = E[\vartheta_{t,s+1}\vartheta_{s,k+1}^2 \mid \mathcal{F}_k] = \Phi_{t,s+1}\Psi_{s,k+1}. \tag{S.101}$$

Hence, by equation (S.73) and the fact that $g_{k,t}g_{k,s} = \gamma_k^2 \Phi_{t,s+1}\Phi_{s,k+1}^2 = \Phi_{t,s+1}g_{k,s}^2$,

$$E[\xi_t \xi_s] = \tau^2 \sum_{k=1}^{s} \gamma_k^2 \Phi_{t,s+1}\Psi_{s,k+1}$$
$$= \tau^2 \sum_{k=1}^{s} g_{k,t}g_{k,s} + \tau^2 \Phi_{t,s+1} \sum_{k=1}^{s} g_{k,s}^2 (V_{t,k+1} - 1), \tag{S.102}$$

where the notation $V_{t,k+1} = \Psi_{t,k+1}/\Phi_{t,k+1}^2$ has been introduced in (S.74). By equation (S.78), $V_{t,k+1} - 1 = O(\gamma_k^{2-1/\eta})$. Hence, by lemma A.1-($d$),

$$E[\xi_t \xi_s] = \tau^2 \sum_{k=1}^{s} g_{k,t}g_{k,s} + O(\gamma_s^{2-1/\eta} g_{s,t}). \tag{S.103}$$

Specifically, the above yields for $t = s$ (cf. lemma A.1-(a)):

$$E[\xi_t^2] = = \tau^2 \phi_t^{ii} + O(\gamma_t^{2-1/\eta}). \tag{S.104}$$

Hence, by lemma A.1,

$$E[B_T^2] = \frac{1}{T}\sum_{t=1}^{T} O(\phi_t^{ii} + \gamma_t^{2-1/\eta}) + \frac{1}{T}\sum_{t=2}^{T} O(\phi_t^{iii} + \gamma_t^{2-1/\eta}) = O(1), \tag{S.105}$$

thereby proving $\mathrm{var}[B_T] = O(1)$, i.e. $B_T = O_p(1)$. This shows (b). □

Proof of lemma S.4-(c): Set $v_t := (x_t^2 - \kappa_x^{(2)})a_{t-1}^*$. Assumption B implies that

$$E[v_t \mid \mathcal{A}_{t-1}] = a_{t-1}^* E[x_t^2 - \kappa_x^{(2)} \mid \mathcal{A}_{t-1}] = a_{t-1}^* E[x_t^2 - \kappa_x^{(2)} \mid \mathcal{F}_{t-1}] = 0, \tag{S.106}$$

with probability one; i.e. $(\{v_t, \mathcal{A}_t\}; t \geq 1)$ is a martingale difference sequence. Hence, by lemma S.3-(b) and equation (A.9),

$$\mathrm{var}[T^{-1/2}\sum_{t=1}^{T} v_t] = T^{-1}\sum_{t=1}^{T} E[v_t^2] = (\kappa_x^{(4)} - \kappa_x^{(2)})T^{-1}\sum_{t=1}^{T} E[a_{t-1}^{*2}] = O(T^{-\eta}). \tag{S.107}$$

By Chebychev's inequality, this completes the proof. □

Proof of lemma S.4-(d): Follows immediately from the proof of lemma A.5-(d). □