

Estimation and Inference in Adaptive Learning Models with Slowly Decreasing Gains

Alexander Mayer

July 2018

Economics Group



Estimation and Inference in Adaptive Learning Models with Slowly Decreasing Gains

Alexander Mayer

WHU – Otto Beisheim School of Management

Working Paper 18/03
July 2018

ISSN 2511-1159

WHU - Otto Beisheim School of Management
Economics Group
Burgplatz 2
56179 Vallendar, Germany
Phone: +49 (261) 65 09 - 0
whu@whu.edu

Any opinions expressed here are those of the author(s) and not those of WHU. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

WHU Working Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

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. It is shown that this procedure is asymptotically equivalent to first estimating the structural parameters jointly via ordinary least-squares (OLS) 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, a limiting normal distribution of the estimators centered at the true parameters is derived. Although this distribution is singular, it can nevertheless be used to draw inferences about joint restrictions by applying results from Andrews (1987) to show that Wald-type statistics remain valid when equipped with a pseudo-inverse. Monte-Carlo evidence confirms the accuracy of the asymptotic theory for the finite sample behaviour of estimators and test statistics discussed here.

JEL-Classification: C12; C22; C51; D83

Keywords: adaptive learning, rational expectations, singular limiting-distribution, non-stationary regression, generalized Wald statistic, degenerate variances.

Corresponding author:

Alexander Mayer, alexander.mayer@whu.edu

Estimation and Inference in Adaptive Learning Models with Slowly Decreasing Gains

Alexander Mayer*

July 9, 2018

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. It is shown that this procedure is asymptotically equivalent to first estimating the structural parameters jointly via ordinary least-squares (OLS) 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, a limiting normal distribution of the estimators centered at the true parameters is derived. Although this distribution is singular, it can nevertheless be used to draw inferences about joint restrictions by applying results from Andrews (1987) to show that Wald-type statistics remain valid when equipped with a pseudo-inverse. Monte-Carlo evidence confirms the accuracy of the asymptotic theory for the finite sample behaviour of estimators and test statistics discussed here.

Keywords: adaptive learning, rational expectations, singular limiting-distribution, non-stationary regression, generalized Wald statistic, degenerate variances.

JEL codes: C12, C22, C51, D83

*WHU – Otto Beisheim School of Management, Burgplatz 2, 56179 Vallendar, Germany, alexander.mayer@whu.edu. I would like to thank Michael Massmann and participants of the WHU Colloquium in Econometrics & Statistics, participants of the research seminar of the chair for statistics and econometrics at the University Cologne and attendees of the IAAE conference 2018 for helpful discussions and suggestions.

1 Introduction

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

$$y_t = \beta y_{t|t-1}^e + \delta x_t + \varepsilon_t \quad \text{for } t = 1, 2, \dots, T, \quad (1)$$

where $y_{t|t-1}^e$ 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 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); for a more detailed account of examples encompassed by (1) see Christopeit and Massmann (2018).

The crucial feature of model (1) is the expectation term $y_{t|t-1}^e$. Under rational expectations (RE), economic agents are presumed to incorporate the information $\mathcal{F}_{t-1} := \sigma(y_s, s < t; x_s, s \leq t)$ in an optimal manner so that

$$y_{t|t-1}^e = \mathbb{E}(y_t | \mathcal{F}_{t-1}), \quad (2)$$

thereby yielding the RE equilibrium

$$y_t = \alpha x_t + \varepsilon_t \quad \text{with } \alpha := \delta / (1 - \beta). \quad (3)$$

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 $\mathbb{E}(y_t | \mathcal{F}_{t-1})$, the economic agent acts like an econometrician forecasting α recursively. Specifically, the agent updates her expectations according to the adaptive scheme

$$y_{t|t-1}^e = a_{t-1} x_t, \quad (4)$$

where the point forecast a_t of α is obtained via stochastic approximation algorithms (see e.g. Sargent (1993) or Evans and Honkapohja (2001)). In particular, the present paper assumes the learning scheme to take the form of a least-squares type stochastic approximation algorithm

$$\begin{aligned} a_t &= a_{t-1} + \gamma_t \frac{x_t}{r_t} (y_t - x_t a_{t-1}) \\ r_t &= r_{t-1} + \gamma_t (x_t^2 - r_{t-1}), \end{aligned} \quad (5)$$

whereby, in addition to estimating α by a_t , a further ‘normalization’ step based on r_t is used to estimate the regressor second moment. The so-called ‘gain’ γ_t reflects the agent’s responsiveness to previous forecast errors. Empirical applications confirm the plausibility of (5) for the formation of expectations (see, e.g., Chakraborty and Evans (2008), Berardi and Galimberti (2014), Berardi and Galimberti (2017) or Markiewicz and Pick (2017)).

Christopeit and Massmann (2018) thoroughly study the statistical properties of the joint OLS estimator of $\lambda := (\beta, \delta)'$ for constant gains ($\gamma_t = \gamma$) and recursive least-squares ($\gamma_t = 1/t$). The present paper adds to their findings in four important aspects: First, while Christopeit and Massmann (2018) consider a deterministic and constant regressor, the current exposition treats x_t as a time-varying random variable. Second, the asymptotic properties of estimators of the equilibrium parameter α are established. Third, the question of hypothesis testing in the presence of a singular variance-covariance matrix is addressed. Finally, the agent, in this case, tries to learn the RE equilibrium parameter α using a sequence of gains that decreases more slowly than t^{-1} , an idea formalized by assumption 1 below. In this paper, it is shown that estimators of the structural parameters β and δ converge at a polynomial rate to its joint normal distribution, with the rate of convergence being inversely related to the degeneration rate of the gain. 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

$$y_t = \beta a_{t-1} x_t + \delta x_t + \varepsilon_t, \tag{6}$$

converges at a faster rate to the ‘long-run’ RE equilibrium parameter α . Finally, single hypotheses about β and δ as well as joint restrictions placed upon λ can be tested when suitable test statistics are used.

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.

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 (5). Specifically, γ_t is assumed to be of polynomial form:

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

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

where $\eta \in (0, 1)$ and $|c\gamma| < 1$ with $c := 1 - \beta$.

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 Christopheit and Massmann (2018). Following the terminology used in the stochastic approximation literature, this polynomial gain with $\eta \in (0, 1)$ considered here will be henceforth referred to as *slowly decreasing* (cf. Polyak and Juditsky (1992), Kushner and Yan (1993) or Chen (1993)).

The next assumption requires $v_t := (x_t, \varepsilon_t)'$ to be independent, identically distributed (*i.i.d.*).

Assumption 2. *The elements of the random vector v_t are independent of $(a_0, r_0)'$ and mutually independent, identically distributed with finite variances so that*

$$v_t \stackrel{i.i.d.}{\sim} \left(\begin{bmatrix} \mu_x \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \right).$$

The small sample experiment in section 3 indicates that the asymptotic approximations derived below continue to hold when the regressor is serially correlated. Serial uncorrelatedness of the error term is, however, a necessary condition for consistent estimability of the least-squares estimators of β and δ discussed below. The reason is that their estimation involves empirical moment conditions of the form

$$\sum_{t=1}^T (a_{t-1} - \alpha)x_t \varepsilon_t. \tag{7}$$

Since a_t contains the complete history of the innovation at time t , the population analog of (7) is non-zero whenever ε_t exhibits serially correlation.

In view of the almost sure convergence of r_t to the regressor second moment (cf. appendix A.7), it is assumed that r_t is centered at the true parameter value $\kappa_x^{(2)} := E(x_t^2)$, i.e.

Assumption 3. *For all t , $r_t = \kappa_x^{(2)}$.*

Although the Monte-Carlo experiment in section 3 suggests that assumption 3 might be relaxed, it will nevertheless be retained for reasons of analytical tractability. Specifically, under assumptions 1, 2 and 3, the agents' updating scheme (5) satisfies the so-called *Robbins-Monro*

conditions. Recursive algorithms of this class have a long tradition in the stochastic approximation literature, where they are usually analyzed 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 known results from this literature in order to establish

$$\mathbb{E}\left[(a_t - \alpha)^\ell\right] = O(\gamma_t^{\ell/2}) \quad \text{for } \ell \in \{2, 4\}; \quad (8)$$

that is, moment bounds required in the development of an appropriate asymptotic theory for the estimation and inference procedures discussed below. Finally, the distributional characteristics of v_t and the initial condition a_0 have to be sharpened.

Assumption 4. *Let $\kappa_a^{(\ell)}$, $\kappa_\varepsilon^{(\ell)}$ and $\kappa_x^{(\ell)}$ denote the ℓ^{th} non-central moment of $a_0 - \alpha$, ε_t and x_t , respectively and assume that $\kappa_a^{(4)}$, $\kappa_\varepsilon^{(4)}$ and $\kappa_x^{(8)}$ are finite and non-zero.*

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

Consider the OLS estimator for $\lambda = (\beta, \delta)'$

$$\hat{\lambda} := \left(\sum_{t=1}^T w_t w_t' \right)^{-1} \sum_{t=1}^T w_t y_t \quad \text{with } w_t := (a_{t-1} x_t, x_t)'. \quad (9)$$

As a starting point for a discussion of the statistical properties of $\hat{\lambda}$, observe that the regressor w_t fails Grenander's conditions (see, for example, Hannan (1970, p. 215)) for well-behaved data as both the sample second moment matrix of the regressor

$$M_T := \sum_{t=1}^T w_t w_t' \quad (10)$$

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

$$M_T/T \xrightarrow{p} \kappa_x^{(2)} \begin{bmatrix} \alpha^2 & \alpha \\ \alpha & 1 \end{bmatrix} \quad \text{and} \quad T^b M_T^{-1} \xrightarrow{p} \frac{2cb}{\gamma\sigma^2} \begin{bmatrix} 1 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix}, \quad (11)$$

where $b := 1 - \eta$ is smaller the more quickly the gain parameter $\gamma_t = \gamma/t^\eta$ approaches zero as $t \rightarrow \infty$; cf. appendix B. 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. lemma A.3 of appendix A.4. The other crucial aspect of (11) is the slower convergence rate of the inverse as compared to the regressor sample moment matrix. This behavior might seem surprising but is analogous to that of regression

models with slowly varying trends as discussed in Phillips (2007) and can best be understood by noting that the determinant of M_T has to be rescaled by T^{1+b} in order to achieve convergence, i.e.

$$\frac{\det M_T}{T^{1+b}} \xrightarrow{p} \frac{\kappa_x^{(2)} \gamma \sigma^2}{2cb}; \quad (12)$$

cf. appendix B. As a consequence, the suitably normalized deviation of $\widehat{\lambda}$ from the true vector λ obeys the approximate asymptotic representation

$$T^{b/2}(\widehat{\lambda} - \lambda) \overset{a}{\approx} \frac{2cb}{\gamma \sigma^2} \begin{bmatrix} 1 \\ -\alpha \end{bmatrix} \frac{1}{T^{b/2}} \sum_{t=1}^T a_{t-1}^* x_t \varepsilon_t, \quad (13)$$

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

Proposition 2.1. *Suppose assumptions 1, 2, 3 and 4 hold. Then*

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

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

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

Proof. See appendix B.

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.

Using a different approach and imposing the assumption that $x_t = 1$, Christopheit 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 Christopheit and Massmann (2018) is equivalent to that stated above up to a factor of proportionality which stems from the different specification of the gain; see also the discussion in Christopheit and Massmann (2015, pp. 19). The crucial difference between their result and proposition 2.1 is the rate of convergence: while Christopheit 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 Christopheit 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 higher¹ than $T^{b/2}$. Note that V^\perp denotes the orthogonal complement of the limiting variance-covariance V , i.e. $V^\perp V = 0'_k$. As summarized by corollary 2.1 below, this linear combination converges at the ‘standard’ rate of $T^{1/2}$ to the RE equilibrium parameter (note that $V^\perp \lambda = \alpha$).

Corollary 2.1. *Let the conditions of proposition 2.1 hold. Define $\tau^2 := \sigma^2/\kappa_x^{(2)}$ and the multivariate normalization*

$$F_T := \begin{bmatrix} T^{b/2} I_2 & 0_2 \\ 0'_2 & T^{1/2} \end{bmatrix},$$

where I_k denotes the k -dimensional identity matrix. Then

$$F_T A (\widehat{\lambda} - \lambda) \xrightarrow{d} \mathcal{N} \left(0_3, \begin{bmatrix} V & 0_2 \\ 0'_2 & \tau^2 \end{bmatrix} \right), \text{ with } A := \begin{bmatrix} I_2 \\ V^\perp \end{bmatrix}.$$

Proof. See appendix C.

The RE equilibrium is in general unknown, so that corollary 2.1 cannot be used in order to estimate α . Instead, consider the simple plug-in estimator

$$\widehat{\lambda}_\alpha := \frac{\widehat{\lambda}_\delta}{1 - \widehat{\lambda}_\beta}, \tag{14}$$

where $\widehat{\lambda}_\beta$ and $\widehat{\lambda}_\delta$ denote respectively the two components of $\widehat{\lambda}$.² Since

$$\widehat{\lambda}_\alpha - \alpha = V^\perp \frac{(\widehat{\lambda}_\alpha - \lambda)}{1 - \widehat{\lambda}_\beta},$$

¹For 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 2.2 below or the discussion in Lütkepohl and Burda (1997) on testing noncausality restrictions using vector autoregressions.

²A simple calculation shows that $\widehat{\lambda}_\alpha = (\widehat{\lambda}_\alpha, 1)\widehat{\lambda}$.

the asymptotic normality of (14) follows as a by-product of the preceding corollary, i.e.

$$T^{1/2}(\widehat{\lambda}_\alpha - \alpha) \xrightarrow{d} \mathcal{N}(0, (\tau/c)^2). \quad (15)$$

The limiting variance $(\tau/c)^2$ coincides with the lower bound established in the literature on stochastic approximation for the averaged iterates estimator for α (see e.g. Polyak and Juditsky (1992), Kushner and Yan (1993) or Chen (1993)). Clearly, the typical limiting variance of the OLS estimator for homoskedastic linear regressions is recovered if the expectation term in (1) is absent (i.e. $\beta = 0$ so that $c = 1$). Note, furthermore, that the asymptotic variance of $\widehat{\lambda}_\alpha$ is strictly larger than that of the infeasible linear combination $V^\perp \widehat{\lambda}$ of corollary 2.1 as $\beta < 1$.

Remark 2.2. *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*

$$y_t = \beta y_{t-1} + \delta x_t + u_t, \quad (\text{ADL})$$

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,

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

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 (3), 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.

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 $\widehat{\lambda}$. As a natural starting-point, consider the textbook t statistics

$$t_\beta := \frac{\widehat{\lambda}_\beta - \beta_0}{\sqrt{m^{11} \widehat{\sigma}^2}} \quad \text{and} \quad t_\delta := \frac{\widehat{\lambda}_\delta - \delta_0}{\sqrt{m^{22} \widehat{\sigma}^2}}, \quad (16)$$

where m^{ii} represents the i^{th} diagonal element of M_T^{-1} (with M_T being defined in (10)). As shown in the the remark to appendix B, 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 t statistic for α based on $\widehat{\lambda}_\alpha$ is similarly defined as $t_\alpha := (\widehat{\lambda}_\alpha - \alpha_0)/\text{SE}(\widehat{\lambda}_\alpha)$ with the standard error given by

$$\text{SE}(\widehat{\lambda}_\alpha) := \frac{\sqrt{\widehat{\sigma}^2 / \sum_{t=1}^T x_t^2}}{1 - \widehat{\lambda}_\beta}. \quad (17)$$

Note that t_α does not approach the standard normal null-distribution at the standard rate $T^{1/2}$, even though $T^{1/2}(\widehat{\lambda}_\alpha - \alpha_0)$ is asymptotically normal under the null, cf. equation (15). The reason is that $T^{1/2}\text{SE}(\widehat{\lambda}_\alpha)$ converges at the slower rate $T^{b/2}$ to the true standard deviation τ/c due to the appearance of $\widehat{\lambda}_\beta$ in the denominator of (17).

The case of joint hypotheses about λ is slightly more involved, since the suitably normalized estimator $\widehat{\lambda}$ has been shown to be asymptotically normal with singular variance-covariance matrix

$$V := (2cb/\gamma)Q \quad \text{with} \quad Q := \begin{bmatrix} 1 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix}, \quad (18)$$

which, for convenience, is restated here from proposition 2.1. As forcefully argued in Andrews (1987), even though the usual variance-covariance estimator V_T consistently estimates V (cf. proposition 2.1), inferences drawn from the classical Wald statistic equipped with V_T is misleading for testing joint restrictions of the form

$$H_0: r(\lambda) = 0, \quad (19)$$

where $r(\lambda) := R\lambda - q$ with $q \in \mathbb{R}^2$ and R representing a suitable 2×2 matrix of linear restrictions. The null (19) can nevertheless be tested using a Wald-type statistic based on a generalized inverse of a suitable variance-covariance estimator, provided a certain rank-condition

is satisfied, see Andrews (1987, theorem 2). To be more specific, recall that for any two-dimensional, singular matrix B the (Moore-Penrose) generalized inverse is defined via

$$B^+ := A \begin{bmatrix} 1/\xi & 0 \\ 0 & 0 \end{bmatrix} A', \quad (20)$$

with A and ξ denoting respectively the 2×2 matrix of eigenvectors and the largest eigenvalue of B , see, e.g., Seber (2008, chapter 7). Yet the pseudo-inverse (20) is not continuous and V_T violates the rank condition of Andrews (1987): $\text{rk}(T^b V_T) > \text{rk}(V)$ for any T with probability one. In consequence,

$$(T^b V_T)^+ = (T^b V_T)^{-1} \xrightarrow{p} V^+$$

even though $T^b V_T$ is consistent for V , cf. proposition 2.1. An alternative variance-covariance estimator is therefore needed. Taking (14) into account, a consistent alternative is given by $T^b \tilde{V}_T$ with

$$\tilde{V}_T := m^{11} \hat{\sigma}^2 \begin{bmatrix} 1 & -\hat{\lambda}_\alpha \\ -\hat{\lambda}_\alpha & \hat{\lambda}_\alpha^2 \end{bmatrix}, \quad (21)$$

see the remark to appendix B. Since $\text{rk}(\tilde{V}_T) = \text{rk}(V)$ by construction, Andrews' (1987) theorem 1 applies and inference can be carried out using

$$\mathcal{W} := r(\hat{\lambda})'(R\tilde{V}_T R')^+ r(\hat{\lambda}). \quad (22)$$

Note that the singularity of V results in \mathcal{W} behaving under the null asymptotically like a χ^2 distributed random variable with *one* degree of freedom. Moreover, it can be inferred from the discussion in Andrews (1987) and Lütkepohl and Burda (1997) that \mathcal{W} has non-trivial power against local alternatives in a $1/T^{b/2}$ -neighborhood of the null; a question pursued further in section 3.

2.4 Extensions

A. MULTIPLE REGRESSORS. The preceding limiting theory extends naturally from the single variable case to a more general setting in which x_t represents a column vector of $k \geq 1$ regressors, so that the actual law of motion is given by

$$y_t = \beta a'_{t-1} x_t + \delta' x_t + \varepsilon_t \quad \text{for } t = 1, 2, \dots, T, \quad (23)$$

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

$$\begin{aligned} a_t &= a_{t-1} + \gamma_t R_t^{-1} x_t (y_t - a'_{t-1} x_t) \\ R_t &= R_{t-1} + \gamma_t (x_t x'_t - R_{t-1}). \end{aligned} \quad (24)$$

Now, let $\widehat{\lambda}$ denote the $(k+1)$ -dimensional OLS estimator of $\lambda := (\beta, \delta)'$ based on (23) and define $V^\perp := (\alpha, I_k)$, a $k \times (k+1)$ matrix resulting from aligning α and I_k columnwise so that $V^\perp \lambda = \alpha$. Just as in the scalar case treated above, it follows that $T^{b/2}(\widehat{\lambda} - \lambda)$ is asymptotically normal while the linear combination $V^\perp(\widehat{\lambda} - \lambda)$ converges to its limiting normal distribution at the faster rate $T^{1/2}$.

Specifically, define the $(2k+1) \times (k+1)$ matrix $A := (I_{k+1}, V^{\perp'})'$ which has the property $A\lambda = (\delta', \alpha)'$ and introduce the multivariate scaling matrix of dimension $(2k+1) \times (2k+1)$

$$F_T := \begin{bmatrix} T^{b/2} I_{k+1} & \mathbf{O} \\ \mathbf{O}' & T^{1/2} I_k \end{bmatrix}, \quad (25)$$

where the \mathbf{O} block represents a $(k+1) \times k$ matrix of zeros. Under the additional condition that

$$R_t = Q := \mathbb{E}(x_t x'_t)$$

has full rank for all t , it can be shown that

$$F_T A (\widehat{\lambda} - \lambda) \xrightarrow{d} \mathcal{N} \left(\mathbf{0}_{2k+1}, \begin{bmatrix} V & \mathbf{O} \\ \mathbf{O}' & \sigma^2 Q^{-1} \end{bmatrix} \right) \quad \text{with } V := \frac{2cb}{k\gamma} \begin{bmatrix} 1 & -\alpha' \\ -\alpha & \alpha\alpha' \end{bmatrix}. \quad (26)$$

Again, it is seen that V^\perp is the orthogonal complement of the rank-one variance-covariance matrix V , i.e. $V^\perp V = \mathbf{O}'$. The proof of equation (26) is delegated to appendix F. As can be seen, the only material difference between the previous display and its counterpart of corollary 2.1 is the appearance of k (the number of regressors) in the denominator of the limiting variance-covariance matrix V . As before, equation (26) be used to carry out inference about $(\beta, \delta)'$ and to construct a plug-in estimator of α analogous to (15).

B. CONSTANT GAIN. In certain cases, the learning literature suggests the plausibility of constant gain sequences instead of decreasing gains, i.e. $\eta = 0$ so that $\gamma_t = \gamma > 0$; see, e.g., Chakraborty and Evans (2008). It follows that the preceding sections continue to hold without material differences as long as γ is ‘sufficiently’ small. The asymptotic theory in this case is based on a limiting expansion of γ and T , whereby $T \rightarrow \infty$ and $\gamma \rightarrow 0^+$. Similar ‘small- γ ’ asymptotics for analyzing constant gain algorithms are commonly used in the literature on

stochastic approximation (e.g. Kushner and Huang (1981)) and macroeconomic learning (e.g. Evans and Honkapohja (2001, chapter 7.4)).³

In order for the impact of the initial value a_0 on the asymptotic distribution of $\widehat{\lambda}$ to vanish, an additional condition on the relative divergence rates of γ and T has to be imposed, namely $\gamma^2 T \rightarrow \infty$. In a sense, this assumption restricts, analogous to the ‘slowly’ decreasing gain assumption 1, the gain sequence not to approach zero too fast in the asymptotic analysis. Under these limiting expansions of γ and T , $\widehat{\lambda}$ is asymptotically normal at the rate $(\gamma T)^{1/2}$, while the linear combination $V^\perp \widehat{\lambda}$ with $V^\perp := (\alpha, 1)$ is ‘super-consistent’ at the faster rate $T^{1/2}$. Specifically, let

$$F_{\gamma T} := \begin{bmatrix} (\gamma T)^{1/2} I_2 & 0_2 \\ 0_2' & T^{1/2} \end{bmatrix},$$

denote the three-dimensional normalizing matrix and assume that assumptions 2, 3 and 4 continue to hold. Then, it follows that

$$F_{\gamma T} A (\widehat{\lambda} - \lambda) \xrightarrow{d} \mathcal{N} \left(0_3, \begin{bmatrix} V & 0_2 \\ 0_2' & \tau^2 \end{bmatrix} \right) \quad \text{with } V := 2c \begin{bmatrix} 1 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix}. \quad (27)$$

where A has the property $A\lambda = (\lambda', \alpha)'$, see the discussion surrounding proposition 2.3. The proof of equation (27) is delegated to appendix G.

2.5 An asymptotically equivalent three-step estimator

This section presents an asymptotically equivalent approach to the estimation procedure for (α, β, δ) discussed at the beginning of this paragraph. Suppose, to begin with, that the true RE equilibrium parameter α is known and the interest lies on 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

$$y_t - \alpha x_t = \beta a_{t-1}^* x_t + \varepsilon_t. \quad (28)$$

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

$$\widehat{\beta}_0 = \frac{\sum_{t=1}^T (y_t - \alpha x_t) a_{t-1}^* x_t}{\sum_{t=1}^T (a_{t-1}^* x_t)^2}, \quad (29)$$

which, indeed, is asymptotically equivalent to $\widehat{\lambda}_\beta$, see proposition 2.1:

³Note that this approach differs from the one in Christopheit and Massmann (2018), who assume γ to be constant in their asymptotic analysis.

Proposition 2.2. *Suppose that assumptions 1, 2, 3 and 4 hold. Then*

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

Proof. See appendix D. □

Since $\widehat{\beta}_0$ depends on the unknown RE equilibrium parameter, this approach is infeasible unless a suitable estimator for α is available. Under RE, one might resort to estimating (3), yielding

$$\widehat{\alpha} = \frac{\sum_{t=1}^T y_t x_t}{\sum_{t=1}^T x_t^2}. \quad (30)$$

However, 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), (4) and (5)—rather than by (3)—the limiting distribution of (30) is normal 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 E. □

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

$$\widehat{\beta} = \frac{\sum_{t=1}^T \tilde{y}_t \tilde{x}_t}{\sum_{t=1}^T \tilde{x}_t^2}, \quad (31)$$

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

$$T^{b/2}(\widehat{\beta} - \beta) = T^{b/2}(\widehat{\beta}_0 - \beta) + o_p(1). \quad (32)$$

Put differently, knowledge of α does not improve the estimation of β . Returning thus to the co-integration analogy drawn in remark 2.2, this estimation procedure is clearly reminiscent of Engle and Granger’s (1987) two-step approach to estimating co-integrating vectors and associated error correction models.

Finally, consider the ‘three-step’ estimator

$$\widehat{\delta} = \widehat{\alpha}(1 - \widehat{\beta}). \quad (33)$$

Exploiting the different convergence rates of $\widehat{\alpha}$ and $\widehat{\beta}$ (cf. proposition 2.2 and 2.3), it is readily verified that

$$T^{b/2}(\widehat{\delta} - \delta) = -\alpha T^{b/2}(\widehat{\beta} - \beta) + o_p(1); \quad (34)$$

cf. remark E.1 to appendix E, and thus

$$T^{b/2} \begin{bmatrix} \widehat{\beta} - \beta \\ \widehat{\delta} - \delta \end{bmatrix} \overset{a}{\sim} \begin{bmatrix} 1 \\ -\alpha \end{bmatrix} T^{b/2}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N} \left(0_2, \frac{2cb}{\gamma} \begin{bmatrix} 1 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix} \right), \quad (35)$$

thereby yielding the same limiting distribution as the one stated in proposition 2.1.⁴

3 Monte-Carlo results

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

$$y_t = \beta a_{t-1} x_t + \delta x_t + \varepsilon_t, \quad (36)$$

with $\delta = 1$ and $\beta = 1/3$, implying for the RE equilibrium that $\alpha = 3/2$. Two different Monte-Carlo designs will be used: design I is the ‘baseline’ setup satisfying assumptions 1–4, while design II deviates from assumption 2 and 4: first, by allowing for an autocorrelated regressor and, second, by not imposing the condition $r_t = \kappa_x^{(2)}$ on the agent’s updating scheme (5).

MONTE-CARLO DESIGN I. The regressor and the error term are drawn respectively from a t distribution with ten degrees of freedom and a χ^2 distribution with one degree of freedom, thereby ensuring that $v_t = (x_t, \varepsilon_t)'$ obeys

$$v_t \overset{i.i.d.}{\sim} \left(\begin{bmatrix} 1/\sqrt{2} \\ 0 \end{bmatrix}, \begin{bmatrix} 1/2 & 0 \\ 0 & 1 \end{bmatrix} \right). \quad (37)$$

In accordance with assumption 3, we set $r_t = \kappa_x^{(2)}$ for all t . Since equation (37) implies $\kappa_x^{(2)} = 1$, the recursion for a_t reduces to

$$a_t = a_{t-1} + \gamma_t x_t (y_t - a_{t-1} x_t) \quad \text{with } a_0 = 1. \quad (38)$$

⁴As a result of a purely algebraic exercise, the joint OLS estimator $\widehat{\lambda}$ can be recovered by using $\widehat{\lambda}_\alpha$ instead of $\widehat{\alpha}$ in order to make the regression based on (28) feasible.

MONTE-CARLO DESIGN II. The regressor evolves over time according to the first-order autoregression

$$x_t = 1 + \rho_x x_{t-1} + u_t, \quad (39)$$

where $|\rho_x| < 1$, and u_t is identically and independently drawn from a mean-zero t distribution with ten degrees of freedom. The variance of u_t and the autoregressive coefficient ρ_x are chosen such that $E(x_t) = 1/\sqrt{2}$ and $\text{var}(x_t) = 1/2$, thereby ensuring that $\kappa_x^{(2)} = 1$. Like in the first design, ε_t is an *i.i.d.* standardized χ^2 random variable with one degree of freedom. The agent uses the recursion (5) (initialized at $(a_0, r_0) = (1, 1)$) to form her beliefs about α .

In both specifications, the gain sequence is given by

$$\gamma_t = (3/2)t^{-\eta} \text{ for } \eta \in \{2/5, 4/5\}. \quad (40)$$

DISCUSSION. Monte-Carlo estimates of the asymptotic variance (asy.var), one-hundred times the bias and size of a two-sided t test at the 5% significance level are evaluated for the joint OLS (j OLS) approach (cf. section 2.2) and the three-step (3STEP) procedure (cf. section 2.5). To begin with, recall the limiting variances of the estimators of (β, δ, α) as a benchmark, against which the following simulation evidence can be judged. The preceding discussion made clear that the asymptotic variances of the estimators of (β, δ) do not depend on the moments of (x_t, ε_t) , but instead on the model parameters (γ, η, β) ; therefore, the limiting variances of the j OLS and 3STEP estimators of (β, δ) do not vary between both Monte-Carlo specifications, cf. proposition 2.1 and 2.2. Furthermore, due to the choice of the parameters, also the limiting variance $(\tau/c)^2$ of the two estimators of α , $\hat{\lambda}_\alpha$ and $\hat{\alpha}$, is invariant with respect to the two Monte-Carlo designs. Table 1 below summarizes the numerical values of the aforementioned limiting variances.

Table 1: Asymptotic variances

		η	
asy.var		2/5	4/5
β	$2cb/\gamma$	0.533	0.178
δ	$\alpha^2 2cb/\gamma$	1.200	0.400
α	$(\tau/c)^2$	2.25	

Results for $T \in \{200, 2,000, 20,000\}$ using 10,000 Monte Carlo repetitions are summarized in table 2.

Table 2: Simulation Results^a

		$\eta = 2/5$			$\eta = 4/5$			
		asy.var	bias	size	asy.var	bias	size	
DESIGN I								
β	<i>j</i> OLS	200	0.299	-2.833	3.95	0.230	-16.491	4.73
			3STEP	0.299	-2.846	4.02	0.218	-16.912
		2,000	0.397	-0.997	4.82	0.188	-9.978	5.30
			0.384	-0.998	4.83	0.186	-10.093	5.31
		20,000	0.497	-0.361	4.84	0.164	-5.938	5.00
			0.501	-0.361	4.84	0.163	-5.982	4.98
δ	<i>j</i> OLS	200	0.781	4.196	4.03	0.525	24.570	4.91
			3STEP	0.782	4.181	5.98	0.504	25.153
		2,000	0.894	1.507	4.84	0.425	14.977	5.26
			0.895	1.507	4.84	0.421	15.149	5.31
		20,000	1.079	0.548	4.91	0.368	8.915	5.03
			1.078	0.547	5.06	0.366	8.984	5.03
α	<i>j</i> OLS	200	2.003	-0.077	6.74	2.025	-0.200	12.93
			3STEP	1.998	-0.127	6.27	2.088	-0.257
		2,000	2.042	0.018	5.86	2.032	0.014	9.68
			2.044	0.008	5.70	2.036	0.011	9.70
		20,000	2.128	0.010	5.18	2.025	0.011	8.16
			2.130	0.009	5.08	2.029	0.015	8.37
DESIGN II								
β	<i>j</i> OLS	200	0.340	-3.744	4.24	0.179	-13.612	4.91
			3STEP	0.336	-3.918	4.35	0.169	-14.206
		2,000	0.429	-1.367	4.50	0.165	-9.004	4.88
			0.429	-1.383	4.51	0.163	-9.143	4.84
		20,000	0.496	-0.386	5.04	0.161	-6.125	5.09
			0.497	-0.387	5.04	0.160	-6.175	5.09
δ	<i>j</i> OLS	200	0.898	5.749	4.21	0.419	20.436	4.95
			3STEP	2.284	5.974	6.16	0.400	21.394
		2,000	1.001	2.031	4.49	0.375	13.559	4.97
			1.034	2.090	5.53	0.370	13.772	5.06
		20,000	1.106	0.592	5.00	0.382	9.198	5.20
			1.107	0.593	5.23	0.380	9.274	5.20
α	<i>j</i> OLS	200	2.056	0.190	7.37	2.507	0.023	12.39
			3STEP	3.544	0.138	7.77	2.252	0.105
		2,000	2.003	-0.029	5.47	1.964	0.070	8.91
			2.604	0.022	5.54	1.996	0.076	8.04
		20,000	2.069	0.020	4.97	2.017	0.014	6.58
			2.031	0.020	4.97	2.029	0.016	6.74

^a Size refers to the 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.

As can be seen, for large T , the Monte-Carlo estimates of the variances of the estimators are close to the values in table 1. In case of β and δ , the asymptotic approximation in terms of bias works better the smaller η . This is due to the fact that, as shown in the previous sections, the convergence rate of the estimators is decreasing in η . Observe that the bias of the estimators for δ is approximately equal to $-\alpha$ times the bias of the estimators for β , thus confirming the singularity of the joint distribution derived previously.

Finally, the generalized Wald statistic \mathcal{W} as defined in (22) is evaluated under the null

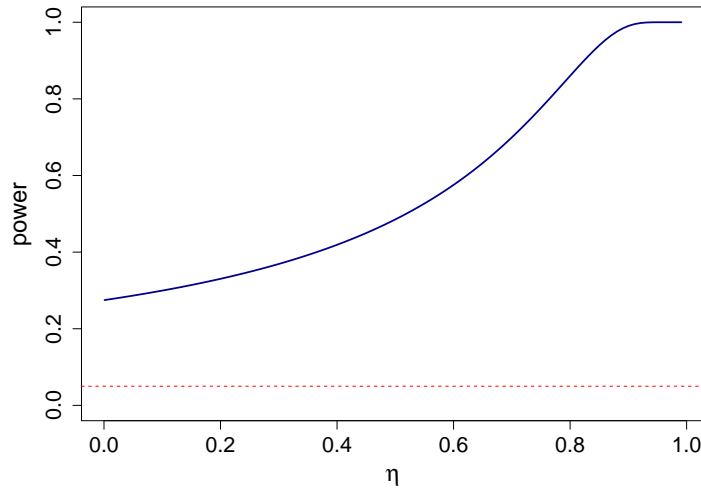
$$H_0: r(\lambda) = 0 \text{ with } r(\lambda) = R\lambda - q, \quad (41)$$

where $R = \text{diag}(1, 1)$ for $q = (1/3, 1)'$ and the local alternatives

$$H_a: r(\lambda) = \mu/T^{b/2}, \quad (42)$$

where $\mu \neq 0_2$. In the latter case, \mathcal{W} follows asymptotically a χ^2 distribution with one degree of freedom and non-centrality parameter $\mu'V^+\mu$ (see, for example, Andrews (1987)). In order to

Figure 1: Local power as a function of η^a



^a The dotted red line indicates the 5% significance level. Local power is evaluated for $\mu = (1/3, 3)'$ and various η .

understand the local power properties, it is useful to take a look at the generalized inverse of V

$$V^+ = \frac{\gamma}{2cb(1 + \alpha^2)^2} \begin{bmatrix} 1 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix}, \quad (43)$$

where equations (18) and (20) have been used.⁵ Note that V^+ is positive semi-definite, i.e. there

⁵It is readily verified that the eigenvalue and eigenvectors of V are given by $\xi = (1 + \alpha)^2 2cb/\gamma$ and

$$A = \frac{1}{\sqrt{1 + \alpha^2}} \begin{bmatrix} -1 & -\alpha \\ \alpha & -1 \end{bmatrix},$$

exist non-trivial local alternatives for which the non-centrality parameter is zero. Specifically, the test has no local power in the direction $\mu' = cV^\perp$ for some constant $c \neq 0$ as $V^\perp V^+ = 0_2'$. On the other hand, if μ is not proportional to V^\perp , power is increasing in η as illustrated by figure 1 for $\mu = (1/2, 3)'$.

Table 3: Simulation Results^a

T	$\eta = 2/5$			$\eta = 4/5$		
	size	power I	power II	size	power I	power II
DESIGN I						
200	4.88	52.06	8.10	5.14	81.52	25.91
2,000	5.11	50.00	5.62	5.34	82.75	13.46
20,000	5.00	46.91	5.35	5.03	83.31	6.59
DESIGN II						
200	7.00	49.18	11.90	5.73	46.92	30.10
2,000	6.03	47.33	8.10	5.00	61.80	15.02
20,000	5.06	43.07	5.50	4.72	72.10	7.41

^a Size and power refers to the rejection frequencies (%) at the 5% significance level using the 0.95 percentile from the χ^2 distribution with one degree of freedom under the null (41) and the (local) alternative (42), respectively.

Monte-Carlo evidence supports these theoretical considerations: table 3 reports rejection frequencies under the null (size) and local power for $\mu = (1/3, 3)'$ (power I) and $\mu = V^{\perp'}$ (power II). It is seen that the size of the test is controlled, while the local power approaches the theoretical values given by 41.91% and 86.01% in case of $\mu = (1/3, 3)'$ for $\eta = 2/5$ and $\eta = 4/5$, and 5.00% in case of $\mu = V^{\perp'}$.

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 macroeconomic model of the form

$$y_t = \beta y_{t|t-1}^e + \delta x_t + \varepsilon_t, \quad (44)$$

respectively.

where the agents expectation formation is boundedly rational in the sense that $y_{t|t-1}^e$ obeys a stochastic approximation algorithm (cf. equation (5)). In contrast to Christopheit 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 $\eta \in (0, 1)$, this intermediate case lying on a continuum between least-squares learning ($\eta = 1$) and constant-gain learning ($\eta = 0$) considered by Christopheit and Massmann (2018).

The estimators of 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 byproduct of the two estimation frameworks. Interestingly, these estimators converge at the ‘standard’ (faster) convergence rate $T^{1/2}$, and are thus not subject to the aforementioned trade-off. Moreover, it is shown how the limiting results can be used to make inferences about both single and joint hypotheses by drawing on ideas from Andrews (1987).

References

- Abadir, K. M. and Magnus, J. R. (2006). *Matrix Algebra*. Cambridge University Press.
- Andrews, D. W. K. (1987). Asymptotic results for generalized wald tests. *Econometric Theory*, 3(3):348–358.
- Apostol, T. M. (1976). *Introduction to Analytic Number Theory*. Springer.
- Apostol, T. M. (1999). An elementary view of euler’s summation formula. *The American Mathematical Monthly*, 106(5):409–418.
- Banerjee, A., Dolado, J. J., Galbraith, J. W., and Hendry, D. F. (1993). *Co-integration, Error Correction, and the Econometric Analysis of Non-stationary Data*. Oxford University Press.
- Benveniste, A., Métivier, M., and Priouret, P. (1990). *Adaptive Algorithms and Stochastic Approximation*. Springer.
- Berardi, M. and Galimberti, J. K. (2014). A note on the representative adaptive learning algorithm. *Economic Letters*, 124:104–107.
- Berardi, M. and Galimberti, J. K. (2017). Empirical calibration of adaptive learning. *Journal of Economic Behavior & Organization*, 144:219–237.
- Chakraborty, A. and Evans, G. W. (2008). Can perpetual learning explain the forward premium puzzle? *Journal of Monetary Economics*, 55(3):477–490.
- Chen, H.-F. (1993). Asymptotically efficient stochastic approximation. *Stochastics and Stochastic Reports*, 45:1–16.
- Chen, H.-F. (2002). *Stochastic Approximation and Its Applications*. Kluwer Academic Publishers.
- Christopeit, N. and Massmann, M. (2015). Estimating structural parameters in regression models with adaptive learning. *Discussion Paper 15-106/III, Tinbergen Institute*.
- Christopeit, N. and Massmann, M. (2018). Estimating structural parameters in regression models with adaptive learning. *Econometric Theory*, 34(1):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(3):402–412.

- Davidson, J. (1994). *Stochastic Limit Theory*. Oxford University Press.
- Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(22):251–276.
- Evans, G. W. and Honkapohja, S. (2001). *Learning and Expectations in Macroeconomics*. Princeton University Press.
- Hannan, E. J. (1970). *Multiple Time Series*. John Wiley and Sons.
- Hassler, U. and Wolters, J. (2006). Autoregressive distributed lag models cointegration. *Advances in Statistical Analysis*, 90(1):59–74.
- Jameson, G. J. (2016). The incomplete gamma functions. *The Mathematical Gazette*, 100(548):298–306.
- Kushner, H. J. and Huang, H. (1981). Asymptotic properties of stochastic approximations with constant coefficients. *SIAM Journal on Control and Optimization*, 19(1):87–105.
- Kushner, H. J. and Yan, J. (1993). Stochastic approximation with averaging of the iterates: Optimal asymptotic rate of convergence for general processes. *Siam Journal on Control and Optimization*, 31(4):1045–1062.
- Lampret, V. (2001). The euler–maclaurin and taylor formulas: Twin, elementary derivations. *Mathematics Magazine*, 74(2):109–122.
- Lütkepohl, H. and Burda, M. M. (1997). Modified wald tests under nonregular conditions. *Journal of Econometrics*, 78(2):315–332.
- Markiewicz, A. and Pick, A. (2017). Adaptive learning and survey data. *Journal of Economic Behavior & Organization*, 107:685–707.
- Pagan, A. (1984). Econometric issues in the analysis of regressions with generated regressors. *International Economic Review*, 25(1):221–247.
- Pesaran, M. H. and Shin, Y. (1998). An autoregressive distributed lag modelling approach to cointegration analysis. In Strom, S., editor, *Econometrics and Economic Theory in the 20th Century. The Ragnar Frisch Centennial Symposium*, pages 371–413. Cambridge University Press, Cambridge.
- Phillips, P. C. (2007). Regression with slowly varying regressors and nonlinear trends. *Econometric Theory*, 23(4):557–614.

- Polyak, B. T. and Juditsky, A. B. (1992). Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization*, 30(4):838–855.
- Sargent, T. J. (1993). *Bounded Rationality in Macroeconomics*. Oxford University Press.
- Seber, G. A. F. (2008). *A Matrix Handbook for Statisticians*. Wiley.
- Stock, J. H. (1987). Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica*, 55(5):1035–1056.
- Thompson, I. (2013). Algorithm 026: Incomplete gamma functions with negative arguments. *ACM Transactions on Mathematical Software*, 39(14):1–9.
- Wickens, M. R. and Breusch, T. S. (1988). Dynamic specification, the long-run and the estimation of transformed regression models. *The Economic Journal*, 98(390):198–205.
- Winitzki, S. (2003). Computing the incomplete gamma function to arbitrary precision. In Goos, G., Hartmanis, J., and van Leeuwen, J., editors, *Proceedings of the International Conference on Computational Science and Its Applications (ICCSA'03)*.

Appendices

The outline of the appendix is as follows: Appendix A establishes six lemmata which are used throughout the remaining parts. The other appendices contain the proofs of the main results of the main text.

A Preliminaries	23
A.1 Working formula for a_t	24
A.2 Lemma A.1	24
A.3 Lemma A.2	25
A.4 Lemma A.3	36
A.5 Lemma A.4	39
A.6 Lemma A.5	40
A.7 Lemma A.6	46
B Proof of proposition 2.1	47
C Proof of corollary 2.1	51
D Proof of proposition 2.2	52
E Proof of proposition 2.3	52
F Extension A: multiple regressors	56
G Extension B: constant gain	66

A Preliminaries

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

A.1 Working formula for a_t

With a little rearrangement, one gets

$$\begin{aligned} a_t^* &= a_{t-1}^* + \frac{\gamma_t}{\kappa_x^{(2)}} x_t (y_t - a_{t-1} x_t) \\ &= a_{t-1}^* (1 - c\gamma_t) + \gamma_t e_t + \gamma_t u_t, \end{aligned} \tag{A.1}$$

where $a_t^* := a_t - \alpha$ while e_t and u_t are respectively given by

$$e_t := c(1 - x_t^{*2}) a_{t-1}^* \quad \text{with} \quad x_t^* := x_t / \sqrt{\kappa_x^{(2)}} \tag{A.2}$$

$$u_t := x_t^* \varepsilon_t / \sqrt{\kappa_x^{(2)}}, \tag{A.3}$$

which are, by construction, mean-zero. Solving (A.1) recursively, yields for $t > 0$

$$a_t^* = a_0^* \Phi(t, 1) + \check{u}_t + \check{e}_t, \tag{A.4}$$

where

$$\Phi(t, k) := \prod_{i=k}^t (1 - c\gamma_i) \quad \text{with} \quad \Phi(t, t+1) := 1, \tag{A.5}$$

while for any scalar z_t

$$\check{z}_t := \sum_{k=1}^t \gamma(k, t) z_k \quad \text{with} \quad \gamma(k, t) := \Phi(t, k+1) \gamma_k. \tag{A.6}$$

A.2 Lemma A.1

Lemma A.1. *For any s and t*

$$\mathbb{E}(u_t e_s) = 0, \tag{A.7}$$

while

$$\mathbb{E}(e_t e_s) = \begin{cases} 0 & \text{if } s \neq t \\ \sigma^2(e_t) := c^2(\kappa_x^{(4)} / \kappa_x^{(2)2} - 1) \mathbb{E}(a_{t-1}^{*2}) & \text{if } s = t. \end{cases} \tag{A.8}$$

Proof. Note that $\mathbb{E}(u_t e_s) = c [\mathbb{E}(u_t a_{s-1}^*) - \mathbb{E}(u_t x_s^{*2} a_{s-1}^*)]$. It thus suffices to show that

$$\mathbb{E}(u_t a_{s-1}^*) = \mathbb{E}(u_t x_s^{*2} a_{s-1}^*).$$

Clearly, $E(u_t a_{s-1}^*) = E(u_t x_s^{*2} a_{s-1}^*) = 0$ for all $s \leq t$. For $s > t$, observe that

$$\begin{aligned} E(u_t x_s^{*2} a_{s-1}^*) &= E\left[u_t a_{s-1}^* E(x_s^{*2} | \mathcal{V}_{s-1})\right] \\ &= E(u_t a_{s-1}^*) E(x_s^{*2}) = E(u_t a_{s-1}^*), \end{aligned} \quad (\text{A.9})$$

where

$$\mathcal{V}_t := \sigma(a_0; v_s, s \leq t). \quad (\text{A.10})$$

Equation (A.8) is a direct consequence of the independence assumption of x_1, \dots, x_t . \square

A.3 Lemma A.2

This lemma summarizes properties of partial sums of the coefficients $\gamma(k, t)$ (cf. equation (A.6)), which are frequently used in the subsequent proofs for the decreasing gain case.

Lemma A.2. *Define*

$$\phi_t^i := \sum_{k=1}^t \gamma(k, t) \quad (\text{A.11})$$

$$\phi_t^{ii} := \sum_{k=1}^t \gamma(k, t)^2 \quad (\text{A.12})$$

$$\phi_t^{iii} := \sum_{k=1}^{t-1} \sum_{s=1}^k \gamma(s, t) \gamma(s, k) \quad (\text{A.13})$$

and suppose that $\beta \in (0, 1)$, $\eta \in (0, 1)$. Then,

- (a) $\phi_t^i = 1/c + o(1)$
- (b) $\phi_t^{ii} = \gamma_t/(2c) + o(\gamma_t)$
- (c) $\phi_t^{iii} = 1/(2c^2) + o(1)$.

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

A.3.1 Part (a)

Proof. Note that

$$c\phi_t^i + \Phi(t, 1) = 1. \quad (\text{A.14})$$

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

$$\mathcal{H}_n(\eta) := \sum_{i=1}^n 1/i^\eta$$

can be written as

$$\mathcal{H}_t(\eta) = \frac{t^{1-\eta}}{1-\eta} + \zeta(\eta) + \eta R_t^\infty(\eta) \text{ with } R_a^b(\eta) := \int_a^b \frac{u - [u]}{u^{1+\eta}} du, \quad (\text{A.15})$$

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

$$\zeta(\eta) := \lim_{t \rightarrow \infty} \left(\mathcal{H}_t(\eta) - \frac{t^{1-\eta}}{1-\eta} \right), \quad (\text{A.16})$$

$[x]$ is the greatest integer smaller or equal x and

$$0 \leq R_t^\infty(\eta) \leq \int_t^\infty \frac{1}{u^{1+\eta}} du = \frac{1}{\eta t^\eta};$$

see, e.g., theorem 3.2 b) in Apostol (1976, chapter 3). 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. $\mathcal{H}_t := \mathcal{H}_t(\eta)$ and $R_a^b := R_a^b(\eta)$.

The claim follows because for any $r \geq 1$

$$\Phi(t, 1)^r \leq C_0 e^{-art^b} \quad (\text{A.17})$$

where $C_0 \in [0, \infty)$, $a := \tilde{c}/b$, $b := 1 - \eta$ and $\tilde{c} := c\gamma$. To see that (A.17) is true, consider

$$\begin{aligned} \Phi(t, 1)^r &= e^{r \sum_{j=1}^t \ln(1 - \tilde{c}/j^\eta)} \\ &= e^{-\tilde{c}r \sum_{j=1}^t 1/j^\eta} e^{-O(\sum_{j=1}^t 1/j^{2\eta})} \\ &\leq e^{-\tilde{c}r \mathcal{H}_t(\eta)} \\ &= e^{-\tilde{c}r \zeta(1-b)} e^{-art^b} e^{-O(t^{-\eta})} \\ &\leq e^{-\tilde{c}r \zeta(1-b)} e^{-art^b}, \end{aligned}$$

where $e^{-\tilde{c}r \zeta(1-b)} \geq 0$ as $\zeta(1-b) \leq 0$. The second and third equality follow, respectively, from the Taylor series approximation

$$\ln(1 - \tilde{c}/j^\eta) = -\tilde{c}/j^\eta - K_t \text{ with } K_t := \tilde{c}^2/(2j^{2\eta}) + \tilde{c}^3/(3j^{3\eta}) + \dots = O(1/j^{2\eta}), \quad (\text{A.18})$$

and equation (A.15), while the inequalities result upon recognizing that the terms $O(\sum_{j=1}^t 1/j^{2\eta})$ and $O(t^{-\eta})$ are both positive.⁶ \square

⁶Note that the conclusion of part (a) holds also for $\eta = 1$. Specifically, $\gamma_{k,t}$ reduces for $\eta = 1$ to $\gamma_{k,t} =$

A.3.2 Part (b)

The main idea of this proof is to approximate the partial sum

$$c^2 \phi_t^{ii} = \sum_{k=1}^t g(k, t) \quad \text{with} \quad g(k, t) := [c\gamma_k \Phi(t, k+1)]^2 \quad (\text{A.20})$$

by the integral of the function

$$f(k, t) := [c\gamma_k e^{-a(t^b - k^b)}]^2. \quad (\text{A.21})$$

Recall that $a = \tilde{c}/b$; hence, $f(k, t)$ is only well-defined if $\eta < 1$. To begin with, consider

$$c^2 \phi_t^{ii} = \int_1^t f(k, t) dk + A_t + B_t, \quad (\text{A.22})$$

where

$$A_t := \sum_{k=1}^t f(k, t) - \int_1^t f(k, t) dk \quad \text{and} \quad B_t := \sum_{k=1}^t [g(k, t) - f(k, t)]. \quad (\text{A.23})$$

The remainder of this proof is as follows: (1) evaluate the integral $\int f(k, t) dk$, (2) show that $A_t = O(\gamma_t^2)$, (3) show that $B_t = O(\gamma_t^2)$.

(1) *Evaluation of the integral.* We seek to establish

$$\int_1^t f(k, t) dk = \frac{c\gamma_t}{2} + o(\gamma_t). \quad (\text{A.24})$$

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

$$\lim_{t \rightarrow \infty} t^{\eta(s-1)} \int_1^t e^{-rz/b(t^b - k^b)} (z/k^{1-b})^s dk = \frac{z^{s-1}}{r}. \quad (\text{A.25})$$

Hence, (A.24) follows from (A.25) with $s = r = 2$ and $z = \tilde{c}$. Before (A.25) is proven, recall the definition of the (upper) incomplete gamma function

$$\Gamma(s, x) = \int_x^\infty t^{s-1} e^{-t} dt, \quad (\text{A.26})$$

$\lambda_t(1-c, 1)\lambda_k(0, 1-c)$ where

$$\lambda_t(\alpha, \beta) := \frac{\Gamma(t+\alpha)}{\Gamma(t+\beta)}, \quad (\text{A.19})$$

where $\Gamma(\cdot)$ denotes the gamma function.

see Jameson (2016) for details. Note that the definition of $\Gamma(s, x)$ extends to arbitrary (possibly complex) s, x ; see, e.g. Winitzki (2003) or Thompson (2013). The proof repeatedly makes use of the following: For any $x, v > 0$

$$\lim_{t \rightarrow \infty} t^{1+v} \Gamma(-1/v, -xt^v) e^{-xt^v} = -x^{(1-v)/v} e^{-i\pi/v}, \quad (\text{A.27})$$

$$\int e^{-xt^v} dt = -\frac{\Gamma(1/v, xt^v)}{vx^{1/v}} + C, \quad (\text{A.28})$$

$$\int e^{xt^v} dt = -\frac{\Gamma(1/v, -xt^v)}{vx^{1/v}} e^{-i\pi/v} + C, \quad (\text{A.29})$$

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

Proof of (A.27)-(A.29). Define $w(t) := \Gamma(-1/v, -xt^v)$ and $q(t) := t^{-1-v} e^{xt^v}$ and denote the first derivatives with respect to t by

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

so that⁷

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

Then, by L'Hôpital's rule,

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

This completes the proof of (A.27). Turning to (A.28), use the u -substitution $u := xt^v$ and observe that $t^{1-v} = u^{1/v-1} x^{1-1/v}$. Hence,

$$\int e^{-xt^v} dt = \frac{1}{vx^{1/v}} \int e^{-u} u^{1/v-1} du = -\frac{1}{vx^{1/v}} \int_u^\infty e^{-k} k^{1/v-1} dk + C, \quad (\text{A.33})$$

where the last equality uses that for any integrable function f $\int f(x) dx = \int_a^x f(t) dt + C$ with a so that the integral converges; thereby proving the claim. (A.29) follows directly from (A.28) upon using the u -substitution $u := ((-1)x)^{1/v} t$. \square

Proof of (A.25). To begin with, note that (ignoring the constant of integration)

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

⁷The convention $(-1)^a = e^{i\pi a}$ for $a > 0$ is used throughout.

where $C_0 := b^{-\eta(s-1)/b} r^{(s\eta-1)/b} z^{(s-1)/b} e^{i\pi\eta(s-1)}$ and $m(k, t) := e^{-t^b r z/b} \Gamma((1-s\eta)/b, -k^b r z/b)$. Hence,

$$\int_1^t e^{-r z/b(t^b-k^b)} (z/k^{1-b})^s dk = C_0 [m(t, t) - m(1, t)] = C_0 m(t, t) + O(e^{-r z/b t^b}). \quad (\text{A.35})$$

Next, it will be shown that

$$\lim_{t \rightarrow \infty} t^{\eta(s-1)} m(t, t) = \frac{C_1}{r z}, \quad (\text{A.36})$$

where $C_1 := b^{\eta(s-1)/b} (r z)^{(1-s\eta)/b} e^{i\pi(b(1+s)+1-s)/b}$. By L'Hôspital's rule,

$$\lim_{t \rightarrow \infty} t^{\eta(s-1)} m(t, t) = \lim_{t \rightarrow \infty} \frac{h_1'(t)}{h_2'(t)} \quad (\text{A.37})$$

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

$$h_1'(t) = C_1 t^{-s\eta} e^{t^b r z/b} \quad (\text{A.38})$$

$$h_2'(t) = r z t^{-\eta(s-1)-\eta} e^{t^b r z/b} \left(1 + \frac{(b-1)(s-1)}{r z t^b} \right), \quad (\text{A.39})$$

thereby verifying (A.36). Some simple calculations reveal further that $C_0 C_1 / (r z) = z^{s-1} / r$. \square

(2) *Negligibility of A_t* . Note that one version of the Euler-Maclaurin formula (see, e.g., Lampret (2001)) states that

$$\sum_{k=1}^t f(k, t) - \int_1^t f(k, t) dk = \frac{1}{2} [f(1, t) + f(t, t)] + \frac{1}{12} [f^{(1)}(t, t) - f^{(1)}(1, t)] + \rho(t; f), \quad (\text{A.40})$$

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

$$|\rho(t; f)| \leq \frac{1}{120} \int_1^t |f^{(3)}(k, t)| dk. \quad (\text{A.41})$$

It will be shown that the three terms on the right-hand side of (A.40) are of order $o(\gamma_t)$. Clearly, $f(1, t) + f(t, t) = O(\gamma_t^2)$ while from

$$f^{(1)}(k, t) = 2(ck^b - \eta)k^{-1} f(k, t) \quad (\text{A.42})$$

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

$$f^{(3)}(k, t) = 2 \left[4(ck^b)^3 - 18(ck^b)^2(1-b) - ck^b(1-b)(19b-26) - 2(2-b)(1-b)(3-2b) \right] \frac{f(k, t)}{k^3}$$

implies the existence of a finite constant $C_0 > 0$ such that $|f^{(3)}(k, t)| \leq C_0 (c\gamma_k)^3 f(k, t)$. Hence, by equation (A.25), $|\rho(t; f)| = O(\gamma_t^4)$ – thereby showing that $A_t = O(\gamma_t^2)$. \square

(3) *Negligibility of B_t* . Taking account of equation (A.15) and $R_t^\infty = R_k^\infty - R_k^t$, it is seen that

$$\sum_{j=k+1}^t \gamma_j = \gamma(\mathcal{H}_t(\eta) - \mathcal{H}_k(\eta)) = \gamma(t^b - k^b)/b - \gamma\eta R_k^t(\eta). \quad (\text{A.43})$$

Now, using equation (A.18) together with (A.43), it follows that

$$\begin{aligned} g(k, t) &= (c\gamma_k)^2 e^{2\sum_{j=k+1}^t \ln(1-c\gamma_j)} \\ &= (c\gamma_k)^2 e^{-2\bar{c}(\mathcal{H}_t - \mathcal{H}_k)} e^{-2\sum_{j=k+1}^t K_j} \\ &= (c\gamma_k)^2 e^{-2\bar{c}(\mathcal{H}_t - \mathcal{H}_k)} + (c\gamma_k)^2 e^{-2\bar{c}(\mathcal{H}_t - \mathcal{H}_k)} \left(e^{-2\sum_{j=k+1}^t K_j} - 1 \right) \\ &= f(k, t) + f(k, t) \left(e^{2\bar{c}\eta R_k^t} - 1 \right) \\ &\quad + f(k, t) \left(e^{-2\sum_{j=k+1}^t K_j} - 1 \right) \\ &\quad + f(k, t) \left(e^{2\bar{c}\eta R_k^t} - 1 \right) \left(e^{-2\sum_{j=k+1}^t K_j} - 1 \right) \\ &=: f(k, t) + B^i(k, t) + B^{ii}(k, t) + B^{iii}(k, t), \end{aligned} \quad (\text{A.44})$$

say. It thus suffices to show that $\sum_k B^\ell(k, t) = O(\gamma_t^2)$ 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 $e^{2c\eta R_k^t} \geq 1$, it follows that $B^i(k, t) \geq 0$. Infer from the discussion surrounding (A.15) that $R_k^t \leq \gamma_k$ and thus $B^i(k, t) \leq f(k, t) (e^{2c\eta R_k^t} - 1)$. Next, suppose without loss of generality that $2c\eta R_k^t \leq \alpha < 1$. Because $e^x < (1-x)^{-1}$ for all $x < 1$, it follows $B^i(k, t) \leq 2(1-\alpha)^{-1} f(k, t) c\eta R_k^t$. By (A.25) 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

$$|B^{ii}(k, t)| = f(k, t) \left(1 - e^{-2\sum_{j=k+1}^t K_j(\eta)} \right). \quad (\text{A.45})$$

The definition of $K_i(\eta)$ (cf. equation (A.18)) implies that there exists a finite constant $C_0 > 0$ so that $\sum_{j=k+1}^t K_j(\eta) \leq 2^{-1} C_0 \sum_{j=k+1}^t j^{-2\eta}$, while $\sum_{j=k+1}^t j^{-2\eta} = \ell(k, t) - 2\eta R_k^t(2\eta)$ with

$$\ell(k, t) = \begin{cases} (1-2\eta)^{-1} (t^{1-2\eta} - k^{1-2\eta}) & \text{if } \eta < 1/2 \\ (2\eta-1)^{-1} (k^{1-2\eta} - t^{1-2\eta}) & \text{if } \eta > 1/2 \\ \ln t - \ln k & \text{if } \eta = 1/2, \end{cases} \quad (\text{A.46})$$

where for $\eta \geq 1/2$ Apostol (1976, theorem 3.2 (a), (b)) has been used and the case $\eta < 1/2$ follows immediately from (A.43). Hence, for some finite constant $C_1 > 0$

$$\begin{aligned} 1 - e^{-2\sum_{j=k+1}^t K_j(\eta)} &\leq 1 - e^{-C_0 \sum_{j=k+1}^t j^{-2\eta}} \\ &= 1 - e^{-C_0 \ell(k,t)} e^{C_0 2\eta R_k^t(2\eta)} \\ &\leq 1 - e^{-C_0 \ell(k,t)} \\ &\leq C_0 \ell(k,t). \end{aligned} \tag{A.47}$$

The second inequality uses that $2\eta R_k^t(2\eta) \leq \ell(k,t)$ while the final one is an elementary property of the exponential function. Since

$$\lim_{t \rightarrow \infty} t^{2\eta} \int_1^t f(k,t) \ell(k,t) dk = \frac{1}{4}, \tag{A.48}$$

it follows from the integral comparison test that $\sum_{k=1}^t B^{ii}(k,t) = O(\gamma_t^2)$. Finally, $\sum_k B^{iii}(k,t) = O(\gamma_t^2)$ is deduced from the analysis of $\sum_k B^{ii}(k,t)$ as

$$\left| 1 - e^{-2\sum_{j=k+1}^t K_j(\eta)} \right| \leq 1. \tag{A.49}$$

This completes the proof of part (b). \square

Remark. More generally, it can be shown that for any $\alpha > 0$ and $\beta \geq 0$

$$\sum_{k=1}^t \gamma(k,t)^\alpha \gamma_k^\beta = O(\gamma_t^{\alpha+\beta-1}). \tag{A.50}$$

Proof of equation A.50. First, note that with $r = \alpha$, $s = \alpha + \beta$ and $z = \tilde{c}$ it follows from (A.25) that

$$\lim_{t \rightarrow \infty} t^{(\alpha+\beta-1)(1-b)} \int_1^t e^{-\alpha a(t^b-k^b)} \gamma_k^{\alpha+\beta} dk = \frac{\tilde{c}^{\alpha+\beta-1}}{\alpha + \beta}. \tag{A.51}$$

Next, note that

$$\begin{aligned} \Phi(t, k+1)^\alpha &= e^{\alpha \sum_{i=k+1}^t \ln(1-c\gamma_i)} \\ &= e^{-\alpha \tilde{c}(H_t(\eta) - H_k(\eta))} e^{-\tilde{c}\alpha \sum_{i=k+1}^t K_i(\eta)} \\ &\leq e^{-\alpha \tilde{c}(H_t(\eta) - H_k(\eta))} \\ &= e^{-\alpha a(t^b-k^b)} + e^{-\alpha a(t^b-k^b)} \left(1 - e^{\alpha \tilde{c}\eta R_k^t(\eta)} \right) \end{aligned} \tag{A.52}$$

It follows from Toeplitz's lemma, the integral comparison test and the use of similar arguments used to show the asymptotic negligibility of $B^i(k,t)$ that

$$\sum_{k=1}^t \gamma(k,t)^\alpha \gamma_k^\beta \leq \int_1^t e^{-\alpha a(t^b-k^b)} \gamma_k^{\alpha+\beta} dk. \tag{A.53}$$

Taking equation (A.51) into account completes the proof of (A.50).

A.3.3 Part (c)

Without loss of generality, assume $\gamma = 1$. To begin with, observe that one gets from part (a) $(\phi_t^i)^2 \rightarrow c^{-2}$. But,

$$(\phi_t^i)^2 = \sum_{k,s=1}^t \gamma(k,t)\gamma(s,t) = \phi_t^{ii} + 2 \sum_{k=2}^t \sum_{s=1}^{k-1} \gamma(k,t)\gamma(s,t). \quad (\text{A.54})$$

Consequently, $2\phi_t^{iii} = (\phi_t^i)^2 + 2R_t - \phi_t^{ii}$, where

$$R_t := \phi_t^{iii} - \sum_{k=2}^t \sum_{s=1}^{k-1} \gamma(s,k)\gamma(s,t). \quad (\text{A.55})$$

The claim follows by part (b) if $R_t = o(1)$. Investigation of R_t reveals that this term decomposes as $R_t = A_t + B_t + C_t$, where

$$\begin{aligned} A_t &:= \sum_{k=1}^{t-2} \gamma(k,t) \sum_{j=k+1}^{t-1} [\gamma_k \Phi(j, k+1) - \gamma_j \Phi(t, j+1)] \\ B_t &:= \sum_{k=1}^{t-2} \gamma(k,t)(\gamma_k - \gamma_t) \\ C_t &:= (1 - c\gamma_t)\gamma_{t-1}(\gamma_{t-1} - \gamma_t). \end{aligned}$$

Clearly, $C_t = o(1)$, while $B_t = o(1)$ is a direct consequence of (A.50). It thus remains to show that $A_t = o(1)$.

Before the proof proceeds, three useful results from the literature on stochastic approximation with averaging of the iterates are stated:

Suppose γ_t satisfies assumption 1, $j \geq k$ and $v \in (1/2, 1)$. There exist finite and positive constants $C_0 > 0$, $C_1 > 0$ and $C_2 \geq 0$ so that

$$\begin{aligned} \gamma_k/\gamma_j &\leq e^{o(1)\sum_{s=k}^j \gamma_s} \\ \Phi(j, k+1) &\leq C_0 e^{-C_1 \sum_{s=k}^j \gamma_s} \\ \sup_{k,t} \sum_{j=k}^t j^{-v} e^{-\lambda \sum_{s=k}^j s^{-v}} &\leq C_2 < \infty \text{ for } \lambda > 0, \end{aligned} \quad (\text{A.56})$$

with $o(1)$ denoting a magnitude that approaches zero as $k \rightarrow \infty$.

Proof of (A.56). See, for example Chen (2002, lemma 3.1.1 & 3.4.1) or Polyak and Juditsky (1992). \square

To begin with, note that

$$\sum_{j=k+1}^{t-1} \Phi(t, j+1) \geq \sum_{i=k+1}^{t-1} \Phi(i, k+1). \quad (\text{A.57})$$

Therefore,

$$|A_t| \leq \sum_{k=1}^{t-2} \gamma(k, t) \sum_{j=k+1}^{t-1} (\gamma_k - \gamma_j) \Phi(j, k+1). \quad (\text{A.58})$$

Since $\lim_{t \rightarrow \infty} t^{1+\eta} (\gamma_{t-1} - \gamma_t) = \gamma\eta$, one arrives at

$$\begin{aligned} \sum_{j=k+1}^{t-1} (\gamma_k - \gamma_j) \Phi(j, k+1) &= \sum_{j=k+1}^{t-1} \sum_{i=k+1}^j (\gamma_{i-1} - \gamma_i) \Phi(j, k+1) \\ &= \sum_{j=k+1}^{t-1} \sum_{i=k+1}^j O\left(\frac{\gamma_i}{i}\right) \Phi(j, k+1). \end{aligned} \quad (\text{A.59})$$

Hence, there exists a finite constant $C_0 > 0$ such that

$$|A_t| \leq C_0 \sum_{k=1}^{t-2} \gamma_{k,t} \gamma_k \sum_{j=k+1}^{t-1} \sum_{i=k+1}^j \frac{\Phi(j, k+1)}{i}. \quad (\text{A.60})$$

Consider first the case of $\eta \neq 1/2$. There exist constants C_1 and C_2 such that

$$\begin{aligned} \sum_{j=k+1}^{t-1} \sum_{i=k+1}^j \frac{\Phi(j, k+1)}{i} &= \sum_{j=k+1}^{t-1} \gamma_j \sum_{i=k+1}^j \frac{\gamma_i \Phi(j, k+1)}{\gamma_j i^b} \\ &\leq \sum_{j=k+1}^{t-1} \gamma_j \sum_{i=k+1}^j \frac{\gamma_{k+1} \Phi(j, k+1)}{\gamma_j i^b} \leq C_1 \sum_{j=k+1}^{t-1} \gamma_j \sum_{i=k+1}^j \frac{e^{-C_2/2 \sum_{s=k+1}^j \gamma_s}}{i^b}, \end{aligned} \quad (\text{A.61})$$

where the first inequality is due to $\gamma_i \leq \gamma_{k+1}$ and the final one uses the first part of (A.56) to approximate γ_{k+1}/γ_j .

Case $\eta < 1/2$. (A.61) can be further bounded by

$$\begin{aligned}
\sum_{j=k+1}^{t-1} \gamma_j \sum_{i=k+1}^j \frac{e^{-C_2/2 \sum_{s=k+1}^j \gamma_s}}{j^b} &= \sum_{j=k+1}^{t-1} \sum_{i=k+1}^j \left(\frac{j}{i}\right)^{1-2\eta} \gamma_i \frac{e^{-C_2/2 \sum_{s=k+1}^j \gamma_s}}{j^b} \\
&\leq \frac{4}{C_2} \left(\frac{t}{k}\right)^{1-2\eta} \sum_{j=k+1}^{t-1} \left(e^{C_2/4 \sum_{i=k+1}^j \gamma_i}\right) \frac{e^{-C_2/2 \sum_{s=k+1}^j \gamma_s}}{j^b} \\
&= \frac{4}{C_2} \left(\frac{t}{k}\right)^{1-2\eta} \sum_{j=k+1}^{t-1} \frac{e^{-C_2/4 \sum_{s=k+1}^j \gamma_s}}{j^b} \\
&\leq \frac{4}{C_2} \left(\frac{t}{k}\right)^{1-2\eta} \sum_{j=k+1}^{t-1} \frac{e^{-C_2/4 \sum_{s=k+1}^j s^{-b}}}{j^b} \leq C_3 \left(\frac{t}{k}\right)^{1-2\eta}, \quad (\text{A.62})
\end{aligned}$$

where C_3 represents a finite constant; the first inequality is due to $j/i \leq t/k$ and $x \leq e^x$ ($x > 0$), the second inequality used that for $\eta < 1/2$ and $s \geq 1$, $\gamma_s \geq s^{-b}$ while the final one is based on the last property of (A.56). Hence, there exists a finite constant $K > 0$ such that

$$|A_t| \leq K t^{1-2\eta} \sum_{k=1}^t \frac{\gamma_{k,t}}{k^b} = K t^{1-2\eta} \sum_{k=1}^t \frac{\Phi(t, k+1)}{k}. \quad (\text{A.63})$$

By equation (A.52),

$$\sum_{k=1}^t \frac{\Phi(t, k+1)}{k} \leq \sum_{k=1}^t \frac{e^{-ar(t^b - k^b)}}{k} + \sum_{k=1}^t \frac{e^{-ar(t^b - k^b)}}{k} \left(e^{cr\eta R_k^t} - 1\right). \quad (\text{A.64})$$

Furthermore,

$$\int_1^t \frac{e^{-a(t^b - k^b)}}{k} dk = e^{-at^b} \left(\frac{\text{Ei}(at^b) - \text{Ei}(a)}{b}\right) \quad (\text{A.65})$$

where

$$\text{Ei}(x) := - \int_{-x}^{\infty} \frac{e^{-t}}{t} dt \quad (\text{A.66})$$

denotes the exponential integral. Because for any $a, v > 0$

$$\lim_{t \rightarrow \infty} t^v e^{-at^v} \text{Ei}(at^v) = 1/a, \quad (\text{A.67})$$

it follows from the integral comparison test and Toeplitz's lemma (see also section A.3.2) that $A_t = O(\gamma_t)$. This completes the proof for the case $\eta < 1/2$. Case $\eta > 1/2$. Using $i^{-b} = i^{2\eta-1}/i^\eta$,

it is seen that (A.61) can similarly be bounded by

$$\begin{aligned} \sum_{j=k+1}^{t-1} \gamma_j \sum_{i=k+1}^j i^{2\eta-1} \gamma_i e^{-C_2/2 \sum_{s=k+1}^j \gamma_s} &\leq t^{2\eta-1} \sum_{j=k+1}^{t-1} \gamma_j \sum_{i=k+1}^j \gamma_i e^{-C_2/2 \sum_{s=k+1}^j \gamma_s} \\ &\leq t^{2\eta-1} C_3, \end{aligned} \quad (\text{A.68})$$

where the existence of some finite constant $C_3 \geq 0$ is ensured by the last property of (A.56). Hence,

$$|A_t| \leq C_3 t^{2\eta-1} \sum_{k=1}^t \gamma_{k,t} \gamma_k \quad (\text{A.69})$$

and by equation (A.50), $A_t = O(\gamma_t)$.

Case $\eta = 1/2$. For $\epsilon \in (0, 1/4)$ and suitable constants C_0, C_1 and C_2 one gets

$$\begin{aligned} \sum_{j=k+1}^{t-1} \sum_{i=k+1}^j \frac{\Phi(j, k+1)}{i} &\leq C_0 \sum_{j=k+1}^{t-1} \sum_{i=k+1}^j \frac{e^{-C_1 \sum_{s=k+1}^j 1/s^{1/2}}}{i} \\ &\leq C_0 \sum_{j=k+1}^{t-1} \sum_{i=k+1}^j \frac{e^{-C_1 \sum_{s=k+1}^j 1/s^{1/2+\epsilon}}}{i} \\ &= C_0 \sum_{j=k+1}^{t-1} \frac{1}{j^{1/2+\epsilon}} \sum_{i=k+1}^j \left(\frac{j}{i}\right)^{1/2+\epsilon} \frac{e^{-C_1 \sum_{s=k+1}^j 1/s^{1/2+\epsilon}}}{i^{1/2-\epsilon}} \\ &\leq C_0 \sum_{j=k+1}^{t-1} \frac{1}{j^{1/2+\epsilon}} \sum_{i=k+1}^j \left(\frac{j}{k+1}\right)^{1/2+\epsilon} \frac{e^{-C_1 \sum_{s=k+1}^j 1/s^{1/2+\epsilon}}}{i^{1/2-\epsilon}} \\ &\leq C_0 \sum_{j=k+1}^{t-1} \frac{1}{j^{1/2+\epsilon}} \sum_{i=k+1}^j \frac{e^{-C_1/2 \sum_{s=k+1}^j 1/s^{1/2+\epsilon}}}{i^{1/2-\epsilon}} \\ &= C_0 \sum_{j=k+1}^{t-1} \frac{1}{j^{1/2+\epsilon}} \sum_{i=k+1}^j i^{2\epsilon} \frac{e^{-C_1/2 \sum_{s=k+1}^j 1/s^{1/2+\epsilon}}}{i^{1/2+\epsilon}} \\ &\leq t^{2\epsilon} \frac{4C_0}{C_1} \sum_{j=k+1}^{t-1} \frac{e^{-C_1/4 \sum_{s=k+1}^j 1/s^{1/2+\epsilon}}}{j^{1/2+\epsilon}} \leq t^{2\epsilon} C_2, \end{aligned} \quad (\text{A.70})$$

where the first inequality is due to the first property of (A.56), the second uses that $1/s^{1/2} > 1/s^{1/2+\epsilon}$, the third follows from $i \geq k+1$, the fourth is based on the first property of (A.56) while the final inequality uses the last property of (A.56). Thus, taking equation (A.50) into account, it follows that

$$|A_t| \leq C_2 t^{2\epsilon} \sum_{k=1}^t \gamma(k, t) \gamma_k = O\left(\frac{1}{t^{1/2-2\epsilon}}\right). \quad (\text{A.71})$$

This completes the proof of part (c). □

A.4 Lemma A.3

Lemma A.3. *Recall that $\tau = \sigma/\sqrt{\kappa_x^{(2)}}$. Then*

$$\mathbb{E}(a_t^{*2}) = O(\gamma_t) \tag{a}$$

$$\mathbb{E}(a_t^{*4}) = O(\gamma_t^2), \tag{b}$$

and

$$\lim_{T \rightarrow \infty} \frac{1}{T^b} \sum_{t=1}^T \mathbb{E}(a_t^{*2}) = \frac{\tau^2 \gamma}{2cb}. \tag{c}$$

Proof of equation (a). The claim follows from Benveniste et al. (1990, theorem 22, part 2). Their theorem provides an L_2 -upper bound for Robbins-Monro algorithms; cf. Benveniste et al. (1990, pp. 243). In order to apply the theorem, deduce from (A.1) that

$$a_t = a_{t-1} + \gamma_t H(a_{t-1}, v_t), \tag{A.72}$$

where the map $a \mapsto H(a, v_t)$ is defined via

$$H(a, v_t) = h(a)x_t^{*2} + u_t, \tag{A.73}$$

with

$$h(a) := -c(a - \alpha). \tag{A.74}$$

The crucial, so-called *Robbins-Monro* properties of algorithms of this class are

$$\mathbb{E}\left[H(a_{t-1}, v_t) - h(a_{t-1}) \mid \mathcal{V}_{t-1}\right] = 0, \tag{A.75}$$

while for some finite constant $C > 0$

$$\mathbb{E}\left[H(a_{t-1}, v_t)^2 \mid \mathcal{V}_{t-1}\right] = \tau^2 + c^2 \frac{\kappa_x^{(4)}}{\kappa_x^{(2)2}} a_{t-1}^{*2} \leq C(1 + a_{t-1}^{*2}), \tag{A.76}$$

where $\mathcal{V}_t := \sigma(a_0; v_s, s \leq t)$. Hence, conditions 1.10.2 and 1.10.4 of the theorem are satisfied. Furthermore, h in (A.74) satisfies clearly the stability, or Lipschitz condition 1.10.5 of Benveniste et al. (1990). It thus suffices to verify that

$$\liminf_{t \rightarrow \infty} 2c \frac{\gamma_t}{\gamma_{t+1}} + \frac{\gamma_{t+1} - \gamma_t}{\gamma_{t+1}^2} > 0 \tag{A.77}$$

for all $\eta \in (0, 1)$ (i.e. condition 1.10.6 in Benveniste et al. (1990)). But, by assumption 1

$$\frac{\gamma_{t+1} - \gamma_t}{\gamma_{t+1}^2} = o(1), \quad (\text{A.78})$$

see, e.g., equation (1.5) in Chen (1993). Since $\gamma_t/\gamma_{t+1} = 1 + o(1)$, (A.77) holds true and (a) follows by the virtue of theorem 22 in Benveniste et al. (1990). \square

Proof of equation (b). The main idea follows Kushner and Yan (1993, lemma 3.1).⁸ Consider

$$\begin{aligned} a_t^{*4} - a_{t-1}^{*4} &= 4a_{t-1}^{*3}\gamma_t H(a_{t-1}, v_t) + \gamma_t^4 H(a_{t-1}, v_t)^4 + 6a_{t-1}^{*2}\gamma_t^2 H(a_{t-1}, v_t)^2 + 4a_{t-1}^*\gamma_t^3 H(a_{t-1}, v_t)^3 \\ &=: B_1 + B_2 + B_3 + B_4. \end{aligned} \quad (\text{A.79})$$

Note that $\mathbb{E}(B_1) = -4c\gamma_t \mathbb{E}(a_{t-1}^{*4})$ while the other terms obey $\mathbb{E}(B_i) = O(\gamma_t^2)\mathbb{E}(a_{t-1}^{*4}) + O(\gamma_t^3)$ for $i = 2, 3, 4$. In order to see this, begin with B_2 and note that

$$H(a_{t-1}, v_t)^4 = u_t^4 + c^4 x_t^{*8} a_{t-1}^{*4} + 6c^2 u_t^2 x_t^{*4} a_{t-1}^{*2} - 4c u_t^3 x_t^{*2} a_{t-1}^* - 4c^3 u_t x_t^{*6} a_{t-1}^{*3} \quad (\text{A.80})$$

implies

$$\mathbb{E}[H(a_{t-1}, v_t)^4 | \mathcal{V}_{t-1}] = \kappa_x^{(2)-4} \left(\kappa_x^{(4)} \kappa_\varepsilon^{(4)} + c^4 \kappa_x^{(8)} a_{t-1}^{*4} + 6c^2 \kappa_x^{(6)} \kappa_\varepsilon^{(2)} a_{t-1}^{*2} - 4c \kappa_x^{(5)} \kappa_\varepsilon^{(3)} a_{t-1}^* \right). \quad (\text{A.81})$$

Since

$$\mathbb{E}(a_t^*) = \mathbb{E}(a_0^*)\Phi(t, 1) = O(e^{-t^b}), \quad (\text{A.82})$$

using equation (A.4) and (A.17), it follows that $\mathbb{E}(B_2) = O(\gamma_t^4)\mathbb{E}(a_{t-1}^{*4}) + O(\gamma_t^4)$ by part (a) of this lemma. Furthermore, one deduces from equation (A.76) and part (a) of this lemma that

$$\mathbb{E}(B_3) = 6\gamma_t^2 (c^2 \kappa_x^{(4)}/\kappa_x^{(2)2} \mathbb{E}(a_{t-1}^{*4}) + \tau^2 \mathbb{E}(a_{t-1}^{*2})) = O(\gamma_t^2)\mathbb{E}(a_{t-1}^{*4}) + O(\gamma_t^4), \quad (\text{A.83})$$

while

$$\mathbb{E}(B_4) = O(\gamma_t^3)\mathbb{E}(a_{t-1}^{*4}) + O(\gamma_t^3). \quad (\text{A.84})$$

Hence, for t sufficiently large

$$\mathbb{E}(a_t^{*4}) - \mathbb{E}(a_{t-1}^{*4}) \leq -c\gamma_t \mathbb{E}(a_{t-1}^{*4}) + O(\gamma_t^3). \quad (\text{A.85})$$

⁸Note, that Kushner and Yan (1993) assume that their counterpart of $v_t = (x_t, \varepsilon_t)'$ is bounded a.s.; while the authors do allow for a time-series dependence in v_t . The proof here does not require v_t to be bounded with probability one because the algorithm a_t is of the Robbins-Monro class.

It thus follows for some finite constant C

$$\begin{aligned} \mathbb{E}(a_t^{*4}) &\leq (1 - c\gamma_t)\mathbb{E}(a_{t-1}^{*4}) + C\gamma_t^3 = \Phi(t, 1)\mathbb{E}(a_0^{*4}) + C \sum_{k=1}^t \gamma(k, t)\gamma_k^2 \\ &= O(e^{-tb}) + O(\gamma_t^2) = O(\gamma_t^2), \end{aligned} \quad (\text{A.86})$$

using equations (A.17) and (A.50). \square

Proof of equation (c). Set $\check{a}_{0t} := a_0^*\Phi(t, 1)$ and recall from equation (A.4) that $a_t^* = \check{a}_{0t} + \check{u}_t + \check{e}_t$. Now, because $\mathbb{E}(u_t e_s) = 0$ for all t and s (cf. lemma A.1) and, by assumption 2, v_t is independent of a_0 , it follows that

$$\mathbb{E}(a_t^{*2}) = \mathbb{E}(\check{a}_{0t}^2) + \mathbb{E}(\check{u}_t^2) + \mathbb{E}(\check{e}_t^2). \quad (\text{A.87})$$

Let us consider the summands in (A.87) one by one: First, equation (A.17) and (A.28) yield

$$\sum_{t=1}^T \mathbb{E}(\check{a}_{0t}^2) = \kappa_a^{(2)} \sum_{t=1}^T \Phi(t, 1)^2 \leq K \int_1^\infty e^{-2at^b} = O(1), \quad (\text{A.88})$$

for some finite constant $K > 0$. Turning to the second summand, note that

$$\mathbb{E}(\check{u}_t \check{u}_s) = \begin{cases} \tau^2 \sum_{k=1}^t \gamma(k, t)^2 & \text{if } s = t \\ \tau^2 \sum_{k=1}^s \gamma(k, t)\gamma(k, s) & \text{if } s < t \\ \tau^2 \sum_{k=1}^t \gamma(k, t)\gamma(k, s) & \text{if } s > t. \end{cases} \quad (\text{A.89})$$

Therefore, taking account of lemma A.2, it follows that

$$\mathbb{E}(\check{u}_t^2) = \tau^2 \phi_t^{ii} = \gamma_t \frac{\tau^2}{2c} + o(\gamma_t). \quad (\text{A.90})$$

Furthermore, (A.8) implies that

$$\mathbb{E}(\check{e}_t \check{e}_s) = \begin{cases} \sum_{k=1}^t \gamma(k, t)^2 \sigma^2(e_k) & \text{if } s = t \\ \sum_{k=1}^s \gamma(k, t)\gamma(k, s) \sigma^2(e_k) & \text{if } s < t \\ \sum_{k=1}^t \gamma(k, t)\gamma(k, s) \sigma^2(e_k) & \text{if } s > t. \end{cases} \quad (\text{A.91})$$

By (a), there exists a constant $K < \infty$ such that

$$\mathbb{E}(\check{e}_t^2) = \sum_{k=1}^t \gamma(k, t)^2 \sigma^2(e_k) = \sum_{k=1}^t \gamma(k, t)^2 O(\gamma_k) = O(\gamma_t^2), \quad (\text{A.92})$$

where the second equality uses $\sigma^2(e_k) = O(\gamma_k)$ by part (a) of this lemma, while the final order of magnitude results equation A.50. Since $\mathcal{H}_T(\eta)/T^b = 1/b + o(1)$, it follows that

$$\begin{aligned} \frac{1}{T^b} \sum_{t=1}^T \mathbb{E} (a_t^{*2}) &= \frac{1}{T^b} \sum_{t=1}^T \left(\gamma_t \frac{\sigma^2}{2c} + o(\gamma_t) \right) \\ &= \frac{\gamma \sigma^2 \mathcal{H}_T(\eta)}{2c T^b} + o\left(\frac{\mathcal{H}_T(\eta)}{T^b}\right) = \gamma \sigma^2 / (2cb) + o(1). \end{aligned} \quad (\text{A.93})$$

A.5 Lemma A.4

Assume α to be known and set $\tilde{x}_{0t} := a_{t-1}^* x_t$. Throughout the following, redefine $a_t = a_t^*$ for notational simplicity.

Lemma A.4. *Set*

$$\tilde{\rho}_{t,t+m} := \text{cov}(\tilde{x}_{0t}^2, \tilde{x}_{0,t+m}^2).$$

Then

$$\tilde{\rho}_{t,t+m} = \begin{cases} \kappa_x^{(2)} \pi_t & \text{if } m = 1 \\ \kappa_x^{(2)} \pi_t \prod_{j=1}^{m-1} \psi_{t+j} & \text{if } m > 1, \end{cases} \quad (\text{A.94})$$

where

$$\pi_t := \mathbb{E}(a_{t-1}^4) \tilde{\psi}_t - \mathbb{E}(a_{t-1}^2) \left[\kappa_x^{(2)} \mathbb{E}(a_{t-1}^2) \psi_t - \sigma^2 \gamma_t^2 (\kappa_x^{(4)} / \kappa_x^{(2)2} - 1) \right] = O(\gamma_t^2),$$

with $\tilde{\psi}_t := \kappa_x^{(2)} - c\gamma_t (2\kappa_x^{(4)} / \kappa_x^{(2)} - c\gamma_t \kappa_x^{(6)} / \kappa_x^{(2)2})$ and $\psi_t := 1 - c\gamma_t (2 - c\gamma_t \kappa_x^{(4)} / \kappa_x^{(2)2})$.

Proof. Irrespective of the nature of the gain sequence, assumption 2 implies

$$\begin{aligned} \mathbb{E}(\tilde{x}_{0t}^2, \tilde{x}_{0,t+m}^2) &= \mathbb{E}(x_t^2 a_{t-1}^2 x_{t+m}^2 a_{t+m-1}^2) \\ &= \mathbb{E}\left[a_{t-1}^2 \mathbb{E}(x_t^2 a_{t+m-1}^2 | \mathcal{V}_{t-1}) \right] \kappa_x^{(2)}, \end{aligned} \quad (\text{A.95})$$

so that

$$\tilde{\rho}_{t,t+m} = \kappa_x^{(2)} \left\{ \mathbb{E}\left[a_{t-1}^2 \mathbb{E}(x_t^2 a_{t+m-1}^2 | \mathcal{V}_{t-1}) \right] - \mathbb{E}(a_{t-1}^2) \kappa_x^{(2)} \mathbb{E}(a_{t+m-1}^2) \right\}. \quad (\text{A.96})$$

Furthermore, equation (A.72) together with (A.73) yields

$$\begin{aligned} x_t^2 a_{t+m}^2 &= x_t^2 \left[a_{t+m-1}^2 + \gamma_{t+m}^2 (u_{t+m}^2 + c^2 x_{t+m}^{*4} a_{t+m-1}^2 - 2c x_{t+m}^{*2} a_{t+m-1} u_{t+m}) \right. \\ &\quad \left. + 2\gamma_{t+m} a_{t+m-1} (u_{t+m} - c x_{t+m}^{*2} a_{t+m-1}) \right]. \end{aligned} \quad (\text{A.97})$$

Now, for $m > 1$

$$\begin{aligned} \mathbb{E}(x_t^2 a_{t+m}^2 | \mathcal{V}_{t-1}) &= \mathbb{E}(x_t^2 a_{t+m-1}^2 | \mathcal{V}_{t-1}) \psi_{t+m} + \gamma_{t+m}^2 \sigma^2 \\ &= \mathbb{E}(x_t^2 a_t^2 | \mathcal{V}_{t-1}) \prod_{j=1}^m \psi_{t+j} + \sigma^2 \left(\sum_{j=1}^{m-1} \gamma_{t+j}^2 \prod_{i=j+1}^m \psi_{t+i} + \gamma_{t+m}^2 \right), \end{aligned} \quad (\text{A.98})$$

where

$$\mathbb{E}(x_t^2 a_t^2 | \mathcal{V}_{t-1}) = a_{t-1}^2 \tilde{\psi}_t + \gamma_t^2 \sigma^2 \kappa_x^{(4)} / \kappa_x^{(2)2}. \quad (\text{A.99})$$

Similarly, it is readily verified that

$$\mathbb{E}(a_{t+m}^2) = \mathbb{E}(a_t^2) \prod_{j=1}^m \psi_{t+j} + \tau^2 \left(\sum_{j=1}^{m-1} \gamma_{t+j}^2 \prod_{i=j+1}^m \psi_{t+i} + \gamma_{t+m}^2 \right). \quad (\text{A.100})$$

Hence,

$$\mathbb{E}[a_{t-1}^2 \mathbb{E}(x_t^2 a_{t+m}^2 | \mathcal{V}_{t-1})] - \mathbb{E}(a_{t-1}^2) \kappa_x^{(2)} \mathbb{E}(a_{t+m}^2) = \pi_t \prod_{j=1}^m \psi_{t+j} \quad (\text{A.101})$$

where

$$\begin{aligned} \pi_t &= \mathbb{E}[a_{t-1}^2 \mathbb{E}(x_t^2 a_t^2 | \mathcal{V}_{t-1})] - \mathbb{E}(a_{t-1}^2) \kappa_x^{(2)} \mathbb{E}(a_t^2) \\ &= \mathbb{E}(a_{t-1}^4) \tilde{\psi}_t - \mathbb{E}(a_{t-1}^2) \left[\kappa_x^{(2)} \mathbb{E}(a_{t-1}^2) \psi_t - \sigma^2 \gamma_t^2 (\kappa_x^{(4)} / \kappa_x^{(2)2} - 1) \right]. \end{aligned} \quad (\text{A.102})$$

Finally, $\pi_t = O(\gamma_t^2)$ is a direct consequence of lemma A.3. \square

A.6 Lemma A.5

Lemma A.5. *Recall that $\tilde{x}_{0t} = x_t a_{t-1}^*$. Then,*

$$\frac{1}{T^b} \sum_{t=1}^T \tilde{x}_{t0}^2 \xrightarrow{p} \sigma^2 \frac{\gamma}{2cb} \quad (\text{a})$$

$$\frac{1}{T^{(1+b)/2}} \sum_{t=1}^T x_t \tilde{x}_{0t} = O_p \left(\frac{1}{T^{b/2}} \right) \quad (\text{b})$$

$$\frac{1}{T^{b/2}} \sum_{t=1}^T \tilde{x}_{0t} \varepsilon_t \xrightarrow{d} \mathcal{N} \left(0, \sigma^4 \frac{\gamma}{2cb} \right). \quad (\text{c})$$

A.6.1 Proof of equation (a).

Because lemma A.3 established already that

$$\lim_{T \rightarrow \infty} \frac{1}{T^b} \sum_{t=1}^T \mathbb{E}(\tilde{x}_{0t}^2) = \sigma^2 \gamma / (2cb),$$

the claim follows (by the weak LLN) if

$$\text{var} \left(\frac{1}{T^b} \sum_{t=1}^T \tilde{x}_{0t}^2 \right) = o(1). \quad (\text{A.103})$$

Consider

$$\text{var} \left(\sum_{t=1}^T \tilde{x}_{0t}^2 \right) = \sum_{t=1}^T \text{var}(\tilde{x}_{0t}^2) + 2 \sum_{t=1}^{T-1} \sum_{s=1}^{T-t} \tilde{\rho}_{t,t+s}, \quad (\text{A.104})$$

where $\tilde{\rho}_{t,t+m} = \text{cov}(\tilde{x}_{0t}^2, \tilde{x}_{0,t+m}^2)$ has been defined in lemma A.4. By lemma A.3, equation (a), there exists a constant $K < \infty$ such that

$$\sum_{t=1}^T \text{var}(\tilde{x}_{0t}^2) \leq K \mathcal{H}_T(2\eta) = \begin{cases} O(T^{2b-1}) & \text{if } \eta \neq 1/2 \\ O(\ln T) & \text{if } \eta = 1/2, \end{cases} \quad (\text{A.105})$$

where $\mathcal{H}_T(\cdot)$ denotes the generalized harmonic number (cf. equation (A.15)). Turning to the second summand on the right-hand side of (A.104), use lemma A.4 to write

$$\sum_{t=1}^{T-1} \sum_{s=1}^{T-t} \tilde{\rho}_{t,t+s} = \kappa_x^{(2)} \sum_{t=1}^{T-1} \pi_t + \sum_{t=1}^{T-2} \sum_{s=2}^{T-t} \tilde{\rho}_{t,t+s}. \quad (\text{A.106})$$

Since $\pi_t = O(\gamma_t^2)$, the first term of the previous display behaves like

$$\sum_{t=1}^{T-1} \pi_t \sim \mathcal{H}_T(2\eta) = \begin{cases} O(T^{2b-1}) & \text{if } \eta \neq 1/2 \\ O(\ln T) & \text{if } \eta = 1/2. \end{cases} \quad (\text{A.107})$$

Before the second term in (A.106) is investigated, recall from lemma A.4

$$\psi_t = 1 - c\gamma_t \left(2 - c\gamma_t \kappa_x^{(4)} / \kappa_x^{(2)2} \right).$$

Next, define

$$t^* := \min \left\{ t : \gamma_t \leq \frac{\kappa_x^{(2)2}}{c\kappa_x^{(4)}} \right\}.$$

Then,

(a) $\psi_t \leq 1 - c\gamma_t$ if $t \geq t^*$ and $\psi_t > 1 - c\gamma_t$ otherwise,

(b) ψ_t is non-negative,

(c) $\psi_t \leq \psi_{t+m}$ for $m > 0$ and $t \geq t^*$.

Proof of statement (a)-(c). Statement (a) follows directly from the definition of τ . Statement (b) is clearly true for $t < t^*$. In case $t \geq t^*$, note that

$$\begin{aligned} \psi_t \geq 0 &\Leftrightarrow 2 - c\gamma_t \kappa_x^{(4)} / \kappa_x^{(2)2} \leq \frac{1}{c\gamma_t} \\ &\Leftrightarrow c\gamma_t (1 - c\gamma_t \kappa_x^{(4)} / \kappa_x^{(2)2}) \leq 1 - c\gamma_t. \end{aligned} \quad (\text{A.108})$$

From $t \geq \tau$, it follows $c\gamma_t \kappa_x^{(4)} / \kappa_x^{(2)2} \leq 1$ and thus $1 - c\gamma_t \kappa_x^{(4)} / \kappa_x^{(2)2} \geq 0$. To see that (A.108) is true, note that $c\gamma_t \leq 1$ and, by the Lyapunov inequality, $\kappa_x^{(2)2} \leq \kappa_x^{(4)}$. Hence, $1 - c\gamma_t \kappa_x^{(4)} / \kappa_x^{(2)2} \leq 1 - c\gamma_t$. Turning to statement (c), observe that

$$\begin{aligned} \psi_t \leq \psi_{t+m} &\Leftrightarrow c\gamma_t (2 - c\gamma_t \kappa_x^{(4)} / \kappa_x^{(2)2}) \geq c\gamma_{t+m} (2 - c\gamma_{t+m} \kappa_x^{(4)} / \kappa_x^{(2)2}) \\ &\Leftrightarrow 2(c\gamma_t - c\gamma_{t+m}) \geq \kappa_x^{(4)} / \kappa_x^{(2)2} ((c\gamma_t)^2 - (c\gamma_{t+m})^2) \\ &\Leftrightarrow 2((c\gamma_t)^2 - (c\gamma_{t+m})^2)^{-1} (c\gamma_t - c\gamma_{t+m}) \geq \kappa_x^{(4)} / \kappa_x^{(2)2} \\ &\Leftrightarrow 2[(c\gamma_t)^{-1} (1 + \gamma_{t+m} / \gamma_t)^{-1} - \kappa_x^{(4)} / (2\kappa_x^{(2)2})] \geq 0. \end{aligned} \quad (\text{A.109})$$

But (A.109) is true because $(c\gamma_t)^{-1} \geq \kappa_x^{(4)} / \kappa_x^{(2)2}$ and $(1 + \gamma_{t+m} / \gamma_t)^{-1} \geq 1/2$. \square

Suppose $T > t^* \geq 3$ and decompose the second sum in (A.106) as

$$\sum_{t=1}^{T-2} \sum_{s=2}^{T-t} \tilde{\rho}_{t,t+s} = \kappa_x^{(2)} \sum_{t=1}^{t^*-2} \pi_t \sum_{s=2}^{T-t} \prod_{j=1}^{s-1} \psi_{t+j} + \kappa_x^{(2)} \sum_{t=t^*-1}^{T-2} \pi_t \sum_{s=2}^{T-t} \prod_{j=1}^{s-1} \psi_{t+j} =: A_T + B_T, \quad (\text{A.110})$$

say. By lemma A.3, there exists a finite constant C_0 such that

$$A_T \leq C_0 \sum_{t=1}^{t^*-2} \gamma_t^2 \sum_{s=2}^{T-t} \prod_{j=1}^{s-1} \psi_{t+j} \leq C_0 \sum_{t=1}^{t^*-2} \gamma_t^2 \sum_{s=2}^T \prod_{j=1}^{s-1} \psi_{t+j}, \quad (\text{A.111})$$

while the second inequality results from the non-negativity of ψ_t ; cf. property (b) above. Next, consider

$$\prod_{j=1}^{s-1} \psi_{t+j} = K(s, t) \prod_{j=1}^{s-1} \psi_j \quad \text{with} \quad K(s, t) := \left(\prod_{l=1}^t \psi_l \right)^{-1} \prod_{i=s}^{s+t-1} \psi_i. \quad (\text{A.112})$$

$K(s, t)$ is uniformly bounded in the second argument as $t < t^*$. Only the numerator of $K(s, t)$ depends on s . The numerator is clearly finite for $s < t^*$ while it behaves like the convergent product $\Phi(s + t - 1, s)$ for $s \geq t^*$; cf. property (a) above. Therefore,

$$C_1 := \sup_{\substack{2 \leq s \leq T \\ 1 \leq t \leq t^* - 2}} K(s, t) < \infty. \quad (\text{A.113})$$

It follows that,

$$A_T \leq C_0 C_1 \mathcal{H}_{t^*}(2\eta) \sum_{s=1}^T \prod_{j=1}^s \psi_j. \quad (\text{A.114})$$

Now,

$$\sum_{s=1}^T \prod_{j=1}^s \psi_j = \sum_{s=1}^{t^*-1} \prod_{j=1}^s \psi_j + \prod_{i=1}^{t^*-1} \psi_i \sum_{s=t^*}^T \prod_{j=t^*}^s \psi_j, \quad (\text{A.115})$$

where, by property (a),

$$\sum_{s=t^*}^T \prod_{j=t^*}^s \psi_j \leq \sum_{s=t^*}^T \Phi(s, t^*) = O(1) \quad (\text{A.116})$$

using (A.17) and (A.28). Therefore, $A_T < \infty$. Similarly, by lemma A.3, there exists a finite constant C_0 such that

$$\begin{aligned} B_T &\leq C_0 \sum_{t=t^*-1}^{T-2} \gamma_t^2 \sum_{s=2}^{T-t} \prod_{j=1}^{s-1} \psi_{t+j} \\ &\leq C_0 \sum_{t=t^*-1}^{T-2} \gamma_t^2 \sum_{s=1}^T \Phi(s + t - 1, t + 1). \end{aligned}$$

where the second inequality follows from property (a) and (b). Because $\sum_{s=1}^T \Phi(s + t - 1, t + 1)$ converges for any $t \geq t^* - 1$ as $T \rightarrow \infty$, it follows that there exists a finite constant C_1 so that

$$B_T \leq C_1 \mathcal{H}_T(2\eta) = \begin{cases} O(T^{2b-1}) & \text{if } \eta \neq 1/2 \\ O(\ln T) & \text{if } \eta = 1/2. \end{cases} \quad (\text{A.117})$$

□

A.6.2 Proof of equation (b).

From equation (A.4) one deduces that

$$\begin{aligned} \frac{1}{T^{(1+b)/2}} \sum_{t=1}^T x_t^2 a_{t-1}^* &= \frac{a_0^*}{T^{(1+b)/2}} \sum_{t=1}^T x_t^2 \Phi(t-1, 1) + \frac{1}{T^{(1+b)/2}} \sum_{t=1}^T x_t^2 \check{u}_{t-1} + \frac{1}{T^{(1+b)/2}} \sum_{t=1}^T x_t^2 \check{e}_{t-1} \\ &=: A_T + B_T + C_T, \end{aligned} \quad (\text{A.118})$$

say. It will be shown that the three terms on the right-hand side of (A.118) are asymptotically negligible. Start by considering A_T . Equations (A.17) and (A.29) yield, by the integral comparison test, that $E|A_T| = O(T^{-(1+b)/2})$. Furthermore, by Cauchy-Schwarz's inequality,

$$E(A_T^2) = \frac{\kappa_a^{(2)}}{T^{1+b}} \sum_{t,s=1}^T E(x_t^2 x_s^2) \Phi(t-1, 1) \Phi(s-1, 1) \leq \frac{\kappa_a^{(2)} \kappa_x^{(4)}}{T^{1+b}} \left(\sum_{t=1}^T \Phi(t-1, 1) \right)^2 = O\left(\frac{1}{T^{1+b}}\right), \quad (\text{A.119})$$

where the approximation of the majorant side of the above display uses again (A.17) and (A.29). Hence, by the weak LLN, $A_T = O_p(T^{-b})$ (note that $(1+b)/2 > b$). Turning to the second term, observe that

$$B_T = \frac{\kappa_x^{(2)}}{T^{(1+b)/2}} \sum_{t=1}^T \check{u}_{t-1} + \frac{\kappa_x^{(2)}}{T^{(1+b)/2}} \sum_{t=1}^T (x_t^{*2} - 1) \check{u}_{t-1} =: B_T^{(1)} + B_T^{(2)}, \quad (\text{A.120})$$

say, where both expressions on the right-hand side have zero mean. Now, taking account of (A.89), it follows

$$\begin{aligned} E(B_T^{(1)2}) &= \frac{\kappa_x^{(2)2}}{T^{1+b}} \sum_{t=1}^T E(\check{u}_t^2) + \frac{2\kappa_x^{(2)2}}{T^{1+b}} \sum_{t=2}^T \sum_{s=1}^{t-1} E(\check{u}_t \check{u}_s) \\ &= \frac{(\tau \kappa_x^{(2)})^2}{T^{1+b}} \sum_{t=1}^T \phi_t^{ii} + \frac{2(\tau \kappa_x^{(2)})^2}{T^{1+b}} \sum_{t=2}^T \phi_t^{iii} = O\left(\frac{1}{T}\right) + O\left(\frac{1}{T^b}\right) = O\left(\frac{1}{T^b}\right), \end{aligned} \quad (\text{A.121})$$

where the order of approximation uses part (b) and (c) of lemma A.2. Next, by construction and assumption 2, the summands of $B_T^{(2)}$ are uncorrelated so that

$$E(B_T^{(2)2}) = \frac{\kappa_x^{(2)2}}{T^{1+b}} \sum_{t=1}^T E((x_t^* - 1)^2 \check{u}_t^2) = \frac{\tau^2(\kappa_x^{(4)} - \kappa_x^{(2)2})}{T^{1+b}} \sum_{t=1}^T \phi_t^{ii} = O\left(\frac{1}{T}\right). \quad (\text{A.122})$$

Consequently, $B_T = O_p(T^{-b/2})$. Next, decompose C_T similar to B_T as

$$C_T = \frac{\kappa_x^{(2)}}{T^{(1+b)/2}} \sum_{t=1}^T \check{e}_{t-1} + \frac{\kappa_x^{(2)}}{T^{(1+b)/2}} \sum_{t=1}^T (x_t^{*2} - 1) \check{e}_{t-1} =: C_T^{(1)} + C_T^{(2)}, \quad (\text{A.123})$$

and note that both terms on the right-hand side are mean zero by construction. Hence,

$$\begin{aligned} \mathbb{E}(C_T^{(1)2}) &= \frac{\kappa_x^{(2)2}}{T^{1+b}} \sum_{t=1}^T \sum_{k=1}^t \gamma(k, t)^2 \sigma^2(e_k) + \frac{2\kappa_x^{(2)2}}{T^{1+b}} \sum_{t=2}^T \sum_{s=1}^{t-1} \sum_{k=1}^s \gamma(k, t) \gamma(k, s) \sigma^2(e_k) \\ &= \left(\frac{1}{T^{1+b}} \sum_{t=1}^T \sum_{k=1}^t \gamma(k, t)^2 O(\gamma_k) + \frac{1}{T^{1+b}} \sum_{t=2}^T \sum_{s=2}^{t-1} \sum_{k=1}^s \gamma(k, t) \gamma(k, s) O(\gamma_k) \right) = o\left(\frac{1}{T^b}\right), \end{aligned} \quad (\text{A.124})$$

where the first equality uses (A.8), while the second equality is due to equation (a) of lemma A.3. The order of magnitude $o(1/T^b)$ results from lemma A.2 and equation (A.50). Analogous to $B_T^{(2)}$, it is seen that

$$\mathbb{E}(C_T^{(2)2}) = \frac{C_1}{T^{1+b}} \sum_{t=1}^T \sum_{k=1}^t \gamma(k, t)^2 O(\gamma_k) = O\left(\frac{1}{T^{1+\eta}}\right). \quad (\text{A.125})$$

This completes the proof of (b). \square

A.6.3 Proof of equation (c).

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 $t \rightarrow \infty$) degenerate variances of the underlying stochastic process. Here, the CLT is applied to scaled partial sums of

$$z_t := \tilde{x}_{0t} \varepsilon_t, \quad (\text{A.126})$$

which form a martingale difference sequence with respect to \mathcal{V}_t and whose variances $\sigma_t^2 := \kappa_x^{(2)} \sigma^2 \mathbb{E}(a_{t-1}^{*2})$ behave approximately like γ_t ; see equation (a) of lemma A.3. In order to apply Davidson's (1994) corollary, it is helpful to introduce the following array notation:

$$Z_{tT} := z_t / s_T \quad \text{with} \quad s_T^2 := \sum_{t=1}^T \sigma_t^2. \quad (\text{A.127})$$

Note that Z_{tT} inherits the martingale difference property from z_t and, taking account of equation (b) of lemma A.3, $s_T^2 \sim T^b$. Now, according to Davidson's (1994) corollary if

1. there exists a positive constant array c_{tT} so that $(\mathbb{E}|Z_{tT}/c_{tT}|^r)^{1/r} < \infty$ uniformly for some $r > 2$;
2. $M_T = o(1)$, where $M_T := \max_{1 \leq t \leq T} c_{tT}$;

3. $\sum_{i=1}^T M_{iT}^2 = O(1)$, where $M_{iT} := \max_{\substack{(i-1) \leq t \leq i \\ i=1, \dots, T}} c_{tT}$;

then

$$\sum_{t=1}^T Z_{tT} \xrightarrow{d} \mathcal{N}(0, 1). \quad (\text{A.128})$$

If one lets $c_{tT} := \sigma_t/s_T$, then, following Davidson's (1994) argumentation surrounding his corollary, it can be seen that condition (2) and (3) 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

$$\sup_{t,T} [\mathbb{E} (Z_{tT}/c_{tT})^4]^{1/4} < \infty. \quad (\text{A.129})$$

But,

$$\mathbb{E} \left(\frac{Z_{tT}}{c_{tT}} \right)^4 \sim \frac{\mathbb{E} (z_t^4)}{\gamma_t^2} = \kappa_x^{(4)} \kappa_\varepsilon^{(4)} \frac{\mathbb{E} (a_{t-1}^{*4})}{\gamma_t^2}, \quad (\text{A.130})$$

which is, by lemma A.3, equation (b), finite. The claim follows upon noting that

$$\sum_{t=1}^T Z_{tT} = \left(\frac{1}{T^b} \sum_{t=1}^T \sigma_t^2 \right)^{-1/2} \sum_{t=1}^T \tilde{x}_t \varepsilon_t \quad (\text{A.131})$$

and, as stated in equation (c) of lemma A.3

$$\lim_{T \rightarrow \infty} \frac{1}{T^b} \sum_{t=1}^T \sigma_t^2 = \frac{\gamma \sigma^4}{2cb}. \quad \square \quad (\text{A.132})$$

A.7 Lemma A.6

Lemma A.6. *Suppose $\mathbb{E} \left[(r_0 - \kappa_x^{(2)})^2 \right] < \infty$. Then,*

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

Proof. In view of equation (A.5) and (A.6), define $\tilde{\Phi}(t, k) := \prod_{i=k}^t (1 - \gamma_i)$ with $\tilde{\Phi}(t, t+1) := 1$ and $\tilde{\gamma}(k, t) := \tilde{\Phi}(t, k+1)\gamma_k$. Then,

$$r_t^* - 1 = (r_0^* - 1)\tilde{\Phi}(t, 1) + \sum_{k=1}^t \tilde{\gamma}(k, t) (x_k^{*2} - 1), \quad (\text{A.133})$$

where $r_t^* := r_t/\kappa_x^{(2)}$ and it has been used that $\sum_{k=1}^t \tilde{\gamma}(k, t) + \tilde{\Phi}(t, 1) = 1$. The independence of $\{r_0, x_t; t \geq 1\}$ implies

$$\begin{aligned} \mathbb{E}\left[(r_t^* - 1)^2\right] &= \mathbb{E}\left[(r_0^* - 1)^2\right] \tilde{\Phi}(t, 1)^2 + \sum_{k,s=1}^t \tilde{\gamma}(k, t)\tilde{\gamma}(s, t)\mathbb{E}\left[(x_k^{*2} - 1)(x_s^{*2} - 1)\right] \\ &= \mathbb{E}\left[(r_0^* - 1)^2\right] \tilde{\Phi}(t, 1)^2 + (\kappa_x^{(4)}/\kappa_x^{(2)2} - 1) \sum_{k=1}^t (\tilde{\gamma}(k, t))^2 = O(e^{-t^b}) + O(\gamma_t), \end{aligned} \quad (\text{A.134})$$

where the final equality is due to equation (A.17) and (A.50) with $\beta = 0$ and $\alpha = 2$. Convergence in mean square and thus convergence in probability follows. Because r_t is a partial sum of the independent sequence r_0, x_1, \dots, x_t , convergence is almost surely, see e.g. Davidson (1994, theorem 20.9). \square

B Proof of proposition 2.1

To begin with, let us restate here for convenience the definition of the regressor 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 (10). The scaled deviation of the joint OLS estimator (9) in deviation from λ can then be written as

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

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

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

and

$$\det M_T := \sum_{t=1}^T x_t^2 \sum_{t=1}^T (x_t a_{t-1})^2 - \left(\sum_{t=1}^T x_t^2 a_{t-1} \right)^2. \quad (\text{B.3})$$

The proof is based on the following three steps

$$\text{plim}_{T \rightarrow \infty} \frac{\det M_T}{T^{1+b}} = \sigma^2 \kappa_x^{(2)} \gamma / (2cb) \quad (a)$$

$$\text{plim}_{T \rightarrow \infty} \frac{Q_T}{T} = \kappa_x^{(2)} \begin{bmatrix} 1 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix} \quad (b)$$

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

Step (a). Recall the definitions of $\tilde{x}_{0t} = x_t a_{t-1}^*$ and set

$$m_T := \frac{1}{T} \sum_{t=1}^T x_t^2,$$

where the latter expression approaches $\kappa_x^{(2)}$ *a.s.* by the strong LLN for *i.i.d.* random variables.

Now, taking

$$\sum_{t=1}^T x_t^2 a_{t-1} = \alpha \sum_{t=1}^T x_t^2 + \sum_{t=1}^T x_t \tilde{x}_{0t} \quad (B.4)$$

$$\sum_{t=1}^T (x_t a_{t-1})^2 = \alpha^2 \sum_{t=1}^T x_t^2 + 2\alpha \sum_{t=1}^T x_t \tilde{x}_{0t} + \sum_{t=1}^T \tilde{x}_{0t}^2 \quad (B.5)$$

into account, one gets, after some rearrangement

$$\frac{\det M_T}{T^{1+b}} = m_T \left[\frac{1}{T^b} \sum_{t=1}^T \tilde{x}_{0t}^2 - \frac{1}{m_T} \left(\frac{1}{T^{(1+b)/2}} \sum_{t=1}^T x_t \tilde{x}_{0t} \right)^2 \right]. \quad (B.6)$$

Now, by step (c) of lemma A.5,

$$\frac{1}{T^b} \sum_{t=1}^T \tilde{x}_{0t}^2 = \sigma^2 \gamma / (2cb) + o_p(1).$$

Consequently, taking equation (b) of lemma A.5, equation (B.6) and $m_T \rightarrow \kappa_x^{(2)}$ *a.s.* into account, yields

$$\frac{\det M_T}{T^{1+b}} = \frac{\kappa_x^{(2)}}{T^b} \sum_{t=1}^T \tilde{x}_{0t}^2 + O_p \left(\frac{1}{T^b} \right) = \frac{\kappa_x^{(2)} \sigma^2 \gamma}{2cb} + o_p(1). \quad (B.7)$$

This completes the proof of step (a). □

Step (b). Again, using equations (B.4) and (B.5) yields

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

In view of $m_T \rightarrow \kappa_x^{(2)}$ a.s., the claim follows because the elements of the second matrix are $o_p(1)$ by the same arguments 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. \square

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

$$\frac{Q_T/T}{T^{b/2}} \sum_{t=1}^T w_t \varepsilon_t \quad (\text{B.8})$$

are respectively given by

$$\frac{1}{T} \sum_{t=1}^T x_t^2 \frac{1}{T^{b/2}} \sum_{t=1}^T a_{t-1} x_t \varepsilon_t - \frac{1}{T} \sum_{t=1}^T x_t^2 a_{t-1} \frac{1}{T^{b/2}} \sum_{t=1}^T x_t \varepsilon_t \quad (\text{B.9})$$

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

Making again use of equations (B.4) and (B.5), it is seen that equations (B.9) and (B.10) can respectively be rewritten as

$$\frac{m_T}{T^{b/2}} \sum_{t=1}^T \tilde{x}_{0t} \varepsilon_t - u_T \left(\frac{\kappa_x^{(2)}}{T^{(1+b)/2}} \sum_{t=1}^T x_t \tilde{x}_{0t} \right) \quad (\text{B.11})$$

and

$$-\frac{\alpha m_T}{T^{b/2}} \sum_{t=1}^T \tilde{x}_{0t} \varepsilon_t + u_T \left(\frac{\alpha \kappa_x^{(2)}}{T^{(1+b)/2}} \sum_{t=1}^T x_t \tilde{x}_{0t} + \frac{\kappa_x^{(2)}}{T^{(1+b)/2}} \sum_{t=1}^T \tilde{x}_{0t}^2 \right) - \frac{1}{T} \sum_{t=1}^T x_t \tilde{x}_{0t} \frac{1}{T^{b/2}} \sum_{t=1}^T \tilde{x}_{0t} \varepsilon_t, \quad (\text{B.12})$$

where

$$u_T := \frac{1}{\sqrt{T}} \sum_{t=1}^T u_t. \quad (\text{B.13})$$

Therefore,

$$\begin{aligned}
\frac{Q_T/T}{T^{b/2}} \sum_{t=1}^T w_t \varepsilon_t &= \begin{bmatrix} 1 \\ -\alpha \end{bmatrix} \frac{m_T}{T^{b/2}} \sum_{t=1}^T \tilde{x}_{0t} \varepsilon_t \\
&\quad - \begin{bmatrix} 1 \\ -\alpha \end{bmatrix} \left(\frac{\kappa_x^{(2)}}{T^{(1+b)/2}} \sum_{t=1}^T x_t \tilde{x}_{0t} \right) u_T \\
&\quad + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \left\{ \left(\frac{\kappa_x^{(2)}}{T^{(1+b)/2}} \sum_{t=1}^T \tilde{x}_{0t}^2 \right) u_T - \left(\frac{1}{T} \sum_{t=1}^T x_t \tilde{x}_{0t} \right) \frac{1}{T^{b/2}} \sum_{t=1}^T \tilde{x}_{0t} \varepsilon_t \right\} \\
&= \begin{bmatrix} 1 \\ -\alpha \end{bmatrix} \frac{m_T}{T^{b/2}} \sum_{t=1}^T \tilde{x}_{0t} \varepsilon_t + O_p \left(\frac{1}{T^{b/2}} \right) + O_p \left(\frac{1}{T^{\eta/2}} \right) + O_p \left(\frac{1}{T^{1/2}} \right), \quad (\text{B.14})
\end{aligned}$$

where again equation (b) of lemma A.5 and step (1) of appendix D have been used together with $u_T = O_p(1)$ (by the CLT for *i.i.d.* random variables) in order to obtain the size of the remainder terms. Hence, it follows that (B.8) is asymptotically equal to

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

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

B.0.1 Remarks

Recall

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

and observe that

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

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 (1) and (2) 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 by recognizing that

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

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

C Proof of corollary 2.1

Begin with the proof of (15) and note that the definition in equation (14) of $\hat{\lambda}_\alpha$ yields

$$\hat{\lambda}_\alpha - \alpha = \frac{\hat{\lambda}_\delta - \delta + \alpha(\hat{\lambda}_\beta - \beta)}{1 - \hat{\lambda}_\beta}. \quad (\text{C.1})$$

Now, taking (B.2) and (B.3) into account, it is seen that

$$\hat{\lambda}_\beta - \beta = \frac{1}{\det M_T} \left(\sum_{t=1}^T x_t^2 \sum_{t=1}^T a_{t-1} x_t \varepsilon_t - \sum_{t=1}^T x_t^2 a_{t-1} \sum_{t=1}^T x_t \varepsilon_t \right) \quad (\text{C.2})$$

$$\hat{\lambda}_\delta - \delta = \frac{1}{\det M_T} \left(\sum_{t=1}^T (x_t a_{t-1})^2 \sum_{t=1}^T x_t \varepsilon_t - \sum_{t=1}^T x_t^2 a_{t-1} \sum_{t=1}^T a_{t-1} x_t \varepsilon_t \right). \quad (\text{C.3})$$

Consequently, one gets from equations (B.4), (B.5) and a little rearrangement

$$\begin{aligned} T^{1/2} (\hat{\lambda}_\alpha - \alpha) &= \frac{(\det M_T / T^{1+b})^{-1}}{1 - \hat{\lambda}_\beta} \frac{1}{T^{1/2}} \sum_{t=1}^T a_{t-1} x_t \varepsilon_t \left(\frac{\alpha}{T^b} \sum_{t=1}^T x_t^2 - \frac{1}{T^b} \sum_{t=1}^T x_t^2 a_{t-1} \right) \\ &\quad + \frac{(\det M_T / T^{1+b})^{-1}}{1 - \hat{\lambda}_\beta} \frac{1}{T^{1/2}} \sum_{t=1}^T x_t \varepsilon_t \left(\frac{1}{T^b} \sum_{t=1}^T (x_t a_{t-1})^2 - \frac{\alpha}{T^b} \sum_{t=1}^T x_t^2 a_{t-1} \right) \\ &= \frac{u_\mathcal{T}}{1 - \hat{\lambda}_\beta} \left(\frac{\kappa_x^{(2)} (\det M_T / T^{1+b})^{-1}}{T^b} \sum_{t=1}^T \tilde{x}_{0t}^2 \right) \\ &\quad - \frac{(\det M_T / T^{1+b})^{-1}}{1 - \hat{\lambda}_\beta} \left(\frac{1}{T^{(1+b)/2}} \sum_{t=1}^T x_t \tilde{x}_{0t} \right) \left(\frac{1}{T^{b/2}} \sum_{t=1}^T \tilde{x}_{0t} \varepsilon_t \right). \end{aligned} \quad (\text{C.4})$$

By the CLT for *i.i.d.* random variables, $u_\mathcal{T}$ is asymptotically normal with mean zero and variance τ^2 . Taking lemma A.5 together with the consistency of $\hat{\lambda}_\beta$ into account, one deduces from Slutsky's theorem and the aforementioned CLT that the first expression after the second equality is asymptotically normal with mean zero and variance $(\tau/c)^2$, while the second summand after the second equality is $o_p(1)$.

The limiting distribution of

$$T^{1/2}V^\perp(\widehat{\lambda} - \lambda) = T^{1/2}(V^\perp\widehat{\lambda} - \alpha)$$

is now easily established once one notes that

$$V^\perp\widehat{\lambda} - \alpha = \alpha\widehat{\lambda}_\beta + \widehat{\lambda}_\delta - \alpha \tag{C.5}$$

$$= (\widehat{\lambda}_\alpha - \alpha)(1 - \widehat{\lambda}_\beta). \tag{C.6}$$

Furthermore, it is clear from the construction of the matrices F_T and A defined in the main text, the first two elements of $F_TA(\widehat{\lambda} - \lambda)$ are given by $T^{b/2}(\widehat{\lambda} - \lambda)$, while the final entry is $T^{1/2}V^\perp(\widehat{\lambda} - \lambda)$. \square

D Proof of proposition 2.2

Assume α to be known and consider

$$T^{b/2}(\widehat{\beta}_0 - \beta) = \left(\frac{1}{T^b} \sum_{t=1}^T \tilde{x}_{0t}^2 \right)^{-1} \frac{1}{T^{b/2}} \sum_{t=1}^T \tilde{x}_{0t} \varepsilon_t, \tag{D.1}$$

where $\tilde{x}_{0t} = x_t a_{t-1}^*$. Equations (a) and (c) of lemma A.5 yield

$$\begin{aligned} \text{plim}_{T \rightarrow \infty} \frac{1}{T^b} \sum_{t=1}^T \tilde{x}_{0t}^2 &= \lim_{T \rightarrow \infty} \frac{1}{T^b} \sum_{t=1}^T \mathbb{E}(\tilde{x}_{0t}^2) = \sigma^2 \gamma / (2cb) \\ \frac{1}{T^{b/2}} \sum_{t=1}^T \tilde{x}_{0t} \varepsilon_t &\xrightarrow{d} \mathcal{N}(0, \sigma^4 \gamma / (2cb)). \end{aligned}$$

By Slutsky's theorem, this completes the proof. \square

E Proof of proposition 2.3

Proof. Define $\omega := \beta/(\beta - 1)$ and note that the actual law of motion can be rewritten as

$$\begin{aligned} y_t &= \beta a_{t-1} x_t + \delta x_t + \varepsilon_t \\ &= \alpha x_t + \omega x_t h(a_{t-1}) + \varepsilon_t, \end{aligned} \tag{E.1}$$

where $\delta = \alpha(1 - \beta)$ and the definition of $h(a) = -c(a - \alpha)$ (cf. equation (A.74)) has been used. In order to proceed further, it is helpful to introduce the following partial sum

$$H_{\mathcal{T}} := \frac{1}{\sqrt{T}} \sum_{t=1}^T H(a_{t-1}, v_t). \quad (\text{E.2})$$

Recall for convenience the definition of the map $a \mapsto H(a, v_t)$

$$H(a, v_t) = h(a)x_t^2 + u_t \text{ with } u_t = x_t^* \varepsilon_t / \sqrt{\kappa_x^{(2)}}; \quad (\text{E.3})$$

cf. equation (A.73). Taking account of (30) and the representation (E.1), it follows that

$$\begin{aligned} T^{1/2}(\widehat{\alpha} - \alpha) &= (\kappa_x^{(2)}/m_T) T^{-1/2} \sum_{t=1}^T \frac{y_t x_t - \alpha x_t^2}{\kappa_x^{(2)}} \\ &= (\kappa_x^{(2)}/m_T) T^{-1/2} \sum_{t=1}^T (\omega x_t^2 h(a_{t-1}) + u_t) \\ &= (\kappa_x^{(2)}/m_T) \left[\omega H_{\mathcal{T}} + c^{-1} u_{\mathcal{T}} \right], \end{aligned} \quad (\text{E.4})$$

where the last equality makes use of the definition of $H(a, v_t)$. The scaled average $u_{\mathcal{T}}$ converges in distribution to a mean-zero Gaussian random variable with variance $\tau^2 = \sigma^2/\kappa_x^{(2)}$ while $(\kappa_x^{(2)}/m_T) \rightarrow 1$ *a.s.* by the CLT and the the strong LLN for *i.i.d.* random variables, respectively. It thus remains to be shown that $H_{\mathcal{T}}$ is asymptotically negligible. Before doing so, recall (A.73) and note that the individual summands of the scaled partial sum $H_{\mathcal{T}}$ can be rewritten as

$$H(a_{t-1}, v_t) = (u_t - ca_{t-1}^*) + e_t, \quad (\text{E.5})$$

where $e_t = c(1 - x_t^{*2})a_{t-1}^*$. Henceforth, write $H_t := H(a_{t-1}, v_t)$ for notational ease.

Negligibility of $H_{\mathcal{T}}$. Using the recursive representation (A.4) of a_t^* , it follows that

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^T (u_t - ca_{t-1}^*) &= u_{\mathcal{T}} - \frac{c}{\sqrt{T}} \left(a_0^* + \sum_{t=1}^{T-1} a_t^* \right) \\ &= u_{\mathcal{T}} - c\check{u}_{\mathcal{T}} - c \left\{ \check{e}_{\mathcal{T}} - \frac{\check{e}_T + \check{u}_T}{\sqrt{T}} + \frac{a_0^*}{\sqrt{T}} \left[1 + \sum_{t=1}^{T-1} \Phi(t, 1) \right] \right\}, \end{aligned} \quad (\text{E.6})$$

where $\check{z}_{\mathcal{T}} := T^{-1/2} \sum_{t=1}^T \check{z}_t$ for $z \in \{\check{e}, \check{u}\}$; and the second equality results from

$$a_0^* + \sum_{t=1}^{T-1} a_t^* = a_0^* \left(1 + \sum_{t=1}^{T-1} \Phi(t, 1) \right) + \sum_{t=1}^T \check{u}_t + \sum_{t=1}^T \check{e}_t - (\check{e}_T + \check{u}_T). \quad (\text{E.7})$$

Taking (E.5) into account it thus follows that

$$\begin{aligned} H_{\mathcal{T}} &= (u_{\mathcal{T}} - c\check{u}_{\mathcal{T}}) + (e_{\mathcal{T}} - c\check{e}_{\mathcal{T}}) + \frac{c}{\sqrt{T}} \left\{ (\check{e}_{\mathcal{T}} + \check{u}_{\mathcal{T}}) - a_0^* \left[1 + \sum_{t=1}^{T-1} \Phi(t, 1) \right] \right\} \\ &=: H_{\mathcal{T}}^{(u)} + H_{\mathcal{T}}^{(e)} + H_{\mathcal{T}}^{(x)}, \end{aligned} \quad (\text{E.8})$$

say. Begin with $H_{\mathcal{T}}^{(u)}$ and note that $\mathbb{E} \left(H_{\mathcal{T}}^{(u)} \right) = 0$, while

$$\text{var} \left(H_{\mathcal{T}}^{(u)} \right) = \mathbb{E} \left(u_{\mathcal{T}}^2 \right) + c^2 \mathbb{E} \left(\check{u}_{\mathcal{T}}^2 \right) - 2c \mathbb{E} \left(\check{u}_{\mathcal{T}} u_{\mathcal{T}} \right). \quad (\text{E.9})$$

Now, equation (A.89) in together with lemma A.2 implies that

$$\mathbb{E} \left(\check{u}_{\mathcal{T}}^2 \right) = \frac{1}{T} \left[\sum_{t=1}^T \mathbb{E} \left(\check{u}_t^2 \right) + 2 \sum_{t=2}^T \sum_{s=1}^{t-1} \mathbb{E} \left(\check{u}_t \check{u}_s \right) \right] = \tau^2 \left(\bar{\phi}^{ii} + 2\bar{\phi}^{iii} \right). \quad (\text{E.10})$$

Next, from

$$\mathbb{E} \left(\check{u}_t u_s \right) = \begin{cases} \tau^2 \gamma(s, t) & \text{if } s \leq t \\ 0 & \text{otherwise,} \end{cases} \quad (\text{E.11})$$

it can be seen that

$$\mathbb{E} \left(\check{u}_{\mathcal{T}} u_{\mathcal{T}} \right) = \frac{1}{T} \sum_{s,t=1}^T \mathbb{E} \left(\check{u}_t u_s \right) = \frac{\tau^2}{T} \sum_{t=1}^T \sum_{s=1}^t \gamma(s, t) = \tau^2 \bar{\phi}^i. \quad (\text{E.12})$$

Consequently, in conjunction with lemma A.2, it follows that

$$\text{var} \left(H_{\mathcal{T}}^{(u)} \right) = \tau^2 \left[1 + c^2 \left(\bar{\phi}^{ii} + 2\bar{\phi}^{iii} \right) - 2c\bar{\phi}^i \right] = o(1) \quad (\text{E.13})$$

and, therefore, $H_{\mathcal{T}}^{(u)} = o_p(1)$. Turning to $H_{\mathcal{T}}^{(e)}$ note that $\mathbb{E} \left(H_{\mathcal{T}}^{(e)} \right) = 0$, while

$$\text{var} \left(H_{\mathcal{T}}^{(e)} \right) = \mathbb{E} \left(e_{\mathcal{T}}^2 \right) + c^2 \mathbb{E} \left(\check{e}_{\mathcal{T}}^2 \right) - 2c \mathbb{E} \left(\check{e}_{\mathcal{T}} e_{\mathcal{T}} \right). \quad (\text{E.14})$$

Equation (A.8) and part (a) of lemma A.3 yield

$$\mathbb{E} \left(e_{\mathcal{T}}^2 \right) = \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left(e_t^2 \right) = O \left(\frac{\mathcal{H}_T(\eta)}{T} \right) = o(1), \quad (\text{E.15})$$

where $\mathcal{H}_T(\eta)$ denotes the generalized harmonic number (cf. equation (A.15)). Again, by equation (A.91) and part (a) of lemma A.3 in conjunction with equation (A.50)

$$\begin{aligned} \mathbb{E}(\check{e}_T^2) &= \frac{1}{T} \left[\sum_{t=1}^T \mathbb{E}(\check{e}_t^2) + 2 \sum_{t=2}^T \sum_{k=1}^{t-1} \mathbb{E}(\check{e}_t \check{e}_k) \right] \\ &= \frac{1}{T} \left[\sum_{t=1}^T \sum_{k=1}^t \gamma(k, t) O(\gamma_k) + 2 \sum_{t=2}^T \sum_{k=1}^{t-1} \sum_{s=1}^k O(\gamma_k) \gamma(s, t) \gamma(s, k) \right] \\ &= O\left(\frac{\mathcal{H}_T(\eta)}{T}\right) + \frac{2}{T} \sum_{t=2}^T \sum_{k=1}^{t-1} \sum_{s=1}^k O(\gamma_k) \gamma(s, t) \gamma(s, k), \end{aligned} \quad (\text{E.16})$$

while the second term on the right-hand side is $o(1)$ by part (c) of lemma A.2 and Toeplitz's lemma. Similarly, since

$$\mathbb{E}(\check{e}_t e_s) = \begin{cases} \sigma(e_s)^2 \gamma(s, t) & \text{if } s \leq t \\ 0 & \text{otherwise,} \end{cases} \quad (\text{E.17})$$

it follows from equation (A.50) and part (a) of lemma A.3 that

$$\mathbb{E}(\check{e}_T e_T) = \frac{1}{T} \sum_{s,t=1}^T \mathbb{E}(\check{e}_t e_s) = \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^t \gamma(s, t) O(\gamma_s) = O\left(\frac{\mathcal{H}_T(\eta)}{T}\right) = o(1). \quad (\text{E.18})$$

Therefore, $H_T^{(e)} = o_p(1)$. $H_T^{(x)} = o_p(1)$ follows immediately. \square

Remark E.1. *Proof of equation (32). Note*

$$\tilde{x}_t = x_t(b_{t-1} - \hat{\alpha}) = \tilde{x}_{0t} - (\hat{\alpha} - \alpha)x_t \quad (\text{E.19})$$

$$\tilde{y}_t = y_t - \hat{\alpha}x_t = \tilde{y}_{0t} - (\hat{\alpha} - \alpha)x_t, \quad (\text{E.20})$$

with $\tilde{x}_{0t} = x_t a_{t-1}^*$ and $\tilde{y}_{0t} = y_t - \alpha x_t$, so that

$$\begin{aligned} \frac{1}{T^b} \sum_{t=1}^T \tilde{x}_t^2 &= \frac{1}{T^b} \sum_{t=1}^T \tilde{x}_{0t}^2 + (\hat{\alpha} - \alpha)^2 \frac{1}{T^b} \sum_{t=1}^T x_t^2 - 2(\hat{\alpha} - \alpha) \frac{1}{T^b} \sum_{t=1}^T x_t^2 a_{t-1}^* \\ &= \frac{1}{T^b} \sum_{t=1}^T \tilde{x}_{0t}^2 + O_p\left(\frac{1}{T^b}\right) \end{aligned} \quad (\text{E.21})$$

$$\begin{aligned} \frac{1}{T^{b/2}} \sum_{t=1}^T \tilde{y}_t \tilde{x}_t &= \frac{1}{T^{b/2}} \sum_{t=1}^T \tilde{y}_{0t} \tilde{x}_{0t} - \frac{(\hat{\alpha} - \alpha)(\beta + 1)}{T^{b/2}} \sum_{t=1}^T x_t^2 a_{t-1}^* \\ &\quad - \frac{\hat{\alpha} - \alpha}{T^{b/2}} \sum_{t=1}^T x_t \varepsilon_t + \frac{(\hat{\alpha} - \alpha)^2}{T^{b/2}} \sum_{t=1}^T x_t^2 \\ &= \frac{1}{T^{b/2}} \sum_{t=1}^T \tilde{y}_{0t} \tilde{x}_{0t} + O_p\left(\frac{1}{T^{b/2}}\right). \end{aligned} \quad (\text{E.22})$$

using (b) and proposition 2.3. Therefore, by Slutsky's theorem

$$T^{b/2}(\widehat{\beta} - \beta) = T^{b/2}(\widehat{\beta}_0 - \beta) + O_p\left(\frac{1}{T^{b/2}}\right). \quad (\text{E.23})$$

Proof of equation (34). Using proposition 2.2 and 2.3, it follows

$$\begin{aligned} T^{b/2}(\widehat{\delta} - \delta) &= T^{b/2}(\alpha(1 - \widehat{\beta}) - \delta) + T^{b/2}(\widehat{\alpha} - \alpha)(1 - \widehat{\beta}) \\ &= -\alpha T^{b/2}(\widehat{\beta} - \beta) + O_p\left(T^{-\frac{\eta}{2}}\right). \end{aligned} \quad (\text{E.24})$$

F Extension A: multiple regressors

This appendix extends the results collected in corollary 2.1 to the multivariate case where x_t represents a k -dimensional column vector of regressors. As mentioned in section 2.4 of the main text, the agent's multivariate recursion (24) for the k -dimensional RE equilibrium vector α is given by

$$a_t = a_{t-1} + \gamma_t Q^{-1} x_t (y_t - a'_{t-1} x_t), \quad (\text{F.1})$$

where $Q := \text{E}(x_t x'_t)$, so that $y_{t|t-1}^e = a'_{t-1} x_t$ represents the agent's forecast of y_t . Furthermore, the following regularity conditions are imposed:

Assumption 2.M. *The random vector $v_t := (x'_t, \varepsilon_t)'$ is independent, identically distributed and x'_t is independent of ε_t .*

Assumption 3.M. *The following matrices exist and are positive definite*

$$\begin{aligned} Q &:= \text{E}(x_t x'_t) \\ Q^{(4)} &:= \text{E}[(x'_t Q^{-1} x_t) x_t x'_t] \\ \tilde{Q}^{(4)} &:= \text{E}(x_t x'_t \otimes x_t x'_t) \\ Q^{(6)} &:= \text{E}[(x'_t Q^{-1} x_t)^2 x_t x'_t] \\ \tilde{Q}^{(6)} &:= \text{E}[(x'_t Q^{-1} x_t)(x_t x'_t \otimes x_t x'_t)] \\ \tilde{Q}^{(8)} &:= \text{E}[(x'_t Q^{-1} x_t)^2 (x_t x'_t \otimes x_t x'_t)], \end{aligned}$$

with \otimes denoting the Kronecker product. Furthermore, $\kappa_\varepsilon^{(4)} < \infty$ and $\text{E}[(a_0^{*'} Q a_0^*)^2] < \infty$, where $a_t^* := a_t - \alpha$ for $t = 0, 1, 2, \dots$

It will be shown that the $(k+1)$ -dimensional OLS estimator $\widehat{\lambda}$ based on

$$y_t = \beta a'_{t-1} x_t + \delta' x_t + \varepsilon_t \quad \text{for } t = 1, 2, \dots, T, \quad (\text{F.2})$$

where $|\beta| < 1$ and $\delta \in \mathbb{R}^k$, has the following properties:

Proposition F.1. *Suppose assumptions 1, 2.M and 3.M hold. Then,*

$$T^{b/2}(\widehat{\lambda} - \lambda) \xrightarrow{d} \mathcal{N}(0_{k+1}, V) \quad \text{with } V = \frac{2cb}{k\gamma} \begin{bmatrix} 1 & -\alpha' \\ -\alpha & \alpha\alpha' \end{bmatrix} \quad (a)$$

$$T^{1/2}V^\perp(\widehat{\lambda} - \lambda) \xrightarrow{d} \mathcal{N}(0_k, \sigma^2Q^{-1}). \quad (b)$$

Proof. To begin with, multivariate equivalents of lemma A.3 and A.5 are stated. The proof of lemma F.1 and F.2 is delegated to the end.

Lemma F.1. *Suppose assumptions 1, 2.M and 3.M hold. Then,*

$$\mathbb{E}(a_t^*Qa_t^*) = O(\gamma_t) \quad (a)$$

$$\mathbb{E}[(a_t^*Qa_t^*)^2] = O(\gamma_t^2), \quad (b)$$

and

$$\lim_{T \rightarrow \infty} \frac{1}{T^b} \sum_{t=1}^T \mathbb{E}(a_t^*a_t^*) = \sigma^2Q^{-1} \frac{\gamma}{2cb}. \quad (c)$$

Lemma F.2. *Let $\tilde{x}_{0t} := a_{t-1}^*x_t$. Suppose assumptions 1, 2.M and 3.M hold. Then,*

$$\frac{1}{T^b} \sum_{t=1}^T \tilde{x}_{0t}^2 \xrightarrow{p} k\sigma^2 \frac{\gamma}{2cb} \quad (a)$$

$$\frac{1}{T^{(1+b)/2}} \sum_{t=1}^T x_t \tilde{x}_{0t} = O_p\left(\frac{1}{T^{b/2}}\right) \quad (b)$$

$$\frac{1}{T^{b/2}} \sum_{t=1}^T \tilde{x}_{0t} \varepsilon_t \xrightarrow{d} \mathcal{N}\left(0, k\sigma^4 \frac{\gamma}{2cb}\right). \quad (c)$$

Recall that $\widehat{\lambda}$ denotes the $(k+1)$ -dimensional OLS estimator of $\lambda = (\beta, \delta)'$ based on

$$y_t = \lambda'w_t + \varepsilon_t \quad \text{for } t = 1, 2, \dots, T, \quad (F.3)$$

where $w_t := (y_{t|t-1}^e, x_t')'$ and $y_{t|t-1}^e = a_{t-1}'x_t$. Specifically, upon defining the second moment matrix

$$M_T := \sum_{t=1}^T w_t w_t' = W'W, \quad (F.4)$$

with $W := (w_1', \dots, w_T')'$ (a $T \times (k+1)$ matrix) and $X := (x_1', \dots, x_T')'$ (a $T \times k$ matrix), we can write

$$\widehat{\lambda} = M_T^{-1}W'y. \quad (F.5)$$

In a first step, it will be shown that

$$T^b M_T^{-1} = \frac{2cb}{k\gamma\sigma^2} \begin{bmatrix} 1 & -\alpha' \\ -\alpha & \alpha\alpha' \end{bmatrix} + o_p(1). \quad (\text{F.6})$$

Note that

$$M_T^{-1} = \frac{1}{\zeta_T} \begin{bmatrix} 1 & -b'_T \\ -b_T & B_T \end{bmatrix}, \quad (\text{F.7})$$

where

$$b_T := (X'X)^{-1}X'y^e \quad (\text{F.8})$$

$$B_T := (X'X)^{-1} \left[I_k \zeta_T + X'y^e y^{e'} X (X'X)^{-1} \right] \quad (\text{F.9})$$

$$\zeta_T := y^{e'} y^e - y^{e'} X (X'X)^{-1} X' y^e, \quad (\text{F.10})$$

with $y^e := (y_{1|0}^e, \dots, y_{T|T-1}^e)'$; see, e.g., Abadir and Magnus (2006, exercise 5.15). It follows readily from the weak LLN and lemma F.2 that

$$\frac{X'X}{T} = Q + o_p(1) \quad (\text{F.11})$$

$$\frac{X'y^e}{T} = \frac{X'X}{T} \alpha + \frac{U_T}{T} = Q\alpha + o_p(1) \quad \text{where } U_T := X'\tilde{X} \quad (\text{F.12})$$

$$\frac{\zeta_T}{T^b} = \frac{\tilde{X}'\tilde{X}}{T^b} - \frac{U_T'}{T^{(1+b)/2}} \left(\frac{X'X}{T} \right)^{-1} \frac{U_T}{T^{(1+b)/2}} = k\sigma^2 \frac{\gamma}{2cb} + o_p(1) \quad (\text{F.13})$$

$$B_T = \left(\frac{X'X}{T} \right)^{-1} \left[I_k (\zeta_T/T) + \left(\frac{X'y^e}{T} \right) \left(\frac{X'y^e}{T} \right)' \left(\frac{X'X}{T} \right)^{-1} \right] = \alpha\alpha' + o_p(1), \quad (\text{F.14})$$

where $\tilde{X} := (\tilde{x}_{01}, \dots, \tilde{x}_{0T})'$ and using repeatedly that $X'y^e = U_T + X'X\alpha$. The result in equation (F.6) follows directly. Now, taking (F.7) into account, the deviation of $\hat{\lambda}$ from the true vector λ can be written as

$$\hat{\lambda} - \lambda = M_T^{-1} W' \varepsilon = \frac{1}{\zeta_T} [(1, -b'_T)' y^{e'} \varepsilon + (-b_T, B_T)' X' \varepsilon], \quad (\text{F.15})$$

where $(-b_T, B_T)$ is a $k \times (k+1)$ matrix and $\varepsilon := (\varepsilon_1, \dots, \varepsilon_T)'$. With a little rearrangement

(similar to that of (B.8)), it follows from the above and lemma F.2 that

$$\begin{aligned}
T^{b/2}(\widehat{\lambda} - \lambda) &= \frac{1}{\widehat{\zeta}_T/T^b} \begin{bmatrix} 1 \\ -\alpha \end{bmatrix} \frac{\widetilde{X}'\varepsilon}{T^{b/2}} \\
&\quad - \frac{1}{\widehat{\zeta}_T/T^b} \begin{bmatrix} 1 \\ -\alpha \end{bmatrix} \frac{U'_T}{T^{(1+b)/2}} \left(\frac{X'X}{T} \right)^{-1} \frac{X'\varepsilon}{T^{1/2}} \\
&\quad + \frac{1}{\widehat{\zeta}_T/T^b} \begin{bmatrix} 0'_k \\ I_k \end{bmatrix} \left[\left(\frac{X'X}{T} \right)^{-1} \frac{\zeta_T}{T^{(1+b)/2}} + \left(\frac{X'X}{T} \right)^{-1} \frac{U_T}{T^{(1+b)/2}} \frac{U'_T}{T} \left(\frac{X'X}{T} \right)^{-1} \right] \frac{X'\varepsilon}{T^{1/2}} \\
&\quad - \frac{1}{\widehat{\zeta}_T/T^b} \begin{bmatrix} 0'_k \\ I_k \end{bmatrix} \left(\frac{X'X}{T} \right)^{-1} \frac{U_T}{T^{(1+b)/2}} \frac{\widetilde{X}'\varepsilon}{T^{1/2}} \\
&= \frac{1}{\widehat{\zeta}_T/T^b} \begin{bmatrix} 1 \\ -\alpha \end{bmatrix} \frac{\widetilde{X}'\varepsilon}{T^{b/2}} + o_p(1) \xrightarrow{d} \mathcal{N}(0_{k+1}, V). \tag{F.16}
\end{aligned}$$

Next, let $\widehat{\lambda}_\beta$ and $\widehat{\lambda}'_\delta$ denote the first entry and the last $k-1$ entries of $\widehat{\lambda}$, respectively. Then,

$$\begin{aligned}
\widehat{\lambda} - \lambda &= \begin{bmatrix} \widehat{\lambda}_\beta - \beta \\ \widehat{\lambda}'_\delta - \beta \end{bmatrix} \\
&= \frac{1}{\widehat{\zeta}_T} \begin{bmatrix} y^{e'}\varepsilon - b'_T X'\varepsilon \\ (-b_T, B_T)W'\varepsilon \end{bmatrix}. \tag{F.17}
\end{aligned}$$

Similar to the proof of corollary 2.1, it is seen that the above in conjunction with the multivariate CLT for *i.i.d.* random vectors applied to $X'\varepsilon/T^{1/2}$ implies

$$\begin{aligned}
T^{1/2}V^\perp(\widehat{\lambda} - \lambda) &= T^{1/2}(\widehat{\lambda}_\delta - \delta + \alpha(\widehat{\lambda}_\beta - \beta)) \\
&= \left(\frac{X'X}{T} \right)^{-1} \frac{X'\varepsilon}{T^{1/2}} \\
&\quad - \frac{1}{\widehat{\zeta}_T/T^b} \left(\frac{X'X}{T} \right)^{-1} \frac{U_T}{T^{1/2}} \frac{\widetilde{X}'\varepsilon}{T^b} \\
&\quad + \frac{1}{\widehat{\zeta}_T/T^b} \left(\frac{X'X}{T} \right)^{-1} \frac{U_T}{T^b} \frac{U'_T}{T} \left(\frac{X'X}{T} \right)^{-1} \frac{X'\varepsilon}{T^{1/2}} \\
&= Q^{-1} \frac{X'\varepsilon}{T^{1/2}} + o_p(1) \xrightarrow{d} \mathcal{N}(0_k, \sigma^2 Q^{-1}). \quad \square \tag{F.18}
\end{aligned}$$

Proof of lemma F.1.

For notational ease, set $a_t := a_t^*$ throughout the proof.

Proof of equation (a). Let

$$H_t := Q^{-1}(x_t \varepsilon_t - c x_t x'_t a_{t-1}) \quad (\text{F.19})$$

so that $a_t = a_{t-1} + \gamma_t H_t$, while for $m \geq 0$

$$\begin{aligned} a'_{t+m} Q a_{t+m} &= a'_{t+m-1} Q a_{t+m-1} \\ &+ 2\gamma_{t+m} (a'_{t+m-1} x_{t+m} \varepsilon_{t+m} - c a'_{t+m-1} x_{t+m} x'_{t+m} a_{t+m-1}) \\ &+ \gamma_{t+m}^2 (\varepsilon_{t+m}^2 x'_{t+m} Q^{-1} x_{t+m} - 2c \varepsilon_{t+m} x'_{t+m} Q^{-1} x_{t+m} x'_{t+m} a_{t+m-1} \\ &\quad + c^2 a'_{t+m-1} x_{t+m} x'_{t+m} Q^{-1} x_{t+m} x'_{t+m} a_{t+m-1}). \end{aligned} \quad (\text{F.20})$$

Hence, for $m = 0$

$$\mathbb{E}(a'_t Q a_t) = \mathbb{E}(a'_{t-1} Q a_{t-1}) (1 - 2c\gamma_t) + \gamma_t^2 k \sigma^2 + (c\gamma_t)^2 \mathbb{E}(a'_{t-1} Q^{(4)} a_{t-1}). \quad (\text{F.21})$$

Using that $Q^{(4)}$ and Q are positive definite, it follows that for any k -dimensional vector a with t sufficiently large, $c\gamma_t a' Q^{(4)} a \leq a' Q a$ so that

$$\begin{aligned} \mathbb{E}(a'_t Q a_t) &\leq \mathbb{E}(a'_{t-1} Q a_{t-1}) (1 - c\gamma_t) + \gamma_t^2 k \sigma^2 \\ &= \Phi(t, 1) \mathbb{E}(a'_0 Q a_0) + k \sigma^2 \sum_{k=1}^t \gamma(k, t) \gamma_k. \end{aligned} \quad (\text{F.22})$$

By equations (A.17) and (A.50), the two summands of the preceding display are respectively of size $O(e^{-t^b})$ and $O(\gamma_t^2)$. \square

Proof of equation (b). Consider

$$\begin{aligned} (a'_t Q a_t)^2 - (a'_{t-1} Q a_{t-1})^2 &= 4\gamma_t a'_{t-1} Q a_{t-1} a'_{t-1} Q H_t \\ &+ 4\gamma_t^2 (a'_{t-1} Q H_t)^2 \\ &+ \gamma_t^4 (H'_t Q H_t)^2 \\ &+ 2\gamma_t^2 a'_{t-1} Q a_{t-1} H'_t Q H_t \\ &+ 4\gamma_t^3 a'_{t-1} Q H_t H'_t Q H_t =: B_1 + B_2 + B_3 + B_4 + B_5, \end{aligned} \quad (\text{F.23})$$

say. First, note that

$$\mathbb{E}(B_1 | \mathcal{V}_{t-1}) = -4c\gamma_t (a'_{t-1} Q a_{t-1})^2, \quad (\text{F.24})$$

i.e. $\mathbb{E}(B_1) = -4c\gamma_t \mathbb{E}[(a'_{t-1} Q a_{t-1})^2]$. The remainder of this proof aims at establishing that

$$\sum_{i=2}^5 \mathbb{E}(B_i) = O(\gamma_t^2) \mathbb{E}[(a'_{t-1} Q a_{t-1})^2] + O(\gamma_t^3). \quad (\text{F.25})$$

Throughout, it will frequently be used that for any n -dimensional vector a and conformable, symmetric matrices A and B

$$(a' A a)(a' B a) = (a \otimes a)(A \otimes B)(a \otimes a). \quad (\text{F.26})$$

Furthermore, the proof makes repeatedly use of the fact that, by assumption 3.M, for any vector a (\tilde{a}) of dimension k (k^2) there exist finite constants $K_i > 0$ ($i = 1, \dots, 5$) so that

$$a' Q^{(4)} a = K_1 a' Q a \quad (\text{F.27})$$

$$\tilde{a}' \tilde{Q}^{(4)} \tilde{a} = K_2 \tilde{a}' (Q \otimes Q) \tilde{a} \quad (\text{F.28})$$

$$a' Q^{(6)} a = K_3 a' Q a \quad (\text{F.29})$$

$$\tilde{a}' \tilde{Q}^{(6)} \tilde{a} = K_4 \tilde{a}' (Q \otimes Q) \tilde{a} \quad (\text{F.30})$$

$$\tilde{a}' \tilde{Q}^{(8)} \tilde{a} = K_5 \tilde{a}' (Q \otimes Q) \tilde{a}. \quad (\text{F.31})$$

Now, consider

$$\text{E}(B_2 | \mathcal{V}_{t-1}) = 4\gamma_t^2 \sigma^2 a'_{t-1} Q a_{t-1} + 4\gamma_t^2 c^2 \text{E} [(a'_{t-1} x_t x'_t a_{t-1})^2 | \mathcal{V}_{t-1}]. \quad (\text{F.32})$$

Equations (F.26) and (F.28) yield

$$\begin{aligned} \text{E} [(a'_{t-1} x_t x'_t a_{t-1})^2 | \mathcal{V}_{t-1}] &= (a_{t-1} \otimes a_{t-1}) \tilde{Q}^{(4)} (a_{t-1} \otimes a_{t-1}) \\ &= K_2 (a'_{t-1} Q a_{t-1})^2. \end{aligned} \quad (\text{F.33})$$

Hence,

$$\text{E}(B_2) = O(\gamma_t^2) \text{E}(a'_{t-1} Q a_{t-1})^2 + O(\gamma_t^3),$$

using that $\text{E}(a'_t Q a_t) = O(\gamma_t)$ by part (a) of this lemma. Turning to B_3 , observe that

$$\begin{aligned} (H'_t Q H_t)^2 &= \varepsilon_t^4 (x'_t Q^{-1} x_t)^2 + 6c^2 \varepsilon_t^2 (x'_t Q^{-1} x_t x'_t a_{t-1})^2 + c^4 (a'_{t-1} x_t x'_t Q^{-1} x_t x'_t a_{t-1})^2 \\ &\quad - 4c \varepsilon_t^3 x'_t Q^{-1} x_t x'_t Q^{-1} x_t x'_t a_{t-1} \\ &\quad - 4c^3 \varepsilon_t x'_t Q^{-1} x_t x'_t a_{t-1} a'_{t-1} x_t x'_t Q^{-1} x_t x'_t a_{t-1} =: A_1 + A_2 + A_3 - A_4 - A_5. \end{aligned} \quad (\text{F.34})$$

Now,

$$\begin{aligned} \text{E}(A_1) &= \kappa_\varepsilon^{(4)} \text{E} [(x_t \otimes x_t)(Q^{-1} \otimes Q^{-1})(x_t \otimes x_t)] \\ &= \kappa_\varepsilon^{(4)} \text{tr} [(Q^{-1} \otimes Q^{-1}) \tilde{Q}^{(4)}] < \infty. \end{aligned} \quad (\text{F.35})$$

Next,

$$\begin{aligned}
\mathbb{E}(A_2) &= 6c^2\kappa_\varepsilon^{(2)}\mathbb{E}[(x'_tQ^{-1}x_t x'_t a_{t-1})^2] \\
&= 6c^2\kappa_\varepsilon^{(2)}\mathbb{E}[a'_{t-1}\mathbb{E}(x_t x'_t Q^{-1}x_t x'_t Q^{-1}x_t x'_t) a_{t-1}] \\
&= 6c^2\kappa_\varepsilon^{(2)}\mathbb{E}[a'_{t-1}Q^{(6)}a_{t-1}] \\
&= K_3 6c^2\kappa_\varepsilon^{(2)}\mathbb{E}(a'_{t-1}Qa_{t-1}) = O(\gamma_t),
\end{aligned} \tag{F.36}$$

where $K_3 > 0$ follows from equation (F.29). Turning to A_3 , consider

$$\begin{aligned}
\mathbb{E}(A_3) &= c^4\mathbb{E}[(a'_{t-1}x_t x'_t Q^{-1}x_t x'_t a_{t-1})^2] \\
&= c^4\mathbb{E}[(a_{t-1} \otimes a_{t-1})Q^{(8)}(a_{t-1} \otimes a_{t-1})] \\
&= c^4 K_5 \mathbb{E}[(a'_{t-1}Qa_{t-1})^2],
\end{aligned} \tag{F.37}$$

where $K_5 > 0$ results from equation (F.31). Next,

$$\begin{aligned}
\mathbb{E}(A_4) &= 4c\kappa_\varepsilon^{(3)}\mathbb{E}(x'_t Q^{-1}x_t x'_t Q^{-1}x_t x'_t)\mathbb{E}(a_{t-1}) \\
&= 4c\kappa_\varepsilon^{(3)}\mathbb{E}(x'_t Q^{-1}x_t x'_t Q^{-1}x_t x'_t)\mathbb{E}(a_0)\Phi_{(t-1,1)} = O(e^{-tb}),
\end{aligned} \tag{F.38}$$

using equation (F.44). Finally, $\mathbb{E}(A_5) = 0$. Hence, putting the above together yields

$$\mathbb{E}(B_3) = O(\gamma_t^4)\mathbb{E}[(a'_{t-1}Qa_{t-1})^2] + O(\gamma_t^4). \tag{F.39}$$

Next,

$$\begin{aligned}
\mathbb{E}(B_4|\mathcal{V}_{t-1}) &= 2\gamma_t k\sigma^2 a'_{t-1}Qa_{t-1} + 2c\gamma_t(a_{t-1} \otimes a_{t-1})(Q \otimes Q^{(4)})(a_{t-1} \otimes a_{t-1}) \\
&= O(\gamma_t^2)(a'_{t-1}Qa_{t-1})^2 + O(\gamma_t^3),
\end{aligned} \tag{F.40}$$

using that there exists a finite constant $K > 0$ so that $\tilde{a}'(Q \otimes Q^{(4)})\tilde{a} < K \tilde{a}'(Q \otimes Q)\tilde{a}$, for any k^2 -dimensional vector \tilde{a} . Finally, from the analysis of $\mathbb{E}(B_2)$ and $\mathbb{E}(B_3)$ in conjunction with Cauchy-Schwarz's inequality one gets

$$\mathbb{E}(a'_{t-1}QH_t H'_t QH_t) \leq \mathbb{E}[(a'_{t-1}QH_t)^2]^{1/2} \mathbb{E}[(H'_t QH_t)^2]^{1/2} = O(\gamma_t^{1/2}), \tag{F.41}$$

i.e. $\mathbb{E}(B_5) = O(\gamma_t^{7/2})$. Therefore, putting $\mathbb{E}(B_i)$ ($i = 1, \dots, 5$) together it follows

$$\begin{aligned}
\mathbb{E}[(a'_t Qa_t)^2] - \mathbb{E}[(a'_{t-1}Qa_{t-1})^2] &= -4c\gamma_t \mathbb{E}[(a'_{t-1}Qa_{t-1})^2] \\
&\quad + O(\gamma_t^2)\mathbb{E}[(a'_{t-1}Qa_{t-1})^2] + O(\gamma_t^3).
\end{aligned} \tag{F.42}$$

Hence, for t sufficiently large and some finite constant K

$$\begin{aligned} \mathbb{E} [(a'_t Q a_t)^2] &= (1 - c\gamma_t) \mathbb{E} [(a'_{t-1} Q a_{t-1})^2] + K\gamma_t^3 \\ &= \Phi(t, 1) \mathbb{E} [(a'_0 Q a_0)^2] + K \sum_{k=1}^t \gamma(k, t) \gamma_k^2 = O(\gamma_t^2), \quad \square \end{aligned} \quad (\text{F.43})$$

where the final equality is due to (A.50).

Proof of equation (c). In order to see this, observe that analogous to (A.4), one gets

$$\begin{aligned} a_t^* &= a_{t-1}^* (1 - c\gamma_t) + \gamma_t Q^{-1} x_t \varepsilon_t + \gamma_t c (I_t - Q^{-1} x_t x_t') a_{t-1}^* \\ &= a_0^* \Phi(t, 1) + \check{u}_t + \check{e}_t, \end{aligned} \quad (\text{F.44})$$

with

$$\check{z}_t := \sum_{k=1}^t \gamma_{k,t} z_k \quad \text{for } z \in \{u, e\}, \quad (\text{F.45})$$

where $u_t := Q^{-1} x_t \varepsilon_t$ and $e_t := c(I_t - Q^{-1} x_t x_t') a_{t-1}^*$. As in the proof of lemma A.1, it is readily verified that u_t and e_s are uncorrelated for all t and s . It thus follows that

$$\mathbb{E}(a_t^* a_t^{*'}) = \mathbb{E}(a_0^* a_0^{*'}) \Phi(t, 1)^2 + \mathbb{E}(\check{u}_t \check{u}_t') + \mathbb{E}(\check{e}_t \check{e}_t'). \quad (\text{F.46})$$

The following arguments mirror the proof of equation (b) of lemma A.3: the first term is of order $O(e^{-tb})$, while the second summand obeys

$$\mathbb{E}(\check{u}_t \check{u}_t') = \sum_{k, s=1}^t \gamma(k, t) \gamma(s, t) \mathbb{E}(u_k u_s') = \sigma^2 Q^{-1} \sum_{k=1}^t \gamma(k, t)^2 = \sigma^2 Q^{-1} \frac{\gamma_t}{2c} + o(\gamma_t), \quad (\text{F.47})$$

and the third term is of order $O(\gamma_t^2)$. This proves the claim. \square

Proof of lemma F.2

Proof of equation (a). To begin with, consider

$$\mathbb{E}(\tilde{x}_{0t}^2) = \mathbb{E}(\text{tr} [x_t x_t' a_{t-1}^* a_{t-1}^{*'}]) = \text{tr} [Q \mathbb{E}(a_{t-1}^* a_{t-1}^{*'})]. \quad (\text{F.48})$$

Next, note that by equation (c) of lemma F.1

$$\frac{1}{T^b} \sum_{t=1}^T \mathbb{E}(a_t^* a_t^{*'}) \rightarrow \sigma^2 Q^{-1} \frac{\gamma}{2cb}. \quad (\text{F.49})$$

The claim thus follows by the weak LLN if

$$\text{var} \left(\frac{1}{T^b} \sum_{t=1}^T \tilde{x}_{0t}^2 \right) = o(1). \quad (\text{F.50})$$

First, define

$$\tilde{\rho}_{t,t+m} := \text{cov} \left(\tilde{x}_{0t}^2, \tilde{x}_{0,t+m}^2 \right). \quad (\text{F.51})$$

Taking account of the proof of equation (a) of lemma A.5, the claim follows if it can be shown that

$$\tilde{\rho}_{t,t+m} = \begin{cases} \pi_t & \text{if } m = 1 \\ \pi_t \prod_{j=1}^{m-1} \phi_{t+j} & \text{if } m > 1, \end{cases} \quad (\text{F.52})$$

where $\phi := 1 - c\gamma_t(2 - Kc\gamma_t)$ and $K > 0$ is some finite constant, while $\pi_t = O(\gamma_t^2)$. To see this, consider first

$$\text{E}(\tilde{x}_{0t}^2 \tilde{x}_{0,t+m}^2) = \text{E} \left[a'_{t-1} \text{E}(a'_{t+m-1} Q a_{t+m-1} x_t x'_t | \mathcal{V}_{t-1}) a_{t-1} \right] \quad (\text{F.53})$$

Taking account of equation (F.20), it follows that

$$\begin{aligned} \text{E}(a'_{t+m} Q a_{t+m} x_t x'_t | \mathcal{V}_{t-1}) &= (1 - 2c\gamma_{t+m}) \text{E}(a'_{t+m-1} Q a_{t+m-1} x_t x_t | \mathcal{V}_{t-1}) \\ &\quad + \gamma_{t+m}^2 \sigma^2 k Q \\ &\quad + (c\gamma_{t+m})^2 \text{E}(a'_{t+m-1} Q^{(4)} a_{t+m-1} x_t x_t | \mathcal{V}_{t-1}) \\ &= \phi_{t+m} \text{E}(a'_{t+m-1} Q a_{t+m-1} x_t x_t | \mathcal{V}_{t-1}) + \gamma_{t+m}^2 \sigma^2 k Q. \end{aligned} \quad (\text{F.54})$$

the second equality is due to equation (F.27). Solving the preceding display recursively yields

$$\text{E}(a'_{t+m} Q a_{t+m} x_t x'_t | \mathcal{V}_{t-1}) = \text{E}(a'_t Q a_t x_t x_t | \mathcal{V}_{t-1}) \prod_{j=1}^m \phi_{t+j} + k\sigma^2 Q \left(\sum_{j=1}^{m-1} \gamma_{t+j}^2 \prod_{i=j+1}^m \phi_{t+i} + \gamma_{t+m}^2 \right),$$

so that

$$\begin{aligned} \text{E}(a'_{t-1} \text{E}(a'_{t+m} Q a_{t+m} x_t x'_t | \mathcal{V}_{t-1}) a_{t-1}) &= \text{E}(a'_{t-1} \text{E}(a'_t Q a_t x_t x'_t | \mathcal{V}_{t-1}) a_{t-1}) \prod_{j=1}^m \phi_{t+j} \\ &\quad + \text{E}(a'_{t-1} Q a_{t-1}) k\sigma^2 \left(\sum_{j=1}^{m-1} \gamma_{t+j}^2 \prod_{i=j+1}^m \phi_{t+i} + \gamma_{t+m}^2 \right). \end{aligned} \quad (\text{F.55})$$

Similarly,

$$\begin{aligned} \mathbb{E}(a'_{t+m} Q a_{t+m}) &= \mathbb{E}(a'_{t+m-1} Q a_{t+m-1}) \phi_{t+m} + k \gamma_{t+m}^2 \sigma^2 \\ &= \mathbb{E}(a'_t Q a_t) \prod_{j=1}^m \phi_{t+j} + k \sigma^2 \left(\sum_{j=1}^{m-1} \gamma_{t+j}^2 \prod_{i=j+1}^m \phi_{t+i} + \gamma_{t+m}^2 \right). \end{aligned} \quad (\text{F.56})$$

Hence, for $m > 1$

$$\begin{aligned} \tilde{\rho}_{t,t+m} &= \mathbb{E}(a'_{t-1} \mathbb{E}(a'_{t+m-1} Q a_{t+m-1} x_t x_t | \mathcal{V}_{t-1}) a_{t-1}) - \mathbb{E}(a'_{t-1} Q a_{t-1}) \mathbb{E}(a'_{t+m-1} Q a_{t+m-1}) \\ &= \pi_t \prod_{j=1}^{m-1} \phi_{t+j}, \end{aligned} \quad (\text{F.57})$$

where

$$\pi_t := \mathbb{E}(a'_{t-1} \mathbb{E}(a'_t Q a_t x_t x_t | \mathcal{V}_{t-1}) a_{t-1}) - \mathbb{E}(a'_{t-1} Q a_{t-1}) \mathbb{E}(a'_t Q a_t) = O(\gamma_t^2). \quad (\text{F.58})$$

In order to see that equation (F.58) holds, consider

$$\begin{aligned} a'_{t-1} \mathbb{E}(a'_t Q a_t x_t x_t | \mathcal{V}_{t-1}) a_{t-1} &= (a'_{t-1} Q a_{t-1})^2 \\ &\quad - 2c \gamma_t (a_{t-1} \otimes a_{t-1}) \tilde{Q}^{(4)}(a_{t-1} \otimes a_{t-1}) \\ &\quad + \gamma_t^2 \sigma^2 a'_{t-1} Q^{(4)} a_{t-1} \\ &\quad + (c \gamma_t)^2 (a_{t-1} \otimes a_{t-1}) \tilde{Q}^{(6)}(a_{t-1} \otimes a_{t-1}) \\ &= (1 - c \gamma_t (2K_1 - c \gamma_t K_2)) (a'_{t-1} Q a_{t-1})^2 + K_3 \gamma_t^2 \sigma^2 a'_{t-1} Q a_{t-1}, \end{aligned} \quad (\text{F.59})$$

where the existence of the constants $K_i > 0$ ($i = 1, 2, 3$) follows from equations (F.27), (F.28) and (F.30). The claim follows upon taking expectations. \square

Proof of equation (b). The proof is an immediate extension of the proof of equation (b) of lemma A.5. \square

Proof of equation (c). Note that

$$\text{var}(z_t) = \sigma^2 \mathbb{E}(\tilde{x}_t^2) = \sigma^2 \text{tr} \left[Q \mathbb{E} \left(a_{t-1}^* a_{t-1}^{*'} \right) \right], \quad (\text{F.60})$$

so that by the preceding lemma

$$\frac{1}{T^b} \sum_{t=1}^T \text{var}(z_t) = \sigma^2 \text{tr} \left[Q \frac{1}{T^b} \sum_{t=1}^T \mathbb{E} \left(a_{t-1}^* a_{t-1}^{*'} \right) \right] \rightarrow k \sigma^4 \frac{\gamma}{2cb}. \quad (\text{F.61})$$

The claim now follows by the same arguments as used in the proof of equation (c) of lemma A.5. \square

G Extension B: constant gain

Throughout assumption 2, 3 and 4 are assumed to hold. Further, suppose assumption 1 holds with $\eta = 0$, i.e.

Assumption 1.C. For all t $\gamma_t = \gamma > 0$ so that $\gamma < 2\kappa_x^{(2)2}/(c\kappa_x^{(4)})$.

The following proposition will be proven.

Proposition G.1. Suppose that assumptions 1.C, 2, 3 and 4 hold. Then, as $T \rightarrow \infty$ and $\gamma \rightarrow 0^+$ so that $\gamma^2 T \rightarrow \infty$

$$(\gamma T)^{1/2}(\widehat{\lambda} - \lambda) \xrightarrow{d} \mathcal{N}\left(0, 2c \begin{bmatrix} 1 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix}\right) \quad (a)$$

$$T^{1/2}t'_\alpha(\widehat{\lambda} - \lambda) \xrightarrow{d} \mathcal{N}(0, (\tau/c)^2). \quad (b)$$

Proof. To begin with, lemma A.5 is extended to the present case of constant gains.

Lemma G.1. Recall that $\tilde{x}_{0t} = x_t a_{t-1}^*$ and let the conditions of proposition G.1 hold. Then,

$$\frac{1}{\gamma T} \sum_{t=1}^T \tilde{x}_{0t}^2 \xrightarrow{p} \frac{\sigma^2}{2c} \quad (a)$$

$$\frac{1}{T^{1/2}} \sum_{t=1}^T x_t \tilde{x}_{0t} = O_p(1) \quad (b)$$

$$\frac{1}{(\gamma T)^{1/2}} \sum_{t=1}^T \tilde{x}_{0t} \varepsilon_t \xrightarrow{d} \mathcal{N}(0, \sigma^4/(2c)). \quad (c)$$

The proof of lemma G.1 is delegated to the end. Closely mirroring the proof of proposition 2.1 contained in appendix B, consider

$$(\gamma T)^{1/2}(\widehat{\lambda} - \lambda) = \gamma T W_T^{-1} \frac{1}{(\gamma T)^{1/2}} \sum_{t=1}^T w_t \varepsilon_t = \frac{\gamma T^2}{\det W_T} \left(\frac{Q_T/T}{(\gamma T)^{1/2}} \sum_{t=1}^T w_t \varepsilon_t \right). \quad (G.1)$$

The main idea is to show that

$$\text{plim}_{\gamma T} \frac{\det W_T}{\gamma T^2} = \sigma^2 \kappa_x^{(2)}/(2c) \quad (i)$$

$$\text{plim}_{\gamma T} Q_T/T = \kappa_x^{(2)} \begin{bmatrix} 1 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix} \quad (ii)$$

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

where the respective limits are taken as $T \rightarrow \infty$ and $\gamma \rightarrow 0^+$ so that $\gamma^2 T \rightarrow \infty$.

Step (i). Analogous to (B.6), consider

$$\frac{\det W_T}{\gamma T^2} = m_T \left[\frac{1}{\gamma T} \sum_{t=1}^T \tilde{x}_{0t}^2 - m_T^{-1} \left(\frac{1}{\gamma^{1/2} T} \sum_{t=1}^T x_t \tilde{x}_{0t} \right)^2 \right]. \quad (\text{G.2})$$

The claim follows by lemma G.1 and the almost sure convergence of m_T . \square

Step (ii). Taking the above into account, the proof is analogous to that of step (a) in appendix B.

Step (iii). Similarly to (B.14), the entries of the 2×1 vector

$$\frac{Q_T/T}{\sqrt{\gamma T}} \sum_{t=1}^T w_t \varepsilon_t \quad (\text{G.3})$$

can be written as

$$\begin{aligned} \frac{Q_T/T}{\sqrt{\gamma T}} \sum_{t=1}^T w_t \varepsilon_t &= \begin{bmatrix} 1 \\ -\alpha \end{bmatrix} \frac{m_T}{\sqrt{\gamma T}} \sum_{t=1}^T \tilde{x}_{0t} \varepsilon_t \\ &\quad - \begin{bmatrix} 1 \\ -\alpha \end{bmatrix} \left(\frac{\kappa_x^{(2)}}{\gamma^{1/2} T} \sum_{t=1}^T x_t \tilde{x}_{0t} \right) u_T \\ &\quad + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \left\{ \left(\frac{\kappa_x^{(2)}}{\gamma^{1/2} T} \sum_{t=1}^T \tilde{x}_{0t}^2 \right) u_T - \left(\frac{1}{T} \sum_{t=1}^T x_t \tilde{x}_{0t} \right) \frac{1}{\sqrt{\gamma T}} \sum_{t=1}^T \tilde{x}_{0t} \varepsilon_t \right\} \\ &= \begin{bmatrix} 1 \\ -\alpha \end{bmatrix} \frac{m_T}{\sqrt{\gamma T}} \sum_{t=1}^T \tilde{x}_{0t} \varepsilon_t + O_p \left(\frac{1}{\sqrt{\gamma T}} \right) + O_p(\sqrt{\gamma}) + O_p \left(\frac{1}{\sqrt{T}} \right), \end{aligned} \quad (\text{G.4})$$

using that $u_T = O_p(1)$ (by the CLT for *i.i.d.* random variables) together with lemma G.1. Hence, it follows from equation (c) of lemma G.1 and the almost sure convergence of m_T that

$$\begin{bmatrix} 1 \\ -\alpha \end{bmatrix} \frac{\kappa_x^{(2)}}{\sqrt{\gamma T}} \sum_{t=1}^T \tilde{x}_{0t} \varepsilon_t \xrightarrow{d} \mathcal{N} \left(0, \frac{\kappa_x^{(2)2} \sigma^4}{2c} \begin{bmatrix} 1 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix} \right). \quad (\text{G.5})$$

Step (i) of this proof in conjunction with Slutsky's theorem gives the stated result. \square

Based on the proof of corollary 2.1 contained in appendix C, note that

$$\begin{aligned}
T^{1/2} \iota'_\alpha (\widehat{\lambda} - \lambda) &= \frac{\gamma T^2}{\det W_T} \frac{1}{T^{1/2}} \sum_{t=1}^T a_{t-1} x_t \varepsilon_t \left(\frac{\alpha}{\gamma T} \sum_{t=1}^T x_t^2 - \frac{1}{\gamma T} \sum_{t=1}^T x_t^2 a_{t-1} \right) \\
&\quad + \frac{\gamma T^2}{\det W_T} \frac{1}{T^{1/2}} \sum_{t=1}^T x_t \varepsilon_t \left(\frac{1}{\gamma T} \sum_{t=1}^T (x_t a_{t-1})^2 - \frac{\alpha}{\gamma T} \sum_{t=1}^T x_t^2 a_{t-1} \right) \\
&= u_\tau \left(\frac{\kappa_x^{(2)} \gamma T^2}{\det W_T} \frac{1}{\gamma T} \sum_{t=1}^T \tilde{x}_{0t}^2 \right) \\
&\quad - \frac{\gamma T^2}{\det W_T} \left(\frac{1}{\gamma^{1/2} T} \sum_{t=1}^T x_t \tilde{x}_{0t} \right) \left(\frac{1}{(\gamma T)^{1/2}} \sum_{t=1}^T \tilde{x}_{0t} \varepsilon_t \right). \tag{G.6}
\end{aligned}$$

Taking account of equations (i) and (a), the first term of the right-hand side of equation (G.6) converges in distribution to a mean-zero normal random variable with variance τ^2 . By equations (i), (b) and (c), the second term is seen to be $o_p(1)$. Finally, it follows from the same arguments as that used in proving corollary 2.1 that

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

□

Proof of lemma G.1.

Proof of equation (a). Recalling that

$$a_t = a_{t-1} + \gamma(u_t - c x_t^* a_{t-1}^*) \tag{G.7}$$

and taking equation (A.76) into account, one obtains

$$\mathbb{E}(a_t^{*2} | \mathcal{V}_{t-1}) = a_{t-1}^{*2} + \gamma^2 (\tau^2 + c^2 \kappa_x^{(4)} / \kappa_x^{(2)2} a_{t-1}^{*2}) - 2c\gamma a_{t-1}^{*2}, \tag{G.8}$$

and hence

$$\mathbb{E}(a_t^{*2}) = \mathbb{E}(a_{t-1}^{*2}) \psi^t + \gamma^2 \tau^2 = \kappa_a^{(2)} \psi^t + \gamma^2 \tau^2 \sum_{i=0}^{t-1} \psi^i, \tag{G.9}$$

where

$$\psi := 1 - \theta \left(2 - \theta \kappa_x^{(4)} / \kappa_x^{(2)2} \right) \quad \text{with } \theta := c\gamma. \tag{G.10}$$

Therefore,

$$\frac{1}{\gamma T} \sum_{t=1}^T \mathbb{E}(a_t^{*2}) = \kappa_a^{(2)} \frac{1}{\gamma T} \sum_{t=1}^T \psi^t + \tau^2 \frac{\gamma}{T} \sum_{t=1}^T \sum_{i=0}^{t-1} \psi^i =: \kappa_a^{(2)} A_T + \tau^2 B_T, \quad (\text{G.11})$$

say. To begin with, consider

$$A_T = \frac{1}{\gamma T} \frac{\psi(1 - \psi^T)}{1 - \psi} = \frac{1}{\gamma T} \frac{\psi(1 - \psi^T)}{\theta(2 - \theta\kappa_x^{(4)}/\kappa_x^{(2)2})}. \quad (\text{G.12})$$

Now, a Taylor-series expansion at $\gamma \rightarrow 0^+$ yields

$$\frac{1}{2 - \theta\kappa_x^{(4)}/\kappa_x^{(2)2}} = \frac{1}{2} + \gamma \frac{c\kappa_x^{(4)}}{4\kappa_x^{(2)2}} + O(\gamma^2). \quad (\text{G.13})$$

Since $\psi(1 - \psi^T) < \infty$ and, by assumption, $\gamma^2 T \rightarrow \infty$, it follows that $A_T = O(\gamma^{-2} T^{-1}) = o(1)$.

Next,

$$B_T = \frac{\gamma}{1 - \psi} - \frac{\gamma}{T} \frac{\psi(1 - \psi^T)\psi}{(1 - \psi)^2}. \quad (\text{G.14})$$

From equation (G.13), we deduce that

$$\frac{\gamma}{1 - \psi} = \frac{\gamma}{\theta(2 - \theta\kappa_x^{(4)}/\kappa_x^{(2)2})} = \frac{1}{2c} + \gamma \frac{\kappa_x^{(4)}}{4\kappa_x^{(2)2}} + O(\gamma^2), \quad (\text{G.15})$$

while, similar to (G.13),

$$\frac{1}{(1 - \psi)^2} = \frac{1}{4} + \gamma \frac{c\kappa_x^{(4)}}{4\kappa_x^{(2)2}} + O(\gamma^2). \quad (\text{G.16})$$

Hence,

$$B_T = \frac{1}{2c} + O\left(\frac{\gamma}{T}\right) \quad (\text{G.17})$$

and

$$\frac{1}{\gamma T} \sum_{t=1}^T \mathbb{E}(\tilde{x}_{0t}^2) = \frac{\kappa_x^{(2)}}{\gamma T} \sum_{t=1}^T \mathbb{E}(a_{t-1}^{*2}) \rightarrow \frac{\sigma^2}{2c}. \quad (\text{G.18})$$

Remark. It follows that $\mathbb{E}(a_t^*)/\gamma$ and $\mathbb{E}(a_t^{*2})/\gamma^2$ are bounded. Since $\gamma < 2\kappa_x^{(2)2}/(c\kappa_x^{(4)})$, $\psi < 1$ so that

$$\lim_{t \rightarrow \infty} \mathbb{E}(a_t^{*2}) = (\tau\gamma)^2 \sum_{i=0}^{\infty} \psi^i = \frac{(\tau\gamma)^2}{2\theta - \theta^2\kappa_x^{(4)}/\kappa_x^{(2)2}} = (\tau\gamma)^2 \left[\frac{1}{2\theta} + \frac{\kappa_x^{(4)}/\kappa_x^{(2)2}}{4} + O(\gamma) \right]. \quad (\text{G.19})$$

Hence,

$$\lim_{\gamma \rightarrow 0^+} \lim_{t \rightarrow \infty} \frac{\mathbb{E}(a_t^{*2})}{\gamma} = \frac{\tau^2}{2c}, \quad (\text{G.20})$$

implying that $\mathbb{E}(a_t^{*2}) = O(\gamma)$. Similarly, it can be shown (using lemma A.3) that for some constant $C > 0$ and t sufficiently large

$$\mathbb{E}(a_t^{*4}) \leq (1 - \theta)\mathbb{E}(a_{t-1}^{*4}) + C\gamma^3 = \kappa_a^{(4)}(1 - \theta)^t + C\gamma^3 \sum_{i=0}^{t-1} (1 - \theta)^i, \quad (\text{G.21})$$

and

$$\lim_{\gamma \rightarrow 0^+} \lim_{t \rightarrow \infty} \frac{\mathbb{E}(a_t^{*4})}{\gamma^2} = O(1). \quad (\text{G.22})$$

For simplicity, set $a_t := a_t^*$. By equation (G.18) and the weak LLN, the claim follows if

$$\text{var} \left(\frac{1}{\gamma T} \sum_{t=1}^T \tilde{x}_{0t}^2 \right) = o(1). \quad (\text{G.23})$$

Analogously to lemma A.5, set

$$\tilde{\rho}_{t,t+m} := \text{cov}(\tilde{x}_{0t}^2, \tilde{x}_{0,t+m}^2). \quad (\text{G.24})$$

Then for $m \geq 1$,

$$\tilde{\rho}_{t,t+m} = \kappa_x^{(2)} \pi_t \psi^{m-1},$$

where

$$\pi_t := \mathbb{E}(a_{t-1}^4) \tilde{\psi} - \mathbb{E}(a_{t-1}^2) \left[\kappa_x^{(2)} \mathbb{E}(a_{t-1}^2) \psi - \sigma^2 \gamma^2 (\kappa_x^{(4)} / \kappa_x^{(2)2} - 1) \right] = O(\gamma^2), \quad (\text{G.25})$$

(using equation (G.22)) with $\tilde{\psi} := \kappa_x^{(2)} - \theta(2\kappa_x^{(4)} / \kappa_x^{(2)} - \theta\kappa_x^{(6)} / \kappa_x^{(2)2})$ and $\psi := 1 - \theta(2 - \theta\kappa_x^{(4)} / \kappa_x^{(2)2})$. Therefore,

$$\text{var} \left(\frac{1}{\gamma T} \sum_{t=1}^T \tilde{x}_t^2 \right) = \frac{1}{\gamma^2 T^2} \sum_{t=1}^T \text{var}(\tilde{x}_{0t}^2) + \frac{2}{\gamma^2 T^2} \sum_{t=1}^{T-1} \sum_{s=1}^{T-t} \tilde{\rho}_{t,t+s}, \quad (\text{G.26})$$

where, as defined above, $\tilde{\rho}_{t,t+m} = \text{cov}(\tilde{x}_{0t}^2, \tilde{x}_{0,t+m}^2)$. Since $\text{var}(\tilde{x}_{0t}^2) = O(\gamma^2)$ by equation (G.22), the first term behaves like $O(1/T)$. Turning to the second summand, recall that $\pi_t = O(\gamma^2)$ so

that

$$\begin{aligned}
\frac{1}{\gamma^2 T^2} \sum_{t=1}^{T-1} \sum_{s=1}^{T-t} \tilde{\rho}_{t,t+s} &= \frac{1}{\gamma^2 T^2} \sum_{t=1}^{T-1} \pi_t \sum_{s=1}^{T-t} \psi^{s-1} \\
&= O\left(\frac{1}{T^2}\right) \sum_{t=1}^{T-1} \sum_{s=1}^{T-t} \psi^{s-1} \\
&= O\left(\frac{1}{T^2}\right) \left[\frac{\psi^T - 1}{(1-\psi)^2} + \frac{T}{1-\psi} \right] = O\left(\frac{1}{\gamma^4 T^2}\right) + O\left(\frac{1}{\gamma^2 T}\right), \tag{G.27}
\end{aligned}$$

which completes the proof of equation (a) because, by assumption, $\gamma^2 T \rightarrow \infty$. \square

Proof of equation (b). A recursive representation similar to (A.4) is readily derived, i.e.

$$a_t^* = (1-\theta)^t a_0^* + \gamma \sum_{i=0}^{t-1} (1-\theta)^i e_{t-i} + \gamma \sum_{i=0}^{t-1} (1-\theta)^i u_{t-i}, \tag{G.28}$$

where e_t and u_t have been defined in (A.2). Let

$$\tilde{u}_t := \sum_{i=0}^{t-1} (1-\theta)^i u_{t-i} \tag{G.29}$$

$$\tilde{e}_t := \sum_{i=0}^{t-1} (1-\theta)^i e_{t-i}, \tag{G.30}$$

so that (G.28) can be written as

$$a_t^* = (1-\theta)^t a_0^* + \gamma \tilde{u}_t + \gamma \tilde{e}_t. \tag{G.31}$$

Taking equation (G.31) into account, it follows that

$$\begin{aligned}
\frac{1}{T^{1/2}} \sum_{t=1}^T x_t^2 a_{t-1}^* &= \frac{a_0^*}{T^{1/2}} \sum_{t=1}^T x_t^2 (1-\theta)^{t-1} + \frac{\gamma}{T^{1/2}} \sum_{t=1}^T x_t^2 \tilde{u}_{t-1} + \frac{\gamma}{T^{1/2}} \sum_{t=1}^T x_t^2 \tilde{e}_{t-1} \\
&=: A_T + B_T + C_T, \tag{G.32}
\end{aligned}$$

We seek to show that the three terms A_T , B_T and C_T are $O_p(1)$. Begin with A_T and note that

$$\mathbb{E}|A_T| = \frac{\mathbb{E}|a_0^*| \kappa_x^{(2)}}{T^{1/2}} \sum_{t=1}^T (1-\theta)^{t-1} = O\left(\frac{1}{\gamma T^{1/2}}\right), \tag{G.33}$$

using assumption 2 and the fact that $\sum_{t=1}^T (1-\theta)^{t-1} = \theta^{-1} (1 - (1-\theta)^T) = O(1/\gamma)$. Furthermore, it is seen that

$$\mathbb{E}(A_T^2) = \frac{\kappa_a^{(2)}}{T} \sum_{t,s=1}^T \mathbb{E}(x_t^2 x_s^2) (1-\theta)^{t-1} (1-\theta)^{s-1} \leq \kappa_a^{(2)} \kappa_x^{(4)} \left(\frac{1}{T^{1/2}} \sum_{t=1}^T (1-\theta)^{t-1} \right)^2, \tag{G.34}$$

where Cauchy Schwarz's inequality has been used. The term in brackets on the majorant side of (G.34) has already been shown to be of size $O(\gamma^{-1}T^{-1/2})$, thereby proving that $A_T = O_p(1)$ if $\gamma^2 T \rightarrow \infty$. Before turning to B_T and C_T , note that the definitions of (G.29) and (G.30) imply that

$$\mathbb{E}(\tilde{u}_t \tilde{u}_s) = \begin{cases} \tau^2 \sum_{i=0}^{t-1} (1-\theta)^{2i} & \text{if } s = t \\ \tau^2 \sum_{i=0}^{s-1} (1-\theta)^{t-s+2i} & \text{if } s < t \end{cases} \quad (\text{G.35})$$

$$\mathbb{E}(\tilde{e}_t \tilde{e}_s) = \begin{cases} C_0 \sum_{i=0}^{t-1} (1-\theta)^{2i} \mathbb{E}(a_{t-1-i}^{*2}) & \text{if } s = t \\ C_0 \sum_{i=0}^{s-1} (1-\theta)^{t-s+2i} \mathbb{E}(a_{s-1-i}^{*2}) & \text{if } s < t, \end{cases} \quad (\text{G.36})$$

where $C_0 := c^2 (\kappa_x^{(4)} / \kappa_x^{(2)2} - 1)$. Consider B_T first and let

$$B_T = \frac{\kappa_x^{(2)} \gamma}{T^{1/2}} \sum_{t=1}^T \tilde{u}_{t-1} + \frac{\kappa_x^{(2)} \gamma}{T^{1/2}} \sum_{t=1}^T (x_t^{*2} - 1) \tilde{u}_{t-1} =: B_T^{(1)} + B_T^{(2)}, \quad (\text{G.37})$$

say. Now, from (G.35) one gets

$$\text{var}(B_T^{(1)}) = \frac{(\kappa_x^{(2)} \gamma)^2 \gamma^2}{T} \sum_{t=1}^T \sum_{i=0}^{t-1} (1-\theta)^{2i} + \frac{2(\kappa_x^{(2)} \gamma)^2 \gamma^2}{T} \sum_{t=2}^T \sum_{s=1}^{t-1} \sum_{i=0}^{s-1} (1-\theta)^{t-s+2i}. \quad (\text{G.38})$$

Note that

$$\begin{aligned} \sum_{t=1}^T \sum_{i=0}^{t-1} (1-\theta)^{2i} &= \frac{T}{(2-\theta)\theta} - \frac{1 - (1-\theta)^{2T} + \theta(2-\theta)((\theta-1)^{2T} - 1)}{(2-\theta)^2 \theta^2} \\ &= O\left(\frac{T}{\gamma}\right) + O\left(\frac{1}{\gamma^2}\right). \end{aligned} \quad (\text{G.39})$$

Similarly,

$$\begin{aligned} \sum_{t=2}^T \sum_{s=1}^{t-1} \sum_{i=0}^{s-1} (1-\theta)^{t-s+2i} &= \left[\frac{(1-\theta)^{T+1}}{\theta^3} - \frac{(\theta-1)^{2(T+1)} + (1-\theta)^2}{(2-\theta)^2 \theta^3} \right] + T \frac{1-\theta}{(2-\theta)\theta^2} - \frac{1-\theta}{(2-\theta)^2 \theta} \\ &= O\left(\frac{1}{\gamma^3}\right) + O\left(\frac{T}{\gamma^2}\right) + O\left(\frac{1}{\gamma}\right). \end{aligned} \quad (\text{G.40})$$

Hence it follows that $\text{var}(B_T^{(1)}) = O(1)$ and thus $B_T^{(1)} = O_p(1)$ as $\mathbb{E}(B_T^{(1)}) = 0$. Turning to the partial sum $B_T^{(2)}$, note that its summands are uncorrelated with mean zero so that

$$\text{var}(B_T^{(2)}) = \frac{\tau^2 (\kappa_x^{(4)} - \kappa_x^{(2)2}) \gamma^2}{T} \sum_{t=1}^T \sum_{i=0}^{t-1} (1-\theta)^{2i} = O(\gamma) + O\left(\frac{1}{T}\right), \quad (\text{G.41})$$

using the same arguments as before. Hence, $B_T = O_p(1)$. Finally, consider C_T and let

$$C_T = \frac{\kappa_x^{(2)}\gamma}{T^{1/2}} \sum_{t=1}^T \check{\varepsilon}_{t-1} + \frac{\kappa_x^{(2)}\gamma}{T^{1/2}} \sum_{t=1}^T (x_t^{*2} - 1)\check{\varepsilon}_{t-1} =: C_T^{(1)} + C_T^{(2)}, \quad (\text{G.42})$$

say. Equation (G.20) in conjunction with equation (G.36) shows that

$$\text{var}(C_T^{(1)}) = O(\gamma)\text{var}(B_T^{(1)}) \quad \text{and} \quad \text{var}(C_T^{(2)}) = O(\gamma)\text{var}(B_T^{(2)}), \quad (\text{G.43})$$

what proves equation (b). □

Proof of equation (c). Note that

$$z_t := \tilde{x}_{0t}\varepsilon_t = a_{t-1}^* x_t \varepsilon_t \quad (\text{G.44})$$

forms a martingale difference sequence with respect to \mathcal{V}_t . Furthermore, define $Z_{tT} := z_t/s_T$, where

$$s_T^2 := \sum_{t=1}^T \sigma_t^2 \quad \text{with} \quad \sigma_t^2 := \text{E}(z_t^2) = \sigma^2 \text{E}(a_{t-1}^{*2}) \quad (\text{G.45})$$

and observe that by definition of z_t and Z_{tT} ,

$$\frac{1}{(\gamma T)^{1/2}} \sum_{t=1}^T \tilde{x}_{0t}\varepsilon_t = \left(\frac{1}{\gamma T} s_T^2 \right)^{1/2} \sum_{t=1}^T Z_{tT}. \quad (\text{G.46})$$

Now, taking account of equation (a) of this lemma, it is seen that

$$\frac{s_T^2/(\gamma T)}{\sigma^4/(2c)} = 1 + o(1). \quad (\text{G.47})$$

Consequently, the claim follows if $\sum_{t=1}^T Z_{tT} \xrightarrow{d} N(0, 1)$. According to Davidson's (1994) theorem 24.3, it suffices to verify that the following two conditions hold:

1. $\sum_{t=1}^T Z_{tT} \xrightarrow{p} 1$;
2. $\max_{1 \leq t \leq T} |Z_{tT}| \xrightarrow{p} 0$.

By part (1), the first condition is satisfied. Using theorem 23.10 and 23.16 in Davidson (1994) together with equation (G.22), also the second condition is seen to hold as z_t^2 is uniformly integrable. This completes the proof. □