Estimation and Inference in Adaptive Learning Models with Slowly Decreasing Gains

Alexander Mayer*

April 18, 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 2018 for helpful discussions and suggestions.

1 Introduction

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

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

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

$$y^e_{t|t-1} = \mathrm{E}(y_t \,|\, \mathcal{F}_{t-1}), \tag{2}$$

thereby yielding the RE equilibrium

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

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

$$y^e_{t|t-1} = a_{t-1} x_t, \tag{4}$$

where the point forecast a_t of α is obtained via stochastic approximation algorithms (see e.g. Sargent (1993) or Evans and Honkapohja (2001)). Motivated by the idea of economic agents aiming at minimizing the expected forecast error recursively, the learning scheme is assumed to take the form of a least-squares type stochastic approximation algorithm

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

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 examined. 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 can be 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 Christopeit 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 mutually independent and identically distributed with finite variances so that*

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

Under assumption 2, the recursion (5) can be viewed as a perturbed Robbins-Monro algorithm, with a small approximation error resulting from the estimation of the regressor second moment. Robbins-Monro algorithms 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), Evans and Honkapohja (2001, pp. 125) or Kushner and Yin (2010, pp. 6).

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 estimatability 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.

Finally, the distributional characteristics of v_t are sharpened in order to develop an appropriate asymptotic theory.

Assumption 3. *Let $\kappa_\varepsilon^{(\ell)}$ and $\kappa_x^{(\ell)}$ denote the ℓ^{th} non-central moment of ε_t and x_t, respectively.*

(a) $\theta_t := (a_t, r_t)'$ *is bounded almost surely on a compact subset of* $\mathbb{R} \times \mathbb{R}_+$.

(b) $\kappa_x^{(16)} < \infty$ *and* $\kappa_\varepsilon^{(8)} < \infty$.

A note on assumption 3 seems warranted. Though clearly strong, part (a) is not uncommon in the stochastic approximation literature where many authors restrict θ_t to a compact neighborhood of $(\alpha, \kappa_x^{(2)})'$ by the use of so-called 'projection-facilities': see, e.g., the discussion in Kushner and Yin (2010) on constrained algorithms. Although this assumption might be regarded as overly restrictive (see, e.g., Evans and Honkapohja (1997)), it will nevertheless be retained for reasons of analytical tractability. The reason for imposing relatively[1] high moment conditions on v_t is the need for controlling the approximation error stemming from replacing $\kappa_x^{(2)}$ by r_t; with the added difficulty that r_t appears in the denominator of a_t. Note that for the mere almost sure convergence of r_t to $\kappa_x^{(2)}$, finite fourth moments of the regressor are sufficient, c.f. lemma F.1 of appendix F. In fact, if, like in Evans and Honkapohja (1998), the update scheme (5) takes the form of the so-called stochastic-gradient algorithm

$$b_t = b_{t-1} + \gamma_t x_t (y_t - x_t b_{t-1}), \tag{SG}$$

i.e. $r_t = 1$ for all t, then assumption 3 might be replaced by the following:

Assumption 3-SG. *Let $\kappa_b^{(\ell)}$ denote the ℓ^{th} non-central moment of $b_0 - \alpha$ and assume that* $\kappa_b^{(2)} < \infty$, $\kappa_\varepsilon^{(4)} < \infty$ *and* $\kappa_x^{(6)} < \infty$.

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

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

$$\widehat{\lambda} := \left(\sum_{t=1}^{T} w_t w_t'\right)^{-1} \sum_{t=1}^{T} w_t y_t \quad \text{with} \quad w_t := (a_{t-1} x_t, x_t)'. \tag{8}$$

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

$$M_T := \sum_{t=1}^{T} w_t w_t' \tag{9}$$

[1]Even stronger moment conditions are imposed by, for example, Kuan and White (1994) who require their counterpart of v_t to be bounded almost surely.

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

$$M_T/T \xrightarrow{p} \kappa_x^{(2)} \begin{bmatrix} \alpha^2 & \alpha \\ \alpha & 1 \end{bmatrix} \quad \text{and} \quad T^b M_T^{-1} \xrightarrow{p} \frac{2cb}{\gamma\sigma^2} \begin{bmatrix} 1 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix}, \tag{10}$$

where $b := 1 - \eta$ is smaller the more quickly the gain parameter $\gamma_t = \gamma/t^\eta$ approaches zero as $t \to \infty$; cf. appendix D. Intuitively, the singularity of the empirical second moment matrix stems from the fact that the entries of w_t are asymptotically collinear, since $a_t \sim \alpha$ for large t. The other crucial aspect of (10) 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.

$$\frac{\det M_T}{T^{1+b}} \xrightarrow{p} \frac{\kappa_x^{(2)}\gamma\sigma^2}{2cb}; \tag{11}$$

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

$$T^{b/2}(\widehat{\lambda} - \lambda) \overset{a}{\sim} \frac{2cb}{\gamma\sigma^2} \begin{bmatrix} 1 \\ -\alpha \end{bmatrix} \frac{1}{T^{b/2}} \sum_{t=1}^{T} a_{t-1}^* x_t \varepsilon_t, \tag{12}$$

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 and 3 hold. Then*

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

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

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

Proof. See appendix D.

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 classical estimator of V 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 $x_t = 1$, Christopeit and Massmann (2018) establish a similar result in their theorem 4 for the case of recursive-least squares learning (i.e. $\eta = 1$). The (singular) variance covariance matrix in Christopeit and Massmann (2018) is equivalent to that stated above up to a factor of proportionality which stems from the different specification of the gain; see also the discussion in Christopeit and Massmann (2015, pp. 19). The crucial difference between their result and proposition 2.1 is the rate of convergence: while Christopeit and Massmann (2018) show that $\widehat{\lambda}$ converges at a logarithmic rate, it is seen that for slowly decreasing gains $\widehat{\lambda}$ converges at a faster, polynomial rate. A trade-off between the rate at which the agent learns α (increasing in η) and the convergence rate of $\widehat{\lambda}$ (decreasing in η) becomes thus apparent–with the limiting case of $\eta = 1$ treated by Christopeit and Massmann (2018).

The singularity of the limiting distribution of $\widehat{\lambda}$ means that a linear combination of its entries, namely, $\iota_\alpha'\widehat{\lambda}$ with $\iota_\alpha := (\alpha,\, 1)'$, converges at a rate higher than $T^{b/2}$.[2] 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 $\iota_\alpha'\lambda = \alpha$).

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

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

where I_k *denotes the* $k-$*dimensional identity matrix. Then*

$$G_T D_\alpha(\widehat{\lambda} - \lambda) \xrightarrow{d} \mathcal{N}\left(0_3, \begin{bmatrix} V & 0_2 \\ 0_2' & \tau^2 \end{bmatrix}\right), \quad \textit{with } D_\alpha := \begin{bmatrix} I_2 \\ \iota_\alpha' \end{bmatrix}.$$

Proof. See appendix E.

[2]For another example from the econometric literature where the singularity of the limiting distribution arises due to the super-consistency of a linear-combination of the joint estimator, see remark 2.4 below or the discussion in Lütkepohl and Burda (1997) on testing noncausality restrictions using vector autoregressions.

Remark 2.2. *In certain cases, the learning literature suggests the plausibility of constant gain sequences instead of decreasing gains, i.e.* $\eta = 0$ *so that* $\gamma_t = \gamma > 0$*. It follows that the preceding results continue to hold without material differences as long as* γ *is 'sufficiently' small. Specifically, the asymptotic theory in this case is based on a limiting expansion of* γ *and* T*, whereby* $\gamma \searrow 0$ *and* $T \to \infty$ *so that* $\gamma^2 T \not\to 0$*. 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] *Again, it follows that the joint OLS estimator* $\widehat{\lambda}$ *is asymptotically normal while the linear combination* $\iota_\alpha' \widehat{\lambda}$ *is 'super-consistent'. To see this, let*

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

denote the three-dimensional normalizing matrix and assume for simplicity that $r_t = \kappa_x^{(2)}$ *for all* t*. Then, given the assumptions of proposition 2.3 hold and* $\gamma \searrow 0$*,* $T \to \infty$ *so that* $\gamma^2 T \not\to 0$*, it follows that*

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

Remark 2.3. *The limiting theory also 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. In this case, the agent's updating scheme for the* k *dimensional RE equilibrium vector* α *is given by*

$$\begin{aligned} a_t &= a_{t-1} + \gamma_t R_t^{-1} x_t (y_t - a_{t-1}' x_t) \\ R_t &= R_{t-1} + \gamma_t (x_t x_t' - R_{t-1}). \end{aligned} \tag{13}$$

Consider the $k+1$ *dimensional OLS estimator* $\widehat{\lambda}$ *of* $\lambda := (\beta, \delta')'$ *based on the actual law of motion*

$$y_t = \beta a_{t-1}' x_t + \delta' x_t + \varepsilon_t \quad \text{for } t = 1, 2, \ldots, T. \tag{14}$$

Just as in the scalar case treated above, it follows that $T^{b/2}(\widehat{\lambda} - \lambda)$ *is asymptotically normal while the linear combination* $I_\alpha(\widehat{\lambda} - \lambda)$*, with* $I_\alpha := (\alpha, I_k)$*, converges to its limiting normal distribution at the faster rate* $T^{1/2}$ *(note that* $I_\alpha \lambda = \alpha$*). Specifically, let* $D_\alpha := (I_{k+1}, I_\alpha')'$ *and define the multivariate scaling matrix of dimension* $(2k+1) \times (2k+1)$

$$G_T := \begin{bmatrix} T^{b/2} I_{k+1} & \mathrm{O} \\ \mathrm{O}' & T^{1/2} I_k \end{bmatrix}, \tag{15}$$

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

where the O *block represents a* $(k+1)\times k$ *matrix of zeros. Under the additional condition that* $Q := \mathrm{E}(x_t x_t')$ *has full rank and assuming, for simplicity, that* $R_t = Q$ *for all* t*, it can be shown that*[4]

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

where $\ell_\alpha := (1, -\alpha')'$.

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

$$\widehat{\lambda}_\alpha := \frac{\widehat{\lambda}_\delta}{1-\widehat{\lambda}_\beta}, \tag{16}$$

where $\widehat{\lambda}_\beta$ and $\widehat{\lambda}_\delta$ denote respectively the two components of $\widehat{\lambda}$.[5] Since $\widehat{\lambda}_\alpha - \alpha = \iota_\alpha'(\widehat{\lambda}_\alpha - \lambda)/(1-\widehat{\lambda}_\beta)$, the asymptotic normality of (16) follows as a by-product of the preceding corollary, i.e.

$$T^{1/2}(\widehat{\lambda}_\alpha - \alpha) \xrightarrow{d} \mathcal{N}\left(0, (\tau/c)^2\right). \tag{17}$$

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 $\iota_\alpha'\widehat{\lambda}$ of corollary 2.1 as $\beta < 1$.

Remark 2.4. *Concerning the different convergence rates of* $\widehat{\lambda}$ *on the one hand and* $\widehat{\lambda}_\alpha$ *on the other, one is reminded of estimating co-integrating vectors based on autoregressive distributed lag* (ADL) *models. Specifically, consider the* ADL(1,0) *model*

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

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

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

[4]...

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

suggests the following two-step approach: 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 estimator of the co-integrating coefficient 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 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: $\delta = \delta_0$ for suitable values $\lambda_0 := (\beta_0, \delta_0)'$ using the joint OLS estimator $\widehat{\lambda}$. As a natural starting-point, consider the textbook t statistics

$$t_\beta := \frac{\widehat{\lambda}_\beta - \beta_0}{\sqrt{m^{11}\,\widehat{\sigma}^{\,2}}} \quad \text{and} \quad t_\delta := \frac{\widehat{\lambda}_\delta - \delta_0}{\sqrt{m^{22}\,\widehat{\sigma}^{\,2}}}, \tag{18}$$

where m^{ii} represents the i^{th} diagonal element of M_T^{-1} (with M_T being defined in (9)). As shown in the appendix, t_β and t_δ are asymptotically standard normally distributed, which allows us to draw inferences from $\widehat{\lambda}$ about the true structural parameters β and δ. The t statistic for α based on $\widehat{\lambda}_\alpha$ is similarly defined via $t_\alpha := (\widehat{\lambda}_\alpha - \alpha_0)/\mathrm{SE}(\widehat{\lambda}_\alpha)$ with a modified standard error given by

$$\mathrm{SE}(\widehat{\lambda}_\alpha) := \frac{\sqrt{\widehat{\sigma}^2/\sum_{t=1}^T x_t^2}}{1-\widehat{\lambda}_\beta}. \tag{19}$$

Note that t_α does not approach a 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 $H_0 : \alpha = \alpha_0$ for some hypothetical value $\alpha_0 \in \mathbb{R}$, cf. equation (17). The reason is that $T^{1/2}\mathrm{SE}(\widehat{\lambda}_\alpha)$ 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 (19).

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

$$V := (2cb/\gamma)Q \text{ with } Q := \begin{bmatrix} 1 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix}, \tag{20}$$

cf. 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

$$H_0\colon\ r(\lambda) = 0, \tag{21}$$

where $r(\lambda) := R\lambda - q$ with $q \in \mathbb{R}^2$ and R represents a suitable 2×2 matrix of linear restrictions. The null (21) can nevertheless be tested using a Wald-type statistic when 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

$$B^+ := A \begin{bmatrix} 1/\xi & 0 \\ 0 & 0 \end{bmatrix} A', \tag{22}$$

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). Since the pseudo-inverse (22) is not continuous and V_T violates Andrews' (1987) rank condition, namely, $\text{rk}(T^b V_T) > \text{rk}(V)$ so that $(T^b V_T)^+ = (T^b V_T)^{-1}$ does not converge in probability to V^+, an alternative variance-covariance estimator is needed. Taking (17) into account, a consistent alternative is given by $T^b \tilde{V}_T$ with

$$\tilde{V}_T := m^{11}\,\widehat{\sigma}^2 \begin{bmatrix} 1 & -\widehat{\lambda}_\alpha \\ -\widehat{\lambda}_\alpha & \widehat{\lambda}_\alpha^2 \end{bmatrix}. \tag{23}$$

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

$$\mathcal{W} := r(\widehat{\lambda})'(R\tilde{V}_T R')^+ r(\widehat{\lambda}). \tag{24}$$

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 against local alternatives in a $1/T^{b/2}-$neighborhood of the null; a question pursued further in section 3.

2.4 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, the true RE equilibrium parameter α is known and the interest lies in estimating β. Then, with a little rearrangement, the actual law of motion $y_t = \beta a_{t-1} x_t + \delta x_t + \varepsilon_t$ is rewritten as

$$y_t - \alpha x_t = \beta a^*_{t-1} x_t + \varepsilon_t. \tag{25}$$

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

$$\widehat{\beta}_0 = \frac{\sum_{t=1}^{T} (y_t - \alpha x_t) a^*_{t-1} x_t}{\sum_{t=1}^{T} (a^*_{t-1} x_t)^2}, \tag{26}$$

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

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

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

Proof. See appendix B. □

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 would resort to

$$\widehat{\alpha} = \frac{\sum_{t=1}^{T} y_t x_t}{\sum_{t=1}^{T} x_t^2}. \tag{27}$$

However, with agents' beliefs departing from RE, it is not clear from the outset why $\widehat{\alpha}$ should possess any desirable statistical properties. Fortunately, the misspecification error from neglecting the expectation $y^e_{t|t-1}$ is minor and the limiting distribution of (27) is normal when suitably scaled.

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

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

Proof. See appendix C. □

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 to estimate β in a second step by

$$\widehat{\beta}=\frac{\sum_{t=1}^{T}\tilde{y}_t\tilde{x}_t}{\sum_{t=1}^{T}\tilde{x}_t^2}, \tag{28}$$

where $\tilde{y}_t = y_t - \widehat{\alpha}x_t$ and $\tilde{x}_t = (a_{t-1}-\widehat{\alpha})x_t$. Returning thus to the co-integration analogy drawn in remark 2.4, 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. Specifically, knowledge of α does not improve estimation of β; see Stock (1987) for an analogous discussion of the co-integration case. This means, in particular, that estimation and inference is not contaminated by a 'generated regressor' problem as discussed in Pagan (1984), i.e. the limiting distribution of $T^{b/2}(\widehat{\beta}-\beta)$ is (asymptotically) unaffected by the first-step estimation of α and equivalent to that stated in proposition 2.2.

Finally, consider the 'three-step' estimator

$$\widehat{\delta}=\widehat{\alpha}(1-\widehat{\beta}). \tag{29}$$

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

$$\begin{aligned} T^{b/2}(\widehat{\delta}-\delta) &= T^{b/2}(\alpha(1-\widehat{\beta})-\delta)+T^{b/2}(\widehat{\alpha}-\alpha)(1-\widehat{\beta}) \\ &= -\alpha T^{b/2}(\widehat{\beta}-\beta)+O_p\left(T^{-\eta/2}\right) \end{aligned} \tag{30}$$

and thus

$$T^{b/2}\begin{bmatrix}\widehat{\beta}-\beta\\ \widehat{\delta}-\delta\end{bmatrix} \overset{a}{\sim} \begin{bmatrix}1\\ -\alpha\end{bmatrix} T^{b/2}(\widehat{\beta}-\beta) \xrightarrow{d} \mathcal{N}\left(0_2, \frac{2cb}{\gamma}\begin{bmatrix}1 & -\alpha\\ -\alpha & \alpha^2\end{bmatrix}\right), \tag{31}$$

thereby yielding the exact same limiting distribution as the one stated in proposition 2.1.[6]

[6] As 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 (25) feasible.

3 Monte-Carlo results

Consider a data generating process given by model (1) with $\delta = 1$ and $\beta = 1/3$ so that $\alpha = 3/2$. The regressor evolves over time according to the first-order autoregression

$$x_t = 1 + (1/2)x_{t-1} + u_t, \tag{32}$$

where $u_t \overset{i.i.d.}{\sim} (0, 1/2)$ is drawn from a t distribution with eight degrees of freedom while the innovation ε_t is *i.i.d.* distributed as a standardized χ^2 random variable with one degree of freedom. The recursion (5) is initialized at $\theta_0 = (1,1)'$ and the gain sequence is given by $\gamma_t = (3/2)t^{-\eta}$, $\eta \in \{1/5, 3/5\}$. The (Monte-Carlo analogue of the) asymptotic variance, bias and size of a two-sided t test at the 5% significance level is evaluated for the joint OLS (jOLS) approach (cf. section 2.2) and the three-step (3STEP) procedure (cf. section 2.4). Results for $T \in \{200, 2000, 20000\}$ using $10,000$ Monte Carlo repetitions are summarized by table 1.

Table 1: Simulation Results[a]

		T	$\eta = 1/5$ asy.var	$\eta = 1/5$ 100bias(b)	$\eta = 1/5$ size	$\eta = 3/5$ asy.var	$\eta = 3/5$ 100bias(b)	$\eta = 3/5$ size
β	jOLS	200	0.400	-1.512	4.396	0.332	-10.167	4.684
	3STEP		0.398	-1.606	4.448	0.327	-10.348	4.674
		2,000	0.512	-0.319	4.632	0.348	-4.580	5.160
			0.512	-0.328	4.620	0.347	-4.604	5.164
		20,000	0.696	-0.053	5.068	0.354	-1.849	5.182
			0.696	-0.054	5.066	0.354	-1.852	5.180
δ	jOLS	200	0.972	2.244	4.462	0.757	15.252	4.788
	3STEP		1.188	2.329	6.648	0.748	15.450	4.922
		2,000	1.198	0.482	4.722	0.784	6.871	5.188
			1.301	0.496	5.652	0.783	6.901	5.230
		20,000	1.473	0.081	5.142	0.797	2.773	5.160
			1.475	0.081	5.502	0.796	2.777	5.162
α	jOLS	200	0.469	-0.036	6.784	0.474	0.003	10.496
	3STEP		1.799	-0.117	6.510	0.469	-0.092	10.652
		2,000	0.465	0.006	5.358	0.463	0.001	7.432
			0.523	0.005	5.122	0.465	-0.008	7.472
		20,000	0.476	0.002	5.384	0.464	0.000	6.068
			0.509	0.000	5.142	0.465	-0.001	6.134

[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.

Recall the limiting variances of the estimators of (β, δ, α) as a benchmark, against which the following simulation evidence can be evaluated. Specifically, one infers from proposition 2.1

(or, equivalently, from proposition 2.2 and equation (30)) that the asymptotic variances for the estimators of $(\beta,\, \delta)$ are given by $(0.711,\, 1.6)$ if $\eta = 1/5$ and by $(0.356,\, 0.8)$ if $\eta = 3/5$, respectively. The asymptotic variance of $\widehat{\alpha}$ (and $\widehat{\lambda}_\alpha$) does not depend explicitly on the nature of the gain sequence and equals 0.482. For large T, the actual variances of the estimators are close to these values. In case of β and δ, the asymptotic approximation 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. Similarly, the bias increases relatively with 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 (24) is evaluated under the null

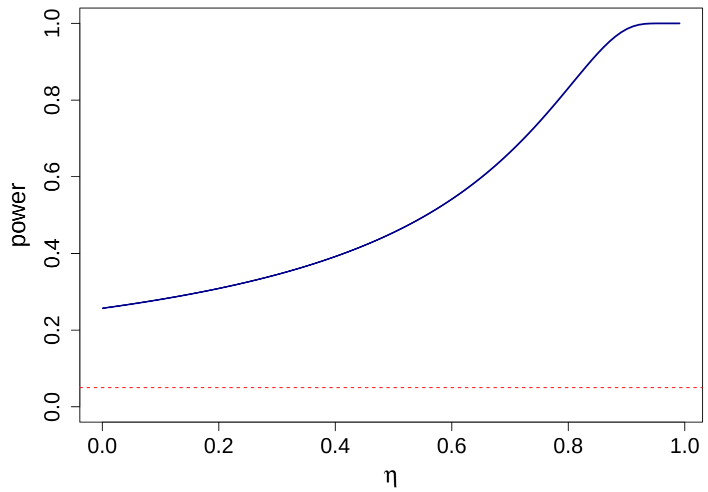
$$H_0\colon\ r(\lambda) = 0 \ \text{ with } \ r(\lambda) = R\lambda - q, \tag{33}$$

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

$$H_a\colon\ r(\lambda) = \mu / T^{b/2}, \tag{34}$$

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 $\mu' V^+ \mu$. In order to understand the local power properties,

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 $\mu = (1/2,\, 3)'$ and various η.

it is useful to take a look at the generalized inverse of V

$$V^+ = \frac{\gamma}{2cb(1+\alpha^2)^2} \begin{bmatrix} 1 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix}, \tag{35}$$

where equations (20) and (22) have been used.[7] Note that V^+ is positive semi-definite, i.e. there exist non-trivial local alternatives for which the non-centrality parameter is zero. Specifically, the test has no local power in the direction $\mu = c\iota_\alpha$ for some constant $c \neq 0$ as $\iota'_\alpha V^+ \iota_\alpha = 0$. On the other hand, if μ is not proportional to ι_α, power is increasing in η as illustrated by figure 1 for $\mu = (1/2, 3)'$.

Table 2: Simulation Results[a]

	$\eta = 1/5$			$\eta = 3/5$		
T	size	power I	power II	size	power I	power II
200	5.31	41.18	7.07	5.22	86.35	24.93
2,000	5.20	31.33	5.86	5.31	47.67	10.29
20,000	5.16	27.07	5.05	5.04	46.31	6.56

[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 (33) and the (local) alternative (34), respectively.

Monte-Carlo evidence supports these theoretical considerations: table 2 reports rejection frequencies under the null (size) and local power for $\mu = (1/2, 3)'$ (power I) and $\mu = \iota_\alpha$ (power II). It is seen that the size of the test is controlled, while the local power approaches the theoretical values given by 24.79% and 43.89% in case of $\mu = (1/2, 3)'$ for $\eta = 1/5$ and $\eta = 3/5$, and 5.00% in case of $\mu = \iota_\alpha$.

4 Concluding remarks

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

$$y_t = \beta y^e_{t|t-1} + \delta x_t + \varepsilon_t, \tag{36}$$

where the agents expectation formation is boundedly rational in the sense that $y^e_{t|t-1}$ obeys a stochastic approximation algorithm (cf. equation (5)). In contrast to Christopeit and Mass-

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

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

respectively.

mann (2018) who set $x_t = 1$, the current exposition treats the regressor as a time-varying random variable while the agent is assumed to use slowly-decreasing gains, i.e. $\gamma_t \sim t^{-\eta}$ with $\eta \in (0, 1)$. This represents the intermediate case lying on a continuum between least-squares learning ($\eta = 1$) and constant-gain learning ($\eta = 0$) considered by Christopeit and Massmann (2018).

The estimators of 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).

As suggested by the small-sample results, the theoretical analysis of this exposition could be extended to allow for serial correlated regressors or by relaxing assumption 3.

References

Abadir, K. M. and Magnus, J. R. (2006). *Matrix Algebra*. Camebridge Univerity 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. (1993). *Co-integration, Error Correction, and the Econometric Analysis of Non-stationary Data*. Oxford University Press.

Benveniste, A., Metivier, 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).

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. (1997). Convergence of learning algorithms without a projection facility. *Journal of Mathematical Economics*, 30:59–86.

Evans, G. W. and Honkapohja, S. (1998). Stochastic gradient learning in the cobweb model. *Economic Letters*, 61:333–337.

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. *AStA Advances in Statistical Analysis*, 90(1):59–74.

Jameson, G. J. (2016). The incomplete gamma functions. *The Mathematical Gazette*, 100(548):298–306.

Kuan, C.-M. and White, H. (1994). Adaptive learning with nonlinear dynamics driven by dependent processes. *Econometrica*, 62(5):1087–1114.

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 proceses. *Siam Journal on Control and Optimization*, 31(4):1045–1062.

Kushner, H. J. and Yin, G. G. (2010). *Stochastic Approximation and Recursive Algorithms and Applications,* 2^{nd} Edition. Springer.

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: Appendices A to E contain the proofs under the auxiliary assumption that r_t is centered at the true value $\kappa_x^{(2)}$, resulting in the reduced algorithm

$$b_t = b_{t-1} + \frac{\gamma_t}{\kappa_x^{(2)}}(y_t - b_{t-1}x_t). \tag{A.1}$$

Specifically, the following assumption replaces assumption 3:

Assumption 3A. *Suppose assumption 3-SG holds and for any t, $r_t = \kappa_x^{(2)}$.*

Thereafter, appendix F extends these results to the general algorithm (5) given assumption 1, 2 and 3 hold true.

A Preliminaries

This section derives the working formula for b_t and collects some helpful auxiliary lemmatae.

A.1 Working formula for b_t

With a little rearrangement, one gets from (A.1)

$$b_t^* = b_{t-1}^*(1 - c\gamma_t) + \gamma_t e_t + \gamma_t u_t, \tag{A.2}$$

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

$$e_t := c(1 - x_t^{*2})b_{t-1}^* \quad \text{with} \quad x_t^* := x_t/\sqrt{\kappa_x^{(2)}} \tag{A.3}$$

$$u_t := x_t^* \varepsilon_t/\sqrt{\kappa_x^{(2)}}, \tag{A.4}$$

which are, by construction, mean-zero. Solving recursively, it is further seen that for $t > 0$

$$b_t^* = a_0^* \Phi_{t,1} + \check{u}_t + \check{e}_t, \tag{A.5}$$

where

$$\Phi_{t,k} := \prod_{i=k}^{t}(1 - c\gamma_i) \quad \text{with} \quad \Phi_{t,t+1} := 1, \tag{A.6}$$

and

$$\gamma_{k,t} := \Phi_{t,k+1}\gamma_k, \tag{A.7}$$

while for any scalar z_t

$$\check{z}_t := \sum_{k=1}^{t} \gamma_{k,t} z_k. \tag{A.8}$$

A.2 Lemma 1

Lemma 1. *Let assumption 2 and 3A hold. For any s and t*

$$\mathrm{E}(u_t e_s) = 0, \tag{A.9}$$

while

$$\mathrm{E}(e_t e_s) = \begin{cases} 0 & \text{if } s \neq t \\ \sigma^2(e_t) := c^2(\kappa_x^{(4)}/\kappa_x^{(2)2} - 1)\mathrm{E}\left(b_{t-1}^{*2}\right) & \text{if } s = t. \end{cases} \tag{A.10}$$

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

$$\mathrm{E}\left(u_t b_{s-1}^*\right) = \mathrm{E}\left(u_t x_s^{*2} b_{s-1}^*\right).$$

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

$$\begin{aligned} \mathrm{E}\left(u_t x_s^{*2} b_{s-1}^*\right) &= \mathrm{E}\left[u_t b_{s-1}^* \mathrm{E}\left(x_s^{*2} | \mathcal{V}_{s-1}\right)\right] \\ &= \mathrm{E}\left(u_t b_{s-1}^*\right) \mathrm{E}\left(x_s^{*2}\right) = \mathrm{E}\left(u_t b_{s-1}^*\right), \end{aligned} \tag{A.11}$$

using repeatedly assumption 2. □

A.3 Lemma 2

This lemma summarizes properties of partial sums of the coefficients $\gamma_{k,t}$ (cf. equation (A.7)), which are frequently used in the subsequent proofs for the decreasing gain case. For simplicity assume $\gamma = 1$ throughout.

Lemma 2. *Define*

$$\phi_t^i := \sum_{k=1}^{t} \gamma_{k,t} \tag{A.12}$$

$$\phi_t^{ii} := \sum_{k=1}^{t} \gamma_{k,t}^2 \tag{A.13}$$

$$\phi_t^{iii} := \sum_{k=1}^{t-1} \sum_{s=1}^{k} \gamma_{s,t}\gamma_{s,k} \tag{A.14}$$

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

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

Remark A.1. *By the Cesàro mean convergence theorem, 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

$$c\phi_t^i + \Phi_{t,1} = 1. \tag{A.15}$$

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

$$\mathcal{H}_t(\eta) = \frac{t^{1-\eta}}{1-\eta} + \zeta(\eta) + \eta R_t^\infty(\eta) \text{ with } R_a^b(\eta) := \int_a^b \frac{u - \lfloor u \rfloor}{u^{1+\eta}} du, \tag{A.16}$$

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

$$\zeta(\eta) := \lim_{t\to\infty} \left(\mathcal{H}_t(\eta) - \frac{t^{1-\eta}}{1-\eta} \right), \tag{A.17}$$

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

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

see, e.g., theorem 3.2 *b*) in Apostol (1976, chap. 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 $r \ge 1$

$$(\Phi_{t,1})^r \le C_0 e^{-rat^b} \tag{A.18}$$

where $C_0 \in [0, \infty)$, $a := c/b$ and $b := 1 - \eta$. To see that (A.18) is true, consider the case $r = 1$ (the following is easily extended to the case $r > 1$) and note that

$$\begin{aligned}
\Phi_{t,1} &= e^{\sum_{j=1}^t \ln(1-c/j^\eta)} \\
&= e^{-c\sum_{j=1}^t 1/j^\eta} e^{-O(\sum_{j=1}^t 1/j^{2\eta})} \\
&\le e^{-c\mathcal{H}_t(\eta)} \\
&= e^{-c\zeta(1-b)} e^{-at^b} e^{-O(t^{-\eta})} \\
&\le e^{-c\zeta(1-b)} e^{-at^b},
\end{aligned}$$

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

$$\ln(1 - c/j^\eta) = -c/j^\eta - K_t \text{ with } K_t := c^2/(2j^{2\eta}) + c^3/(3j^{3\eta}) + \cdots = O(1/j^{2\eta}), \tag{A.19}$$

and equation (A.16), 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. [8]

A.3.2 Part (b)

The main idea of this proof is to approximate the partial sum $c^2\phi_t^{ii}$ by the integral of the function

$$f(k,t) := (c\gamma_k)^2 e^{-2a\left(t^b-k^b\right)}. \tag{A.21}$$

To begin with, let

$$g(k,t) := (c\gamma_k)^2 \Phi_{t,k+1}^2 \tag{A.22}$$

so that

$$c^2\phi_t^{ii} = \int_1^t f(k,t)dk + A_t + B_t, \tag{A.23}$$

where

$$A_t := \sum_{k=1}^{t} f(k,t) - \int_1^t f(k,t)dk \quad \text{and} \quad B_t := \sum_{k=1}^{t} \Big[g(k,t) - f(k,t)\Big]. \tag{A.24}$$

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

$$\int_1^t f(k,t)dk = \frac{c\gamma_t}{2} + o(\gamma_t). \tag{A.25}$$

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

$$\lim_{t\to\infty} t^{\eta(s-1)} \int_1^t e^{-rz/b\left(t^b-k^b\right)} \left(z/k^{1-b}\right)^s dk = \frac{z^{s-1}}{r}. \tag{X.0}$$

Before (X.0) is proven, recall the definition of the (upper) incomplete gamma function

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

[8] Note that the conclusion of part (a) holds also for $\eta = 1$. Specifically, $\gamma_{k,t}$ reduces for $\eta = 1$ to $\gamma_{k,t} = \lambda_t(1-c,1)\lambda_k(0,1-c)$ where

$$\lambda_t(\alpha,\beta) := \frac{\Gamma(t+\alpha)}{\Gamma(t+\beta)}. \tag{A.20}$$

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 $x,v>0$*

$$\lim_{t\to\infty} t^{1+v}\Gamma\left(-1/v,-xt^{v}\right)e^{-xt^{v}} = -\,x^{(1-v)/v}e^{-i\pi/v}, \tag{X.1}$$

$$\int e^{-xt^{v}}dt = -\,\frac{\Gamma\left(1/v,xt^{v}\right)}{vx^{1/v}}+C, \tag{X.2}$$

$$\int e^{xt^{v}}dt = -\,\frac{\Gamma\left(1/v,-xt^{v}\right)}{vx^{1/v}}e^{-i\pi/v}+C, \tag{X.3}$$

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

Proof of (X.1)-(X.3). Define $w(t) := \Gamma\left(-1/v,-xt^{v}\right)$ and $q(t) := t^{-1-v}e^{xt^{v}}$ and denote the first derivatives with respect to t by

$$w'(t) = -\frac{ve^{xt^{v}-i\pi/v}}{t^{2}x^{1/v}} \quad \text{and} \quad q'(t) = t^{-v-2}e^{xt^{v}}\left(v(xt^{v}-1)-1\right) \tag{A.27}$$

so that

$$\frac{w'(t)}{q'(t)} = -\frac{ve^{-i\pi/v}}{x^{1/v}\left(vx-(1+v)/t^{b}\right)} \tag{A.28}$$

Then, by L'Hôspital's rule,

$$\lim_{t\to\infty} t^{1+v}\Gamma\left(-1/v,-xt^{v}\right)e^{-xt^{v}} = \lim_{t\to\infty}\frac{w'(t)}{q'(t)} = -x^{(1-v)/v}e^{-i\pi/v}. \tag{A.29}$$

This completes the proof of (X.1). Turning to (X.2), use the u-substitution $u := xt^{v}$ and observe that $t^{1-v} = u^{1/v-1}x^{1-1/v}$. Hence,

$$\int e^{-xt^{v}}dt = \frac{1}{vx^{1/v}}\int e^{-u}u^{1/v-1}du = -\frac{1}{vx^{1/v}}\int_{u}^{\infty} e^{-k}k^{1/v-1}dk + C, \tag{A.30}$$

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. (X.3) follows directly from (X.2) upon using the u-substitution $u := ((-1)x)^{1/v}t$. □

Proof of (X.0). To begin with, note that

$$\int e^{-rz/b\left(t^{b}-k^{b}\right)}\left(z/k^{1-b}\right)^{s}dk = C_0 m(k,t), \tag{A.31}$$

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

$$\int_1^t e^{-rz/b\left(t^{b}-k^{b}\right)}\left(z/k^{1-b}\right)^{s}dk = C_0 m(t,t) + O(e^{-t^{b}}). \tag{A.32}$$

Next, it will be shown that

$$\lim_{t\to\infty} t^{\eta(s-1)}m(t,t) = \frac{C_1}{rz}, \tag{A.33}$$

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

$$\lim_{t\to\infty} t^{\eta(s-1)}m(t,t) = \lim_{t\to\infty}\frac{h_1'(t)}{h_2'(t)} \tag{A.34}$$

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

$$h_1'(t) = C_1 t^{-s\eta}e^{t^b rz/b} \tag{A.35}$$

$$h_2'(t) = rzt^{-\eta(s-1)-\eta}e^{t^b rz/b}(1+o(1)), \tag{A.36}$$

thereby verifying (A.33). Some simple calculations reveal further that $C_0C_1/(rz) = z^{s-1}/r$. □

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

$$\sum_{k=1}^{t} f(k,t) - \int_1^t f(k,t)dk = \frac{1}{2}\Big[f(1,t)+f(t,t)\Big] + \frac{1}{12}\Big[f^{(1)}(t,t) - f^{(1)}(1,t)\Big] + \rho(t;f), \tag{A.37}$$

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

$$|\rho(t;f)| \le \frac{1}{120}\int_1^t |f^{(3)}(k,t)|dk. \tag{A.38}$$

It will be shown that the three terms on the right-hand side of (A.37) are of order $o(\gamma_t)$. Clearly, $f(1,t)+f(t,t) = O(\gamma_t^2)$ while from $f^{(1)}(k,t) = 2(ck^b-\eta)k^{-1}f(k,t)$ one gets $f^{(1)}(t,t)-f^{(1)}(1,t) = o(\gamma_t^2)$. Furthermore,

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

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

(3) *Negligibility of* B_t. Taking account of equation (A.16) and $R_t^\infty = R_k^\infty - R_k^t$, it is seen that

$$\sum_{j=k+1}^{t}\gamma_j = \mathcal{H}_t(\eta) - \mathcal{H}_k(\eta) = (t^b-k^b)/b - \eta R_k^t(\eta). \tag{A.39}$$

Now, using equation (A.19) together with (A.39), it follows that

$$\begin{aligned}
g(k,t) &= (c\gamma_k)^2\, e^{2\sum_{j=k+1}^t \ln(1-c\gamma_j)} \\
&= (c\gamma_k)^2\, e^{-2c(\mathcal{H}_t-\mathcal{H}_k)}e^{-2\sum_{j=k+1}^t K_j} \\
&= (c\gamma_k)^2\, e^{-2c(\mathcal{H}_t-\mathcal{H}_k)} + (c\gamma_k)^2\, e^{-2c(\mathcal{H}_t-\mathcal{H}_k)}\left(e^{-2\sum_{j=k+1}^t K_j}-1\right) \\
&= f(k,t) + f(k,t)\left(e^{2c\eta R_k^t}-1\right) \\
&\qquad + f(k,t)\left(e^{-2\sum_{j=k+1}^t K_j}-1\right) \\
&\qquad + f(k,t)\left(e^{2c\eta R_k^t}-1\right)\left(e^{-2\sum_{j=k+1}^t K_j}-1\right) \\
&=: f(k,t) + B^i(k,t) + B^{ii}(k,t) + B^{iii}(k,t),
\end{aligned} \tag{A.40}$$

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 $k \le t$ and $e^{2c\eta R_k^t} \ge 1$, it follows that $B^i(k,t) \ge 0$. Infer from the discussion surrounding (A.16) that $R_k^t \le \gamma_k$ and thus $B^i(k,t) \le f(k,t)\left(e^{2c\gamma_k} - 1\right)$. Next, suppose without loss of generality that $2c\gamma_k \le \alpha < 1$. Because $e^x < (1-x)^{-1}$ for all $x < 1$, it follows $B^i(k,t) \le 2(1-\alpha)^{-1} f(k,t) c\gamma_k$. By (X.0) and the integral comparison test, $\sum_k B^i(k,t) = O(\gamma_t^2)$.

Next, consider $B^{ii}(k,t)$, which is non-positive so that

$$\left|B^{ii}(k,t)\right| = f(k,t)\left(1 - e^{-2\sum_{j=k+1}^t K_j(\eta)}\right). \tag{A.41}$$

The definition of $K_i(\eta)$ (cf. equation (A.19)) implies that there exists a finite constant $C_0 > 0$ so that $\sum_{j=k+1}^t K_j(\eta) \le 2^{-1} C_0 \sum_{j=k+1}^t j^{-2\eta}$, while $\sum_{j=k+1}^t j^{-2\eta} = \ell(k,t) - 2\eta R_k^t(2\eta)$ with

$$\ell(k,t) = \begin{cases} (1-2\eta)^{-1}\left(t^{1-2\eta} - k^{1-2\eta}\right) & \text{if } \eta < 1/2 \\ (2\eta - 1)^{-1}\left(k^{1-2\eta} - t^{1-2\eta}\right) & \text{if } \eta > 1/2 \\ \ln t - \ln k & \text{if } \eta = 1/2, \end{cases} \tag{A.42}$$

where for $\eta \ge 1/2$ Apostol (1976, theorem 3.2 $(a),(b)$) has been used and the case $\eta < 1/2$ follows immediately from (A.39). Hence, for some finite constant $C_1 > 0$

$$\begin{aligned} 1 - e^{-2\sum_{j=k+1}^t K_j(\eta)} &\le 1 - e^{-C_0 \sum_{j=k+1}^t j^{-2\eta}} \\ &\le 1 - e^{-C_1 \ell(k,t)} e^{C_0 2\eta R_k^t(2\eta)} \\ &\le C_1 \ell(k,t). \end{aligned} \tag{A.43}$$

Since

$$\lim_{t\to\infty} t^{2\eta} \int_1^t f(k,t)\ell(k,t)dk = \frac{1}{4}, \tag{A.44}$$

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

$$\left|1 - e^{-2\sum_{j=k+1}^t K_j(\eta)}\right| \le 1. \tag{A.45}$$

This completes the proof of part (b). □

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

$$\sum_{k=1}^t \gamma_{k,t}^\alpha \gamma_k^\beta = O(\gamma_t^{\alpha+\beta-1}). \tag{X.4}$$

Proof of equation X.4. First, note that with $r = \alpha$ and $s = \alpha + \beta$ it follows from (X.0) that

$$\lim_{t\to\infty} t^{(\alpha+\beta-1)(1-b)} \int_1^t e^{-\alpha z/b\left(t^b - k^b\right)} \left(z/k^{1-b}\right)^{\alpha+\beta} dk = \frac{z^{\alpha+\beta-1}}{r}. \tag{A.46}$$

Next, note that

$$\begin{aligned}\Phi^{\alpha}_{t,k+1} &= e^{\alpha \sum_{i=k+1}^{t} \ln(1-c\gamma_i)} \\ &= e^{-\alpha c(H_t(\eta)-H_k(\eta))} e^{-c\alpha \sum_{i=k+1}^{t} K_i(\eta)} \\ &\leq e^{-\alpha c(H_t(\eta)-H_k(\eta))} \\ &= e^{-\alpha a\left(t^b-k^b\right)} + e^{-\alpha a\left(t^b-k^b\right)} \left(1 - e^{\alpha c \eta R^t_k(\eta)}\right) \end{aligned} \tag{A.47}$$

Taking equation (A.46) into account, the claim follows from Toeplitz's lemma using similar arguments as to show the asymptotic negligibility of $B^i(k,t)$.

A.3.3 Part (c)

To begin with, observe that one gets from part (a)

$$\left(\phi^i_t\right)^2 = \sum_{k,\,s\,=\,1}^{t} \gamma_{k,t}\gamma_{s,t} = \phi^{ii}_t + 2\sum_{k\,=\,2}^{t}\sum_{s\,=\,1}^{k-1} \gamma_{k,t}\gamma_{s,t} \to \frac{1}{c^2}. \tag{A.48}$$

Consequently, $2\phi^{iii}_t = \left(\phi^i_t\right)^2 + 2R_t - \phi^{ii}_t$, where

$$R_t := \phi^{iii}_t - \sum_{k\,=\,2}^{t}\sum_{s\,=\,1}^{k-1} \gamma_{s,k}\gamma_{s,t}. \tag{A.49}$$

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

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

Clearly, $C_t = o(1)$, while $B_t = o(1)$ is a direct consequence of (X.4). 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

$$\begin{aligned}
&\gamma_k/\gamma_j \leq e^{o(1)\sum_{s=k}^{j}\gamma_s} \\
&\Phi_{j,k+1} \leq C_0 e^{-C_1 \sum_{s=k}^{j}\gamma_s} \\
&\sup_{k,\,t} \sum_{j=k}^{t} j^{-v}\, e^{-\lambda \sum_{s=k}^{j} s^{-v}} \leq C_2 < \infty \quad \text{for } \lambda > 0,
\end{aligned} \tag{X.5}$$

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

Proof of (X.5). See, for example Chen (2002, lemma 3.1.1 *&* 3.4.1) or Polyak and Juditsky (1992). □

Now, for any k and j sufficiently large $\Phi_{t,j+1} \geq \Phi_{j,k+1}$, so that

$$|A_t| \leq \sum_{k=1}^{t-2} \gamma_{k,t} \sum_{j=k+1}^{t-1} (\gamma_k - \gamma_j)\Phi_{j,\,k+1}. \tag{A.50}$$

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

$$\begin{aligned}
\sum_{j=k+1}^{t-1} (\gamma_k - \gamma_j)\Phi_{j,\,k+1} &= \sum_{j=k+1}^{t-1} \sum_{i=k+1}^{j} (\gamma_{i-1} - \gamma_i)\Phi_{j,\,k+1} \\
&= \sum_{j=k+1}^{t-1} \sum_{i=k+1}^{j} O\left(\frac{\gamma_i}{i}\right) \Phi_{j,\,k+1}.
\end{aligned} \tag{A.51}$$

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

$$|A_t| \leq C_0 \sum_{k=1}^{t-2} \gamma_{k,t}\gamma_k \sum_{j=k+1}^{t-1} \sum_{i=k+1}^{j} \frac{\Phi_{j,\,k+1}}{i}. \tag{A.52}$$

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

$$\begin{aligned}
\sum_{j=k+1}^{t-1} \sum_{i=k+1}^{j} \frac{\Phi_{j,\,k+1}}{i} &= \sum_{j=k+1}^{t-1} \gamma_j \sum_{i=k+1}^{j} \frac{\gamma_i}{\gamma_j} \frac{\Phi_{j,\,k+1}}{i^b} \\
&\leq \sum_{j=k+1}^{t-1} \gamma_j \sum_{i=k+1}^{j} \frac{\gamma_{k+1}}{\gamma_j} \frac{\Phi_{j,\,k+1}}{i^b} \leq C_1 \sum_{j=k+1}^{t-1} \gamma_j \sum_{i=k+1}^{j} \frac{e^{-C_2/2\sum_{s=k+1}^{j}\gamma_s}}{i^b},
\end{aligned} \tag{A.53}$$

where the first inequality is due to $\gamma_i \leq \gamma_{k+1}$ and the final one uses the first part of (X.5) to approximate γ_{k+1}/γ_j.

Case $\eta < 1/2$. (A.53) can be further bounded by

$$\begin{aligned}\sum_{j=k+1}^{t-1} \gamma_j \sum_{i=k+1}^{j} \frac{e^{-C_2/2\sum_{s=k+1}^{j}\gamma_s}}{i^b} &= \sum_{j=k+1}^{t-1} \sum_{i=k+1}^{j} \left(\frac{j}{i}\right)^{1-2\eta} \gamma_i \frac{e^{-C_2/2\sum_{s=k+1}^{j}\gamma_s}}{j^b} \\ &\leq \frac{4}{C_2}\left(\frac{t}{k}\right)^{1-2\eta} \sum_{j=k+1}^{t-1} \left(e^{C_2/4\sum_{i=k+1}^{j}\gamma_i}\right) \frac{e^{-C_2/2\sum_{s=k+1}^{j}\gamma_s}}{j^b} \\ &= \frac{4}{C_2}\left(\frac{t}{k}\right)^{1-2\eta} \sum_{j=k+1}^{t-1} \frac{e^{-C_2/4\sum_{s=k+1}^{j}\gamma_s}}{j^b} \\ &\leq \frac{4}{C_2}\left(\frac{t}{k}\right)^{1-2\eta} \sum_{j=k+1}^{t-1} \frac{e^{-C_2/4\sum_{s=k+1}^{j}s^{-b}}}{j^b} \leq C_3\left(\frac{t}{k}\right)^{1-2\eta}, \end{aligned} \tag{A.54}$$

where $C_3 \geq$ 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 (X.5). Hence, there exists a finite constant $K > 0$ such that

$$|A_t| \leq K\, t^{1-2\eta} \sum_{k=1}^{t} \frac{\gamma_{k,t}}{k^b} = K\, t^{1-2\eta} \sum_{k=1}^{t} \frac{\Phi_{t,k+1}}{k}. \tag{A.55}$$

By (A.47),

$$\sum_{k=1}^{t} \frac{\Phi_{t,k+1}}{k} \leq \sum_{k=1}^{t} \frac{e^{-ar(t^b-k^b)}}{k} + \sum_{k=1}^{t} \frac{e^{-ar(t^b-k^b)}}{k}\left(e^{cr\eta R_k^t} - 1\right). \tag{A.56}$$

Furthermore,

$$\int_1^t \frac{e^{-a(t^b-k^b)}}{k} dk = e^{-at^b}\left(\frac{\mathrm{Ei}(at^b) - \mathrm{Ei}(a)}{b}\right) = O\left(\frac{1}{t^b}\right), \tag{A.57}$$

where

$$\mathrm{Ei}(x) := -\int_{-x}^{\infty} \frac{e^{-t}}{t} dt \tag{A.58}$$

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

$$\lim_{t\to\infty} t^v e^{-at^v}\mathrm{Ei}(at^v) = 1/a, \tag{A.59}$$

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 $\eta > 1/2$. Note that (A.53) can similarly be bounded by

$$\begin{aligned}\sum_{j=k+1}^{t-1} \gamma_j \sum_{i=k+1}^{j} i^{2\eta-1}\gamma_i e^{-C_2/2\sum_{s=k+1}^{j}\gamma_s} &\leq t^{2\eta-1} \sum_{j=k+1}^{t-1} \gamma_j \sum_{i=k+1}^{j} \gamma_i e^{-C_2/2\sum_{s=k+1}^{j}\gamma_s} \\ &\leq t^{2\eta-1}C_3, \end{aligned} \tag{A.60}$$

where the existence of some finite constant $C_3 \geq 0$ is ensured by the last property of (X.5). Hence,

$$|A_t| \leq \; C_3 \, t^{2\eta-1} \sum_{k=1}^{t} \Phi_{t,k+1}\gamma_k^2 \tag{A.61}$$

and by equation (X.4), $A_t = O(\gamma_t)$.

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

$$\begin{aligned}
\sum_{j=k+1}^{t-1} \sum_{i=k+1}^{j} \frac{\Phi_{j,k+1}}{i} &\leq C_0 \sum_{j=k+1}^{t-1} \sum_{i=k+1}^{j} \frac{e^{-C_1 \sum_{s=k+1}^{j} 1/s^{1/2}}}{i} \\
&\leq C_0 \sum_{j=k+1}^{t-1} \sum_{i=k+1}^{j} \frac{e^{-C_1 \sum_{s=k+1}^{j} 1/s^{1/2+\epsilon}}}{i} \\
&= C_0 \sum_{j=k+1}^{t-1} \frac{1}{j^{1/2+\epsilon}} \sum_{i=k+1}^{j} \left(\frac{j}{i}\right)^{1/2+\epsilon} \frac{e^{-C_1 \sum_{s=k+1}^{j} 1/s^{1/2+\epsilon}}}{i^{1/2-\epsilon}} \\
&\leq C_0 \sum_{j=k+1}^{t-1} \frac{1}{j^{1/2+\epsilon}} \sum_{i=k+1}^{j} \left(\frac{j}{k+1}\right)^{1/2+\epsilon} \frac{e^{-C_1 \sum_{s=k+1}^{j} 1/s^{1/2+\epsilon}}}{i^{1/2-\epsilon}} \\
&\leq C_0 \sum_{j=k+1}^{t-1} \frac{1}{j^{1/2+\epsilon}} \sum_{i=k+1}^{j} \frac{e^{-C_1/2 \sum_{s=k+1}^{j} 1/s^{1/2+\epsilon}}}{i^{1/2-\epsilon}} \\
&= C_0 \sum_{j=k+1}^{t-1} \frac{1}{j^{1/2+\epsilon}} \sum_{i=k+1}^{j} i^{2\epsilon} \frac{e^{-C_1/2 \sum_{s=k+1}^{j} 1/s^{1/2+\epsilon}}}{i^{1/2+\epsilon}} \\
&\leq t^{2\epsilon} \frac{4C_0}{C_1} \sum_{j=k+1}^{t-1} \frac{e^{-C_1/4 \sum_{s=k+1}^{j} 1/s^{1/2+\epsilon}}}{j^{1/2+\epsilon}} \leq t^{2\epsilon} \, C_2,
\end{aligned} \tag{A.62}$$

where the first inequality is due to the first property of (X.5), 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 (X.5) while the final inequality uses the last property of (X.5). Thus, taking equation (X.4) into account, it follows that

$$|A_t| \leq \; C_2 \, t^{2\epsilon} \sum_{k=1}^{t} \gamma_k^2 \Phi_{t,k+1} = O\left(\frac{1}{t^{1/2-2\epsilon}}\right). \tag{A.63}$$

This completes the proof of part (c). □

A.4 Lemma 3

Lemma 3. *Let assumptions 1, 2 and 3A hold and recall that* $\tau = \sigma/\sqrt{\kappa_x^{(2)}}$. *Then for finite and non-zero constants* C_1 *and* C_2,

$$\begin{aligned} \mathrm{E}\left(b_t^{*2}\right) &\leq C_1 \gamma_t \\ \mathrm{E}\left(b_t^{*4}\right) &\leq C_2 \gamma_t^2 \end{aligned} \tag{a}$$

and

$$\lim_{T\to\infty}\frac{1}{T^b}\sum_{t=1}^{T}\mathrm{E}\left(b_t^{*2}\right)=\frac{\tau^2\gamma}{2cb}. \tag{b}$$

Proof of (a). Begin with (a) which is an application of theorem 22 of part 2 in Benveniste et al. (1990). Their theorem provides an L_2−upper bound for Robbins-Monro algorithms (cf. pp. 243 in Benveniste et al. (1990)). In order to apply the theorem, deduce from (A.1) that

$$b_t = b_{t-1} + \gamma_t H(b_{t-1}, \boldsymbol{v}_t), \tag{A.64}$$

where the function $b \mapsto H(b, \boldsymbol{v}_t)$ is defined via

$$H(b, \boldsymbol{v}_t) = h(b)x_t^{*2} + u_t, \tag{A.65}$$

with

$$h(b) := -c(b-\alpha). \tag{A.66}$$

It is readily verified that

$$\begin{aligned}\mathrm{E}\Big[H(b_{t-1}, \boldsymbol{v}_t) - h(b_{t-1})|\mathcal{V}_{t-1}\Big] &= 0\\ \mathrm{E}\Big[H(b_{t-1}, \boldsymbol{v}_t)^2|\mathcal{V}_{t-1}\Big] &= \tau^2 + c^2\frac{\kappa_x^{(4)}}{\kappa_x^{(2)2}}b_{t-1}^{*2},\end{aligned} \tag{A.67}$$

where

$$\mathcal{V}_{t-1} := \sigma(a_0; \boldsymbol{v}_s, s \le t-1) \tag{A.68}$$

so that conditions 1.10.2 and 1.10.4 are satisfied. Furthermore, h in (A.66) satisfies clearly the Lipschitz condition 1.10.5 of Benveniste et al. (1990). It thus suffices to verify that

$$\liminf_{t\to\infty} 2c\frac{\gamma_t}{\gamma_{t+1}} + \frac{\gamma_{t+1}-\gamma_t}{\gamma_{t+1}^2} > 0 \tag{A.69}$$

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

$$\frac{\gamma_{t+1}-\gamma_t}{\gamma_{t+1}^2} = o(1) \tag{A.70}$$

while $\gamma_t/\gamma_{t+1} = 1 + o(1)$, which proves (A.69). Similarly, it follows from Kushner and Yan's (1993) lemma 3.1 that $\mathrm{E}\left(b_t^{*4}\right)/\gamma_t^2$ is bounded. □

Proof of (b). Set $\check{a}_{0t} := a_0^*\Phi_{t,1}$ and recall from equation (A.5) that $b_t^* = \check{a}_{0t} + \check{u}_t + \check{e}_t$. Now, because $\mathrm{E}(u_t e_s) = 0$ for all t and s (cf. lemma 1), it follows that

$$\mathrm{E}\left(b_t^{*2}\right) = \mathrm{E}\left(\check{a}_{0t}^2\right) + \mathrm{E}\left(\check{u}_t^2\right) + \mathrm{E}\left(\check{e}_t^2\right). \tag{A.71}$$

Let us consider the summands in (A.71) one by one: First, equation (A.18) and (X.2) yield

$$\sum_{t=1}^{T} \mathrm{E}\left(\check{a}_{0t}^{2}\right) = \kappa_{a}^{(2)} \sum_{t=1}^{T} \Phi_{t,1}^{2} \leq K \int_{1}^{\infty} e^{-2at^{b}} = O(1), \tag{A.72}$$

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

$$\mathrm{E}\left(\check{u}_t \check{u}_s\right) = \begin{cases} \tau^2 \sum_{k=1}^{t} \gamma_{k,t}^2 & \text{if } s = t \\ \tau^2 \sum_{k=1}^{s} \gamma_{k,t}\gamma_{k,s} & \text{if } s < t \\ \tau^2 \sum_{k=1}^{t} \gamma_{k,t}\gamma_{k,s} & \text{if } s > t. \end{cases} \tag{A.73}$$

Therefore, taking account of lemma 2, it follows that

$$\mathrm{E}\left(\check{u}_t^2\right) = \tau^2 \phi_t^{ii} = \gamma_t \frac{\sigma^2}{2c} + o(\gamma_t). \tag{A.74}$$

Furthermore, (A.10) implies that

$$\mathrm{E}\left(\check{e}_t \check{e}_s\right) = \begin{cases} \sum_{k=1}^{t} \gamma_{k,t}^2 \sigma^2(e_k) & \text{if } s = t \\ \sum_{k=1}^{s} \gamma_{k,t}\gamma_{k,s} \sigma^2(e_k) & \text{if } s < t \\ \sum_{k=1}^{t} \gamma_{k,t}\gamma_{k,s} \sigma^2(e_k) & \text{if } s > t. \end{cases} \tag{A.75}$$

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

$$\mathrm{E}\left(\check{e}_t^2\right) = \sum_{k=1}^{t} \gamma_{k,t}^2 \sigma^2(e_k) \leq C \sum_{k=1}^{t} \gamma_{k,t}^2 \gamma_k = O\left(\gamma_t^2\right), \tag{A.76}$$

where the last equality is due to Toeplitz's lemma and part (b) of lemma X.4. □

A.5 Lemma 4

Assume α to be known and set $\tilde{x}_t = b_{t-1}^{*} x_t$. Henceforth, redefine $b_t = b_t^{*}$ for notational simplicity.

Lemma 4. *Let assumption 1,2 and 3A hold and set*

$$\tilde{\rho}_{t,t+m} := \mathrm{cov}\left(\tilde{x}_t^2, \tilde{x}_{t+m}^2\right).$$

Then

$$\tilde{\rho}_{t,t+m} = \begin{cases} \kappa_x^{(2)} \pi_t & \text{if } m = 1 \\ \kappa_x^{(2)} \pi_t \prod_{j=1}^{m-1} \psi_{t+j} & \text{if } m > 1, \end{cases} \tag{A.77}$$

where

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

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

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

$$\begin{aligned}
\mathrm{E}\left(\tilde{x}_t^2, \tilde{x}_{t+m}^2\right) &= \mathrm{E}\left(x_t^2 b_{t-1}^2 x_{t+m}^2 b_{t+m-1}^2\right) \\
&= \mathrm{E}\Big[b_{t-1}^2 x_t^2 b_{t+m-1}^2 \mathrm{E}\left(x_{t+m}^2 | \mathcal{V}_{t+m-1}\right)\Big] \\
&= \mathrm{E}\left(b_{t-1}^2 x_t^2 b_{t+m-1}^2\right) \mathrm{E}\left(x_{t+m}^2\right) \\
&= \mathrm{E}\left(b_{t-1}^2 x_t^2 b_{t+m-1}^2\right) \kappa_x^{(2)} \\
&= \mathrm{E}\Big[b_{t-1}^2 \mathrm{E}\left(x_t^2 b_{t+m-1}^2 | \mathcal{V}_{t-1}\right)\Big] \kappa_x^{(2)},
\end{aligned} \tag{A.78}$$

so that

$$\tilde{\rho}_{t,t+m} = \kappa_x^{(2)} \Big\{ \mathrm{E}\Big[b_{t-1}^2 \mathrm{E}\left(x_t^2 b_{t+m-1}^2 | \mathcal{V}_{t-1}\right)\Big] - \mathrm{E}\left(b_{t-1}^2\right) \kappa_x^{(2)} \mathrm{E}\left(b_{t+m-1}^2\right) \Big\}. \tag{A.79}$$

Furthermore, equation (A.64) together with (A.65) yields

$$\begin{aligned}
x_t^2 b_{t+m}^2 = x_t^2 \Big[& b_{t+m-1}^2 + \gamma_{t+m}^2 (u_{t+m}^2 + c^2 x_{t+m}^{*4} b_{t+m-1}^2 - 2c x_{t+m}^{*2} b_{t+m-1} u_{t+m}) \\
& + 2\gamma_{t+m} b_{t+m-1} (u_{t+m} - c x_{t+m}^{*2} b_{t+m-1}) \Big].
\end{aligned} \tag{A.80}$$

Now, for $m > 1$

$$\begin{aligned}
\mathrm{E}\left(x_t^2 b_{t+m}^2 | \mathcal{V}_{t-1}\right) &= \mathrm{E}\left(x_t^2 b_{t+m-1}^2 | \mathcal{V}_{t-1}\right) \psi_{t+m} + \gamma_{t+m}^2 \sigma^2 \\
&= \mathrm{E}\left(x_t^2 b_t^2 | \mathcal{V}_{t-1}\right) \prod_{j=1}^{m} \psi_{t+j} + \sigma^2 \left(\sum_{j=1}^{m-1} \gamma_{t+j}^2 \prod_{i=j+1}^{m} \psi_{t+i} + \gamma_{t+m}^2 \right),
\end{aligned} \tag{A.81}$$

where

$$\mathrm{E}\left(x_t^2 b_t^2 | \mathcal{V}_{t-1}\right) = b_{t-1}^2 \tilde{\psi}_t + \gamma_t^2 \sigma^2 \kappa_x^{(4)} / \kappa_x^{(2)2}. \tag{A.82}$$

Similarly, it is readily verified that

$$\mathrm{E}\left(b_{t+m}^2\right) = \mathrm{E}\left(b_t^2\right) \prod_{j=1}^{m} \psi_{t+j} + \tau^2 \left(\sum_{j=1}^{m-1} \gamma_{t+j}^2 \prod_{i=j+1}^{m} \psi_{t+i} + \gamma_{t+m}^2 \right). \tag{A.83}$$

Hence,

$$\mathrm{E}\left[b_{t-1}^2 \mathrm{E}\left(x_t^2 b_{t+m}^2 | \mathcal{V}_{t-1}\right)\right] - \mathrm{E}\left(b_{t-1}^2\right) \kappa_x^{(2)} \mathrm{E}\left(b_{t+m}^2\right) = \pi_t \prod_{j=1}^{m} \psi_{t+j} \tag{A.84}$$

where

$$\begin{aligned}
\pi_t &= \mathrm{E}\left[b_{t-1}^2 \mathrm{E}\left(x_t^2 b_t^2 | \mathcal{V}_{t-1}\right)\right] - \mathrm{E}\left(b_{t-1}^2\right) \kappa_x^{(2)} \mathrm{E}\left(b_t^2\right) \\
&= \mathrm{E}\left(b_{t-1}^4\right) \tilde{\psi}_t - \mathrm{E}\left(b_{t-1}^2\right) \left[\kappa_x^{(2)} \mathrm{E}\left(b_{t-1}^2\right) \psi_t - \sigma^2 \gamma_t^2 \left(\kappa_x^{(4)} / \kappa_x^{(2)2} - 1 \right) \right].
\end{aligned} \tag{A.85}$$

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

B Proof of proposition 2.2

Suppose α is known and set $\tilde{x}_t := x_t b^*_{t-1}$. Now, consider

$$T^{b/2}(\widehat{\beta}_0 - \beta) = \left(\frac{1}{T^b}\sum_{t=1}^{T}\tilde{x}_t^2\right)^{-1}\frac{1}{T^{b/2}}\sum_{t=1}^{T}\tilde{x}_t\varepsilon_t, \tag{A.86}$$

where $\widehat{\beta}_0$ is defined in (26). The proof is based on two steps:

$$(1) \quad \underset{T\to\infty}{\text{plim}}\,\frac{1}{T^b}\sum_{t=1}^{T}\tilde{x}_t^2 = \lim_{T\to\infty}\frac{1}{T^b}\sum_{t=1}^{T}\mathrm{E}\left(\tilde{x}_t^2\right) = \sigma^2\gamma/(2cb)$$

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

B.0.1 Step (1)

In view of lemma 3 (*b*) and the weak LLN, we seek to establish that

$$\text{var}\left(\frac{1}{T^b}\sum_{t=1}^{T}\tilde{x}_t^2\right) = o(1). \tag{A.87}$$

Consider

$$\text{var}\left(\sum_{t=1}^{T}\tilde{x}_t^2\right) = \sum_{t=1}^{T}\text{var}\left(\tilde{x}_t^2\right) + 2\sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\tilde{\rho}_{t,t+s}, \tag{A.88}$$

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

$$\sum_{t=1}^{T}\text{var}\left(\tilde{x}_t^2\right) \le K\mathcal{H}_T(2\eta) = \begin{cases} O\left(T^{2b-1}\right) & \text{if } \eta \neq 1/2 \\ O\left(\ln T\right) & \text{if } \eta = 1/2, \end{cases} \tag{A.89}$$

where $\mathcal{H}_T(\cdot)$ denotes the generalized harmonic number (cf. equation (A.16)). Turning to the second summand on the right-hand side of (A.88), use lemma 4 to write

$$\sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\tilde{\rho}_{t,t+s} = \kappa_x^{(2)}\sum_{t=1}^{T-1}\pi_t + \sum_{t=1}^{T-2}\sum_{s=2}^{T-t}\tilde{\rho}_{t,t+s}. \tag{A.90}$$

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

$$\sum_{t=1}^{T-1}\pi_t \sim \mathcal{H}_T(2\eta) = \begin{cases} O\left(T^{2b-1}\right) & \text{if } \eta \neq 1/2 \\ O\left(\ln T\right) & \text{if } \eta = 1/2. \end{cases} \tag{A.91}$$

Before the second term in (A.90) is investigated, recall from lemma 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

$$\begin{aligned}\psi_t \geq 0 &\Leftrightarrow 2 - c\gamma_t \kappa_x^{(4)} / \kappa_x^{(2)2} \leq \frac{1}{c\gamma_t} \\ &\Leftrightarrow c\gamma_t \left(1 - c\gamma_t \kappa_x^{(4)} / \kappa_x^{(2)2}\right) \leq 1 - c\gamma_t. \end{aligned} \tag{A.92}$$

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.92) 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

$$\begin{aligned}\psi_t \leq \psi_{t+m} &\Leftrightarrow c\gamma_t(2 - c\gamma_t \kappa_x^{(4)} / \kappa_x^{(2)2}) \geq c\gamma_{t+m}(2 - c\gamma_{t+m} \kappa_x^{(4)} / \kappa_x^{(2)2}) \\ &\Leftrightarrow 2\left(c\gamma_t - c\gamma_{t+m}\right) \geq \kappa_x^{(4)} / \kappa_x^{(2)2} \left((c\gamma_t)^2 - (c\gamma_{t+m})^2\right) \\ &\Leftrightarrow 2\left((c\gamma_t)^2 - (c\gamma_{t+m})^2\right)^{-1} \left(c\gamma_t - c\gamma_{t+m}\right) \geq \kappa_x^{(4)} / \kappa_x^{(2)2} \\ &\Leftrightarrow 2\left[(c\gamma_t)^{-1}(1 + \gamma_{t+m}/\gamma_t)^{-1} - \kappa_x^{(4)} / (2\kappa_x^{(2)2})\right] \geq 0. \end{aligned} \tag{A.93}$$

But (A.93) is true because $(c\gamma_t)^{-1} \geq \kappa_x^{(4)} / \kappa_x^{(2)2}$ and $(1 + \gamma_{t+m}/\gamma_t)^{-1} \geq 1/2$. □

Suppose $T > t^* \geq 3$ and decompose the second sum in (A.90) as

$$\sum_{t=1}^{T-2} \sum_{s=2}^{T-t} \tilde{\rho}_{t,t+s} = \kappa_x^{(2)} \sum_{t=1}^{t^*-2} \pi_t \sum_{s=2}^{T-t} \prod_{j=1}^{s-1} \psi_{t+j} + \kappa_x^{(2)} \sum_{t=t^*-1}^{T-2} \pi_t \sum_{s=2}^{T-t} \prod_{j=1}^{s-1} \psi_{t+j} =: A_T + B_T, \tag{A.94}$$

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

$$A_T \leq C_0 \sum_{t=1}^{t^*-2} \gamma_t^2 \sum_{s=2}^{T-t} \prod_{j=1}^{s-1} \psi_{t+j} \leq C_0 \sum_{t=1}^{t^*-2} \gamma_t^2 \sum_{s=2}^{T} \prod_{j=1}^{s-1} \psi_{t+j}, \tag{A.95}$$

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

$$\prod_{j=1}^{s-1} \psi_{t+j} = K(s,t) \prod_{j=1}^{s-1} \psi_j \text{ with } K(s,t) := \frac{\prod_{i=s}^{s+t-1} \psi_i}{\prod_{l=1}^{t} \psi_l}. \tag{A.96}$$

Since $t < t^*$, $K(s,t)$ is uniformly bounded in the second argument. Similarly, the numerator of $K(s,t)$ is clearly finite for $s < t^*$ while for $s \geq t^*$ the product-sequence is convergent by property (a) above. Hence,

$$A_T \leq C_1 \mathcal{H}_{t^*}(2\eta) \sum_{s=1}^{T} \prod_{j=1}^{s} \psi_j \quad \text{with} \quad C_1 := C_0 \max_{\substack{2 \leq s \leq T \\ 1 \leq t \leq t^*-2}} K(s,t). \tag{A.97}$$

Now,

$$\sum_{s=1}^{T} \prod_{j=1}^{s} \psi_j = \sum_{s=1}^{t^*-1} \prod_{j=1}^{s} \psi_j + \prod_{i=1}^{t^*-1} \psi_i \sum_{s=t^*}^{T} \prod_{j=t^*}^{s} \psi_j, \tag{A.98}$$

where, by property (a),

$$\sum_{s=t^*}^{T} \prod_{j=t^*}^{s} \psi_j \leq \sum_{s=t^*}^{T} \Phi_{s,t^*} = O(1) \tag{A.99}$$

using (A.18) and (X.2). Therefore, $A_T < \infty$. Similarly, by lemma 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 $T \to \infty$, it follows that there exists a finite constant C_1 so that

$$B_T \leq C_1 \mathcal{H}_T(2\eta) = \begin{cases} O\left(T^{2b-1}\right) & \text{if } \eta \neq 1/2 \\ O\left(\ln T\right) & \text{if } \eta = 1/2. \end{cases} \tag{A.100}$$

□

B.0.2 Step (2)

The following proof is based on corollary 24.14 which accompanies Central Limit Theorem (CLT) 24.6 in Davidson (1994). The corollary allows for asymptotically (as $t \to \infty$) degenerate variances of the underlying stochastic process. Here, the CLT is applied to scaled partial sums of

$$z_t := \tilde{x}_t \varepsilon_t, \tag{A.101}$$

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

$$Z_{tT} := z_t / s_T \quad \text{with} \quad s_T^2 := \sum_{t=1}^{T} \sigma_t^2. \tag{A.102}$$

Note that Z_{tT} inherits the martingale difference property from z_t and, taking account of equation (b) of lemma 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 $(\mathrm{E}|Z_{tT}/c_{tT}|^r)^{1/r} < \infty$ uniformly for some $r > 2$;
2. $M_T = o(1)$, where $M_T := \max\limits_{1 \le t \le T} c_{tT}$;
3. $\sum_{i=1}^{T} M_{iT}^2 = O(1)$, where $M_{iT} := \max\limits_{\substack{(i-1) \le t \le i \\ i=1,\ldots,T}} c_{tT}$;

then

$$\sum_{t=1}^{T} Z_{tT} \xrightarrow{d} \mathcal{N}(0,1). \tag{A.103}$$

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

$$\sup_{t,T} \left[\mathrm{E}\left(Z_{tT}/c_{tT}\right)^4\right]^{1/4} < \infty. \tag{A.104}$$

But,

$$\mathrm{E}\left(\frac{Z_{tT}}{c_{tT}}\right)^4 \sim \frac{\mathrm{E}\left(z_t^4\right)}{\gamma_t^2} = \kappa_x^{(4)} \kappa_\varepsilon^{(4)} \frac{\mathrm{E}\left(b_{t-1}^4\right)}{\gamma_t^2}, \tag{A.105}$$

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

$$\sum_{t=1}^{T} Z_{tT} = \left(\frac{1}{T^b} \sum_{t=1}^{T} \sigma_t^2\right)^{-1/2} \sum_{t=1}^{T} \tilde{x}_t \varepsilon_t \tag{A.106}$$

and, as stated in equation (b) of lemma 3

$$\lim_{T \to \infty} \frac{1}{T^b} \sum_{t=1}^{T} \sigma_t^2 = \frac{\sigma^4}{2cb}. \quad \square \tag{A.107}$$

C Proof of proposition 2.3

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

$$\begin{aligned} y_t &= \beta b_{t-1}x_t + \delta x_t + \varepsilon_t \\ &= \alpha x_t + \omega\, x_t h(b_{t-1}) + \varepsilon_t, \end{aligned} \tag{A.108}$$

where $\delta = \alpha(1-\beta)$ and the definition of $h(b) = -c(b-\alpha)$ (cf. equation (A.66)) has been used. In order to proceed further, it is helpful to introduce the following notation for suitably scaled partial sums

$$u_{\mathcal{T}} := T^{-1/2}\sum_{t=1}^{T} u_t, \tag{A.109}$$

$$H_{\mathcal{T}} := T^{-1/2}\sum_{t=1}^{T} H(b_{t-1}, v_t), \tag{A.110}$$

$$m_{\mathcal{T}} := T^{-1}\sum_{t=1}^{T} x_t^2. \tag{A.111}$$

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

$$H(b, v_t) = h(b)x_t^{*2} + u_t \text{ with } u_t = x_t^*\varepsilon_t/\sqrt{\kappa_x^{(2)}}; \tag{A.112}$$

cf. equation (A.65). Taking account of (27) and the representation (A.108), it follows that

$$\begin{aligned} T^{1/2}(\widehat{\alpha}-\alpha) &= (\kappa_x^{(2)}/m_{\mathcal{T}})\,T^{-1/2}\sum_{t=1}^{T}\frac{y_t x_t - \alpha x_t^2}{\kappa_x^{(2)}} \\ &= (\kappa_x^{(2)}/m_{\mathcal{T}})\,T^{-1/2}\sum_{t=1}^{T}(\omega\, x_t^{*2}h(b_{t-1}) + u_t) = (\kappa_x^{(2)}/m_{\mathcal{T}})\left[\omega H_{\mathcal{T}} + c^{-1}u_{\mathcal{T}}\right], \end{aligned} \tag{A.113}$$

where the last equality makes use of (A.65). 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 $(\kappa_x^{(2)}/m_{\mathcal{T}}) = 1$ with probability one 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.65) and note that the individual summands of the scaled partial sum $H_{\mathcal{T}}$ can be rewritten as

$$H(a_{t-1}, v_t) = (u_t - cb_{t-1}^*) + e_t, \tag{A.114}$$

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

Negligibility of $H_{\mathcal{T}}$. Using the recursive representation (A.5) of b_t^*, it follows that

$$\begin{aligned} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t - cb_{t-1}^*) &= u_{\mathcal{T}} - \frac{c}{\sqrt{T}}\left(b_0^* + \sum_{t=1}^{T-1} b_t^*\right) \\ &= u_{\mathcal{T}} - c\check{u}_{\mathcal{T}} - c\left\{\check{e}_{\mathcal{T}} - \frac{\check{e}_T + \check{u}_T}{\sqrt{T}} + \frac{b_0^*}{\sqrt{T}}\left[1 + \sum_{t=1}^{T-1}\Phi_{t,1}\right]\right\}, \end{aligned} \tag{A.115}$$

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

$$b_0^* + \sum_{t=1}^{T-1} b_t^* = b_0^*\left(1 + \sum_{t=1}^{T-1}\Phi_{t,1}\right) + \sum_{t=1}^{T}\check{u}_t + \sum_{t=1}^{T}\check{e}_t - (\check{e}_T + \check{u}_T). \tag{A.116}$$

Taking (A.114) into account it thus follows that

$$\begin{aligned} H_{\mathcal{T}} &= (u_{\mathcal{T}} - c\check{u}_{\mathcal{T}}) + (e_{\mathcal{T}} - c\check{e}_{\mathcal{T}}) + \frac{c}{\sqrt{T}}\left\{(\check{e}_T + \check{u}_T) - b_0^*\left[1 + \sum_{t=1}^{T-1}\Phi_{t,1}\right]\right\} \\ &=: H_{\mathcal{T}}^{(u)} + H_{\mathcal{T}}^{(e)} + H_{\mathcal{T}}^{(x)}, \end{aligned} \tag{A.117}$$

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

$$\operatorname{var}\left(H_{\mathcal{T}}^{(u)}\right) = \mathrm{E}\left(u_{\mathcal{T}}^2\right) + c^2\mathrm{E}\left(\check{u}_{\mathcal{T}}^2\right) - 2c\,\mathrm{E}\left(\check{u}_{\mathcal{T}}u_{\mathcal{T}}\right). \tag{A.118}$$

Now, equation (A.73) in together with lemma 2 implies that

$$\mathrm{E}\left(\check{u}_{\mathcal{T}}^2\right) = \frac{1}{T}\left[\sum_{t=1}^{T}\mathrm{E}\left(\check{u}_t^2\right) + 2\sum_{t=2}^{T}\sum_{s=1}^{t-1}\mathrm{E}\left(\check{u}_t\check{u}_s\right)\right] = \tau^2\left(\bar{\phi}^{ii} + 2\bar{\phi}^{iii}\right). \tag{A.119}$$

Next, from

$$\mathrm{E}\left(\check{u}_t u_s\right) = \begin{cases}\tau^2\gamma_{s,t} & \text{if } s \le t\\ 0 & \text{otherwise,}\end{cases} \tag{A.120}$$

it can be seen that

$$\mathrm{E}\left(\check{u}_{\mathcal{T}}u_{\mathcal{T}}\right) = \frac{1}{T}\sum_{s,t=1}^{T}\mathrm{E}\left(\check{u}_t u_s\right) = \frac{\tau^2}{T}\sum_{t=1}^{T}\sum_{s=1}^{t}\gamma_{s,t} = \tau^2\bar{\phi}^{i}. \tag{A.121}$$

Consequently, in conjunction with lemma 2, it follows that

$$\operatorname{var}\left(H_{\mathcal{T}}^{(u)}\right) = \tau^2\left[1 + c^2\left(\bar{\phi}^{ii} + 2\bar{\phi}^{iii}\right) - 2c\bar{\phi}^{i}\right] = o(1) \tag{A.122}$$

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

$$\operatorname{var}\left(H_{\mathcal{T}}^{(e)}\right) = \mathrm{E}\left(e_{\mathcal{T}}^2\right) + c^2\mathrm{E}\left(\check{e}_{\mathcal{T}}^2\right) - 2c\,\mathrm{E}\left(\check{e}_{\mathcal{T}}e_{\mathcal{T}}\right). \tag{A.123}$$

Equation (A.10) and lemma 3 ensure the existence of a finite constant C such that

$$\mathrm{E}\left(e_{\mathcal{T}}^2\right) = \frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left(e_t^2\right) \le \frac{C\mathcal{H}_T(\eta)}{T} = o(1), \tag{A.124}$$

where $\mathcal{H}_T(\eta)$ denotes the generalized harmonic number (cf. equation (A.16)). Furthermore, by equation (A.75)

$$\begin{aligned}
\mathrm{E}\left(\check{e}_{\mathcal{T}}^2\right) &= \frac{1}{T}\left[\sum_{t=1}^{T}\mathrm{E}\left(\check{e}_t^2\right) + 2\sum_{t=2}^{T}\sum_{k=1}^{t-1}\mathrm{E}\left(\check{e}_t\check{e}_k\right)\right] \\
&= \frac{1}{T}\left[\sum_{t=1}^{T}\sum_{k=1}^{t}\gamma_{k,t}\sigma^2(e_k) + 2\sum_{t=2}^{T}\sum_{k=1}^{t-1}\sum_{s=1}^{k}\sigma^2(e_k)\gamma_{s,t}\gamma_{s,k}\right] \\
&= o(1), \qquad \text{(A.125)}
\end{aligned}$$

by lemma 2 and Toeplitz's lemma. Similarly, since

$$\mathrm{E}\left(\check{e}_t e_s\right) = \begin{cases} \sigma(e_s)^2\gamma_{s,t} & \text{if } s \leq t \\ 0 & \text{otherwise,} \end{cases} \qquad \text{(A.126)}$$

it follows from lemma 2 and Toeplitz's that

$$\mathrm{E}\left(\check{e}_{\mathcal{T}} e_{\mathcal{T}}\right) = \frac{1}{T}\sum_{s,t=1}^{T}\mathrm{E}\left(\check{e}_t e_s\right) = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{t}\sigma(e_s)^2\gamma_{s,t} = o(1). \qquad \text{(A.127)}$$

Therefore, $H_T^{(e)} = o_p(1)$. $H_T^{(x)} = o_p(1)$ follows imediately. □

D 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 $w_t = (x_t b_{t-1}, x_t)'$; cf. equation (9). The scaled deviation of the joint OLS estimator (8) in deviation from λ can then be written as

$$T^{b/2}(\widehat{\lambda} - \lambda) = T^b M_T^{-1}\frac{1}{T^{b/2}}\sum_{t=1}^{T} w_t\varepsilon_t = \frac{T^{1+b}}{\det M_T}\left(\frac{Q_T/T}{T^{b/2}}\sum_{t=1}^{T} w_t\varepsilon_t\right), \qquad \text{(A.128)}$$

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

$$Q_T := \begin{bmatrix} \sum_{t=1}^{T} x_t^2 & -\sum_{t=1}^{T} x_t^2 b_{t-1} \\ -\sum_{t=1}^{T} x_t^2 b_{t-1} & \sum_{t=1}^{T}(x_t b_{t-1})^2 \end{bmatrix} \qquad \text{(A.129)}$$

and

$$\det M_T := \sum_{t=1}^{T} x_t^2 \sum_{t=1}^{T}(x_t b_{t-1})^2 - \left(\sum_{t=1}^{T} x_t^2 b_{t-1}\right)^2. \qquad \text{(A.130)}$$

The proof is based on the following three steps

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

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

$$(3) \qquad \frac{Q_T/T}{T^{b/2}} \sum_{t=1}^{T} w_t \varepsilon_t \xrightarrow{d} \mathcal{N}\left(0, \frac{\kappa_x^{(2)2} \gamma \sigma^4}{2cb} \begin{bmatrix} 1 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix}\right).$$

Step (1). Recall the definitions of $\tilde{x}_t = x_t b_{t-1}^*$ and $m_{\mathcal{T}} = T^{-1} \sum_{t=1}^T x_t^2$, where the latter expression equals $\kappa_x^{(2)}$ with probability one by the strong LLN for *i.i.d.* random variables. Now, with some rearrangement, it is seen that

$$\frac{\det M_T}{T^{1+b}} = m_{\mathcal{T}} \left[\frac{1}{T^b} \sum_{t=1}^T \tilde{x}_t^2 - m_{\mathcal{T}}^{-1} \left(\frac{1}{T^{(1+b)/2}} \sum_{t=1}^T x_t^2 b_{t-1}^* \right)^2 \right]. \tag{A.131}$$

The following is repeatedly used: *Let assumption* 1, 2 *and* 3A *hold. Then,*

$$\frac{1}{T^{(1+b)/2}} \sum_{t=1}^T x_t^2 b_{t-1}^* = O_p\left(\frac{1}{T^{b/2}}\right). \tag{A-D.1}$$

Proof of (A-D.1). From equation (A.5) one deduces that

$$\begin{aligned} \frac{1}{T^{(1+b)/2}} \sum_{t=1}^T x_t^2 b_{t-1}^* &= \frac{b_0^*}{T^{(1+b)/2}} \sum_{t=1}^T x_t^2 \Phi_{t-1,1} + \frac{1}{T^{(1+b)/2}} \sum_{t=1}^T x_t^2 \check{u}_{t-1} + \frac{1}{T^{(1+b)/2}} \sum_{t=1}^T x_t^2 \check{e}_{t-1} \\ &=: A_T + B_T + C_T, \end{aligned} \tag{A.132}$$

say. It will be shown that the three terms on the right-hand side of (A.132) are asymptotically negligible. Start by considering A_t. Assumption 3A in conjunction with (A.18) and (X.3) yields by the integral comparison test that $\mathrm{E}|A_T| = O\left(T^{-(1+b)/2}\right)$. Furthermore, by Cauchy-Schwarz's inequality,

$$\|A_T\|_2^2 = \frac{\kappa_b^{(2)}}{T^{1+b}} \sum_{t,s=1}^T \mathrm{E}(x_t^2 x_s^2) \Phi_{t-1,1} \Phi_{s-1,1} \leq \frac{\kappa_b^{(2)} \kappa_x^{(4)}}{T^{1+b}} \left(\sum_{t=1}^T \Phi_{t-1,1} \right)^2 = O\left(\frac{1}{T^{1+b}}\right), \tag{A.133}$$

where the approximation of the majorant side of the above display uses again (A.18) and (X.3). Hence, by the weak LLN, $A_T = O_p(T^{-b})$ (note that $(1+b)/2 > b$). Turning to the second term, observe that

$$B_T = \frac{\kappa_x^{(2)}}{T^{(1+b)/2}} \sum_{t=1}^T \check{u}_{t-1} + \frac{\kappa_x^{(2)}}{T^{(1+b)/2}} \sum_{t=1}^T (x_t^{*2} - 1) \check{u}_{t-1} =: B_T^{(1)} + B_T^{(2)}, \tag{A.134}$$

say, where both expressions on the right-hand side have zero mean. Now, taking account of (A.73), it follows

$$\|B_T^{(1)}\|_2^2 = \frac{\kappa_x^{(2)2}}{T^{1+b}}\sum_{t=1}^{T}\mathrm{E}(\breve{u}_t^2) + \frac{2\kappa_x^{(2)2}}{T^{1+b}}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\mathrm{E}(\breve{u}_t\breve{u}_s)$$
$$= \frac{(\tau\kappa_x^{(2)})^2}{T^{1+b}}\sum_{t=1}^{T}\phi_t^{ii} + \frac{2(\tau\kappa_x^{(2)})^2}{T^{1+b}}\sum_{t=2}^{T}\phi_t^{iii} = O\left(\frac{1}{T}\right) + O\left(\frac{1}{T^b}\right) = O\left(\frac{1}{T^b}\right), \tag{A.135}$$

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

$$\|B_T^{(2)}\|_2^2 = \frac{\kappa_x^{(2)2}}{T^{1+b}}\sum_{t=1}^{T}\mathrm{E}\left((x_t^* - 1)^2\breve{u}_t^2\right) = \frac{\tau^2(\kappa_x^{(4)} - \kappa_x^{(2)2})}{T^{1+b}}\sum_{t=1}^{T}\phi_t^{ii} = O\left(\frac{1}{T}\right). \tag{A.136}$$

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

$$C_T = \frac{\kappa_x^{(2)}}{T^{(1+b)/2}}\sum_{t=1}^{T}\breve{e}_{t-1} + \frac{\kappa_x^{(2)}}{T^{(1+b)/2}}\sum_{t=1}^{T}(x_t^{*2} - 1)\breve{e}_{t-1} =: C_T^{(1)} + C_T^{(2)}, \tag{A.137}$$

and note that both terms on the right-hand side are mean zero by construction. Hence, for some finite constant $C_0 > 0$

$$\|C_T^{(1)}\|_2^2 = \frac{\kappa_x^{(2)2}}{T^{1+b}}\sum_{t=1}^{T}\sum_{k=1}^{t}\gamma_{k,t}^2\sigma^2(e_k) + \frac{2\kappa_x^{(2)2}}{T^{1+b}}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{k=1}^{s}\gamma_{k,t}\gamma_{k,s}\sigma^2(e_k)$$
$$\le C_0\left(\frac{1}{T^{1+b}}\sum_{t=1}^{T}\sum_{k=1}^{t}\gamma_{k,t}^2\gamma_k + \frac{1}{T^{1+b}}\sum_{t=2}^{T}\sum_{s=2}^{t-1}\sum_{k=1}^{s}\gamma_{k,t}\gamma_{k,s}\gamma_k\right) = o\left(\frac{1}{T^b}\right), \tag{A.138}$$

where the first equality uses (A.10), while the inequality is due to equation (a) of lemma 3. The size $o\left(1/T^b\right)$ results from lemma 2 and (X.4). Analogous to $B_T^{(2)}$, it is seen that

$$\|C_T^{(2)}\|_2^2 \le \frac{C_1}{T^{1+b}}\sum_{t=1}^{T}\sum_{k=1}^{t}\gamma_k\gamma_{k,t}^2 = O\left(\frac{1}{T^{1+\eta}}\right), \tag{A.139}$$

for some finite constant $C_1 > 0$. This completes the proof of (A-D.1). □

Now, by step (1) of appendix B,

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

Consequently, taking (A-D.1), (A.131) and the almost sure convergence of m_T into account,

$$\frac{\det M_T}{T^{1+b}} = \frac{\kappa_x^{(2)}}{T^b}\sum_{t=1}^{T}\tilde{x}_t^2 + O_p\left(\frac{1}{T^b}\right) = \frac{\kappa_x^{(2)}\sigma^2\gamma}{2cb} + o_p(1). \tag{A.140}$$

This completes the proof of step (1). □

Step (2). Observe that

$$\frac{1}{T}\sum_{t=1}^{T} x_t^2 b_{t-1} = \alpha m_{\mathcal{T}} + \frac{1}{T}\sum_{t=1}^{T} x_t^2 b_{t-1}^* \tag{A.141}$$

$$\frac{1}{T}\sum_{t=1}^{T} (x_t b_{t-1})^2 = \alpha^2 m_{\mathcal{T}} + \frac{2\alpha}{T}\sum_{t=1}^{T} x_t^2 b_{t-1}^* + \frac{1}{T}\sum_{t=1}^{T} \tilde{x}_t^2 \tag{A.142}$$

implies,

$$Q_T/T = m_{\mathcal{T}} \begin{bmatrix} 1 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix} + \begin{bmatrix} 0 & -\frac{1}{T}\sum_{t=1}^{T} x_t^2 b_{t-1}^* \\ -\frac{1}{T}\sum_{t=1}^{T} x_t^2 b_{t-1}^* & \frac{2\alpha}{T}\sum_{t=1}^{T} x_t^2 b_{t-1}^* + \frac{1}{T}\sum_{t=1}^{T} \tilde{x}_t^2 \end{bmatrix}.$$

In view of the almost sure convergence of $m_{\mathcal{T}}$, 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. □

Step (3). The entries of the 2×1 vector

$$\frac{Q_T/T}{T^{b/2}} \sum_{t=1}^{T} w_t \varepsilon_t \tag{A.143}$$

are respectively given by

$$\frac{1}{T}\sum_{t=1}^{T} x_t^2 \frac{1}{T^{b/2}} \sum_{t=1}^{T} b_{t-1} x_t \varepsilon_t - \frac{1}{T}\sum_{t=1}^{T} x_t^2 b_{t-1} \frac{1}{T^{b/2}} \sum_{t=1}^{T} x_t \varepsilon_t \tag{A.144}$$

$$\frac{1}{T}\sum_{t=1}^{T} (x_t b_{t-1})^2 \frac{1}{T^{b/2}} \sum_{t=1}^{T} x_t \varepsilon_t - \frac{1}{T}\sum_{t=1}^{T} x_t^2 b_{t-1} \frac{1}{T^{b/2}} \sum_{t=1}^{T} b_{t-1} x_t \varepsilon_t. \tag{A.145}$$

Note that (A.144) and (A.145) can be rewritten as

$$\frac{m_{\mathcal{T}}}{T^{b/2}} \sum_{t=1}^{T} \tilde{x}_t \varepsilon_t - u_{\mathcal{T}} \left(\frac{\kappa_x^{(2)}}{T^{(1+b)/2}} \sum_{t=1}^{T} x_t^2 b_{t-1}^* \right) \tag{A.146}$$

$$-\frac{\alpha m_{\mathcal{T}}}{T^{b/2}} \sum_{t=1}^{T} \tilde{x}_t \varepsilon_t + u_{\mathcal{T}} \left(\frac{\alpha \kappa_x^{(2)}}{T^{(1+b)/2}} \sum_{t=1}^{T} x_t^2 b_{t-1}^* + \frac{\kappa_x^{(2)}}{T^{(1+b)/2}} \sum_{t=1}^{T} \tilde{x}_t^2 \right) - \frac{1}{T}\sum_{t=1}^{T} x_t^2 b_{t-1}^* \frac{1}{T^{b/2}} \sum_{t=1}^{T} \tilde{x}_t \varepsilon_t \tag{A.147}$$

so that

$$\frac{Q_T/T}{T^{b/2}}\sum_{t=1}^{T} w_t\varepsilon_t = \begin{bmatrix}1\\-\alpha\end{bmatrix}\frac{m_{\mathcal{T}}}{T^{b/2}}\sum_{t=1}^{T}\tilde{x}_t\varepsilon_t$$
$$-\begin{bmatrix}1\\-\alpha\end{bmatrix}\left(\frac{\kappa_x^{(2)}}{T^{(1+b)/2}}\sum_{t=1}^{T}x_t^2 b_{t-1}^*\right)u_{\mathcal{T}}$$
$$+\begin{bmatrix}0\\1\end{bmatrix}\left\{\left(\frac{\kappa_x^{(2)}}{T^{(1+b)/2}}\sum_{t=1}^{T}\tilde{x}_t^2\right)u_{\mathcal{T}}-\left(\frac{1}{T}\sum_{t=1}^{T}x_t^2 b_{t-1}^*\right)\frac{1}{T^{b/2}}\sum_{t=1}^{T}\tilde{x}_t\varepsilon_t\right\}$$
$$=\begin{bmatrix}1\\-\alpha\end{bmatrix}\frac{m_{\mathcal{T}}}{T^{b/2}}\sum_{t=1}^{T}\tilde{x}_t\varepsilon_t + O_p\left(\frac{1}{T^{b/2}}\right)+O_p\left(\frac{1}{T^{\eta/2}}\right)+O_p\left(\frac{1}{T^{1/2}}\right), \tag{A.148}$$

where again (A-D.1) and step (1) of appendix B together with $u_{\mathcal{T}} = T^{-1/2}\sum_{t=1}^{T}u_t = O_p(1)$ (cf. (A.109)) has been used in order to obtain the size of the remainder terms. Hence, it follows that (A.143) is asymptotically equal to

$$\begin{bmatrix}1\\-\alpha\end{bmatrix}\frac{\kappa_x^{(2)}}{T^{b/2}}\sum_{t=1}^{T}\tilde{x}_t\varepsilon_t \xrightarrow{d} \mathcal{N}\left(0,\frac{\kappa_x^{(2)2}\gamma\sigma^4}{2cb}\begin{bmatrix}1 & -\alpha\\ -\alpha & \alpha^2\end{bmatrix}\right). \tag{A.149}$$

The limiting distribution is a direct consequence of step (2) of appendix B. Using step (1) of this proof in conjunction with Slutzky's theorem gives the stated result. □

D.0.3 Remarks

Recall

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

and observe that

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

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 be recognizing that

$$T^b m^{11}\hat{\sigma}^2 = \frac{m_{\mathcal{T}}\hat{\sigma}^2}{\det M_T/T^{1+b}} = \frac{2cb}{\gamma} + o_p(1) \tag{A.150}$$

$$T^b m^{22}\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}(x_t b_{t-1})^2\frac{\hat{\sigma}^2}{\det M_T/T^{1+b}} = \alpha^2\frac{2cb}{\gamma} + o_p(1). \tag{A.151}$$

E Proof of corollary 2.1

Begin with the proof of (17) and note that the definition in equation (16) of $\widehat{\lambda}_\alpha$ yields

$$\widehat{\lambda}_\alpha - \alpha = \frac{\widehat{\lambda}_\delta - \delta + \alpha(\widehat{\lambda}_\beta - \beta)}{1 - \widehat{\lambda}_\beta}. \tag{A.152}$$

Now, taking (A.129) and (A.130) into account, it is seen that

$$\widehat{\lambda}_\beta - \beta = \frac{1}{\det W_T}\left(\sum_{t=1}^T x_t^2 \sum_{t=1}^T b_{t-1}x_t\varepsilon_t - \sum_{t=1}^T x_t^2 b_{t-1}\sum_{t=1}^T x_t\varepsilon_t\right) \tag{A.153}$$

$$\widehat{\lambda}_\delta - \delta = \frac{1}{\det W_T}\left(\sum_{t=1}^T (x_t b_{t-1})^2 \sum_{t=1}^T x_t\varepsilon_t - \sum_{t=1}^T x_t^2 b_{t-1}\sum_{t=1}^T b_{t-1}x_t\varepsilon_t\right). \tag{A.154}$$

Consequently, one gets with a little rearrangement,

$$\begin{aligned}
T^{1/2}\left(\widehat{\lambda}_\alpha - \alpha\right) &= \frac{(\det W_T/T^{1+b})^{-1}}{1-\widehat{\lambda}_\beta}\frac{1}{T^{1/2}}\sum_{t=1}^T b_{t-1}x_t\varepsilon_t\left(\frac{\alpha}{T^b}\sum_{t=1}^T x_t^2 - \frac{1}{T^b}\sum_{t=1}^T x_t^2 b_{t-1}\right)\\
&\quad + \frac{(\det W_T/T^{1+b})^{-1}}{1-\widehat{\lambda}_\beta}\frac{1}{T^{1/2}}\sum_{t=1}^T x_t\varepsilon_t\left(\frac{1}{T^b}\sum_{t=1}^T (x_t b_{t-1})^2 - \frac{\alpha}{T^b}\sum_{t=1}^T x_t^2 b_{t-1}\right)\\
&= \frac{u_{\mathcal{T}}}{1-\widehat{\lambda}_\beta}\left(\frac{\kappa_x^{(2)}(\det W_T/T^{1+b})^{-1}}{T^b}\sum_{t=1}^T \tilde{x}_t^2\right)\\
&\quad - \frac{(\det W_T/T^{1+b})^{-1}}{1-\widehat{\lambda}_\beta}\left(\frac{1}{T^{(1+b)/2}}\sum_{t=1}^T x_t^2 b_{t-1}^*\right)\left(\frac{1}{T^{b/2}}\sum_{t=1}^T \tilde{x}_t\varepsilon_t\right).
\end{aligned} \tag{A.155}$$

By the CLT for *i.i.d.* random variables, $u_{\mathcal{T}}$ is asymptotically normal with mean zero and variance τ^2; see also the discussion following the definition of $u_{\mathcal{T}}$ in (A.109). Taking step (1) of appendix B and D together with the consistency of $\widehat{\lambda}_\beta$ into account, one deduces from Slutzky's theorem and the aforementioned CLT that the first expression after the second equality is asymptotically normal with mean zero and variance $(\tau/c)^2$. The the second summand after the second equality is $o_p(1)$ due to (A-D.1) and the second step of appendix B.

The limiting distribution of $T^{1/2}(\iota_\alpha'\widehat{\lambda} - \alpha)$ is now easily established once one notes that

$$\begin{aligned}
T^{1/2}(\iota_\alpha'\widehat{\lambda} - \alpha) &= u_{\mathcal{T}}\left(\frac{\kappa_x^{(2)}(\det W_T/T^{1+b})^{-1}}{T^b}\sum_{t=1}^T \tilde{x}_t^2\right)\\
&\quad - (\det W_T/T^{1+b})^{-1}\left(\frac{1}{T^{(1+b)/2}}\sum_{t=1}^T x_t^2 b_{t-1}^*\right)\left(\frac{1}{T^{b/2}}\sum_{t=1}^T \tilde{x}_t\varepsilon_t\right). \quad \square
\end{aligned} \tag{A.156}$$

F The general algorithm

This section extends the conclusions of the previous analysis to the general algorithm (5) without restricting r_t provided the conditions of assumption 1, 2 and 3 are met. For future reference, define the L_p−norm $\|X\|_p := (\mathrm{E}|X|^p)^{1/p}$ for any random variable X with $\mathrm{E}|X|^p < \infty$ and real $p > 0$. To begin with, a useful result on the recursion r_t of the learning scheme (5) is established:

Lemma F.1. *Suppose assumption 1 and 2 hold. If, in addition, $\kappa_x^{(2s)} < \infty$, then*

$$\|r_t - \kappa_x^{(2)}\|_s = O(\gamma_t^{1-1/s}) \text{ for } s > 0.$$

Proof. In view of (A.6) and (A.7) define $\tilde{\Phi}_{t,1} = \prod_{i=k}^{t}(1-\gamma_i)$ with $\tilde{\Phi}_{t,t+1} := 1$ and $\tilde{\gamma}_{k,t} := \tilde{\Phi}_{t,k+1}\gamma_k$. Then,

$$r_t^* - 1 = (r_0^* - 1)\tilde{\Phi}_{t,1} + \sum_{k=1}^{t} \tilde{\gamma}_{k,t}\left(x_k^{*2} - 1\right), \tag{A.157}$$

where it has been used that $\sum_{k=1}^{t} \tilde{\gamma}_{k,t} + \tilde{\Phi}_{t,1} = 1$. For simplicity, assume $\gamma = 1$, then $\tilde{\Phi}_{t,1} = 0$ and, by construction, $\mathrm{E}(r_t^* - 1) = 0$. The independence of $(x_t)_{t\geq 1}$ implies

$$\|r_t^* - 1\|_s^s = \sum_{k_1,\ldots,k_s=1}^{t} \tilde{\gamma}_{k_1,t}\cdots\tilde{\gamma}_{k_s,t}\left\|\left(x_{k_1}^{*2}-1\right)\cdots\left(x_{k_s}^{*2}-1\right)\right\| = \left\|x_t^{*2}-1\right\|_s^s \sum_{k=1}^{t}\left(\tilde{\gamma}_{k,t}\right)^s, \tag{A.158}$$

where, by the c_r inequality, $\left\|x_t^{*2} - 1\right\|_s^s \leq 2^{s-1}(1+\kappa_x^{(2s)}) < \infty$. The claim follows by equation (X.4). □

Remark. *For $r = 2$, convergence in mean square and thus convergence in probability follows. Because r_t is a partial sum of the independent sequence $r_0, x_1, \ldots, x_t$, convergence is almost surely, see e.g. Davidson (1994, theorem 20.9). In view of the almost sure convergence of r_t, there exists a time t_0 (random) after which the recursions b_t and a_t coincide with probability one.*

First, recall for convenience from (A.1) and (A.2) the simplified algorithm which results upon imposing assumption 3A, i.e.

$$\begin{aligned} b_t^* &= b_{t-1}^* + \gamma_t H(b_{t-1}, v_t) \\ &= b_{t-1}^*(1 - c\gamma_t) + \gamma_t u_t + \gamma_t e_t \\ &= a_0^*\Phi_{t,1} + \check{u}_t + \check{e}_t. \end{aligned} \tag{A.159}$$

where definition of the map $b \mapsto H(b, v_t)$ is given by (A.65); $e_t = c(1 - x_t^{*2})b_{t-1}^*$, $b_t^* = b_t - \alpha$ and $b_0 = a_0$. On the other hand, the general algorithm is seen to be given by

$$\begin{aligned} a_t^* &= a_{t-1}^* + \gamma_t H(a_{t-1}, v_t) + \gamma_t(1/r_t^* - 1)H(a_{t-1}, v_t) \\ &= a_{t-1}^*(1 - c\gamma_t) + \gamma_t u_t + \gamma_t\xi_t^{(1)} + \gamma_t\xi_t^{(2)} \\ &= a_0^*\Phi_{t,1} + \check{u}_t + \check{\xi}_t^{(1)} + \check{\xi}_t^{(2)}, \end{aligned} \tag{A.160}$$

where

$$\xi_t^{(1)} := c(1 - x_t^{*2})a_{t-1}^* \tag{A.161}$$

$$\xi_t^{(2)} := (1/r_t^* - 1)H(a_{t-1}, v_t). \tag{A.162}$$

The strengthening of the conditions of assumption 3 relative to assumption 3A is due to the need of bounding the fourth moment of $\xi_t^{(2)}$. As shown by lemma F.2, the r^{th} moment of $\xi_t^{(2)}$ inherits the properties of the $2r^{th}$ moment of $r_t - \kappa_x^{(2)}$ under assumption 3.

Lemma F.2. *Let assumption 1, 2 and 3 hold, then*

$$\left\|\xi_t^{(2)}\right\|_r = O(\gamma_t^{1-1/(2r)}) \quad for \quad r \in (0, 4].$$

Proof. For notational ease, write $H_t := H(a_{t-1}, v_t)$. Developing $f(x) := 1/x - 1$ in a mean-value expansion around 1 yields

$$\xi_t^{(2)} = f(r_t^*)H_t = -(r_t^* - 1)H_t\lambda_t^{-2}, \tag{A.163}$$

where the mean-value λ_t (random) is on the line segment connecting r_t^* and 1. Cauchy-Schwarz's inequality yields furthermore

$$\left\|\xi_t^{(2)}\right\|_r \leq \|r_t^* - 1\|_{2r} \left\|H_t\lambda_t^{-2}\right\|_{2r}. \tag{A.164}$$

By part (a) of assumption 3, λ_t is bounded a.s., while part (b) ensures that $\|H_t\|_{2r} < \infty$, thereby showing that the previous display is of size $O(\|r_t^* - 1\|_{2r})$. Taking account of lemma F.1 thus completes the proof. □

Remark. *From Minkowski's inequality one gets*

$$\left\|\check{\xi}_t^{(2)}\right\|_r \leq \sum_{k=1}^{t} \gamma_{k,t} \left\|\xi_k^{(2)}\right\|_r. \tag{A.165}$$

Hence, it follows as a corollary (using the previous lemma and equation (X.4)) that

$$\left\|\check{\xi}_t^{(2)}\right\|_r = O(\gamma_t^{1-1/(2r)}).$$

The key result of this section is summarized by the following proposition.

Proposition F.1. *Let assumption 1, 2 and 3 hold, then*

$$(1) \qquad \frac{1}{T^{b/2}} \sum_{t=1}^{T} x_t\varepsilon_t \left(a_{t-1}^* - b_{t-1}^*\right) = o_p(1)$$

$$(2) \qquad \frac{1}{T^{b}} \sum_{t=1}^{T} \left(a_{t-1}^{*2} - b_{t-1}^{*2}\right) = o_p(1).$$

Remark. *In order to see how this proposition can be applied, consider the proof of proposition 2.2 given in appendix B. Let* $\tilde{x}_t := x_t a^*_{t-1}$ *and note that*

$$\frac{1}{T^{b/2}}\sum_{t=1}^{T}\tilde{x}_t\varepsilon_t = \frac{1}{T^{b/2}}\sum_{t=1}^{T}x_t b^*_{t-1}\varepsilon_t + \frac{1}{T^{b/2}}\sum_{t=1}^{T}x_t\varepsilon_t(a^*_{t-1}-b^*_{t-1})$$
$$\frac{1}{T^{b}}\sum_{t=1}^{T}\tilde{x}_t^2 = \frac{1}{T^{b}}\sum_{t=1}^{T}(x_t b^*_{t-1})^2 + \frac{1}{T^{b}}\sum_{t=1}^{T}x_t^2(a^{*2}_{t-1}-b^{*2}_{t-1}).$$

Since the leading terms of the above display are treated in appendix B and the remainder terms have been shown to be $o_p(1)$, *the conclusion of proposition 2.2 remains to hold under the general algorithm* (5) *without constraining* r_t. *Similar arguments can be applied to the results contained in appendices C, D and E in order to show that the corresponding results extend to the general algorithm* (5) *given assumptions 1, 2 and 3 are met.*

Proof of part (1). Note that the summands

$$z_t := x_t\varepsilon_t\left(a^*_{t-1}-b^*_{t-1}\right)$$

form a martingale difference sequence with respect to $\mathcal{V}_t$ (cf. equation (A.68)) so that

$$\left\|\frac{1}{T^{b/2}}\sum_{t=1}^{T}z_t\right\|_2^2 = \frac{1}{T^b}\sum_{t=1}^{T}\|z_t\|_2^2. \tag{A.166}$$

It will be shown that $\|z_t\|_2^2 = O(\gamma_t^{3/2})$. First, note that the definitions of a_t and b_t imply that

$$\|z_t\|_2^2 = \kappa_x^{(2)}\sigma^2\left\|a^*_{t-1}-b^*_{t-1}\right\|_2^2 = \kappa_x^{(2)}\sigma^2\left\|\check{\xi}_t^{(1)}-\check{e}_t+\check{\xi}_t^{(2)}\right\|_2^2 \le 2\kappa_x^{(2)}\sigma^2\Big(\left\|\check{\xi}_t^{(1)}-\check{e}_t\right\|_2^2+\left\|\check{\xi}_t^{(2)}\right\|_2^2\Big).$$

Now,

$$\begin{aligned}
\left\|\check{\xi}_t^{(1)}-\check{e}_t\right\|_2^2 &= c^2\left\|\sum_{k=1}^{t}\gamma_{k,t}\left(1-x_k^{*2}\right)\left(a^*_{k-1}-b^*_{k-1}\right)\right\|_2^2\\
&= c^2\left(\kappa_x^{(4)}/\kappa_x^{(2)2}-1\right)\sum_{k=1}^{t}\gamma_{k,t}^2\left\|a^*_{k-1}-b^*_{k-1}\right\|_2^2\\
&\le c^2\left(\kappa_x^{(4)}/\kappa_x^{(2)2}-1\right)\sum_{k=1}^{t}\gamma_{k,t}^2\left(\left\|a^*_{k-1}\right\|_2^2+\left\|b^*_{k-1}\right\|_2^2\right)\\
&\le 2\,C_0\,c^2\left(\kappa_x^{(4)}/\kappa_x^{(2)2}-1\right)\sum_{k=1}^{t}\gamma_{k,t}^2\gamma_k = O(\gamma_t^2),
\end{aligned} \tag{A.167}$$

where the first equality follows from assumption 2, the final inequality uses that $\mathrm{E}\left(a_t^{*2}\right) = O(\gamma_t)$ (cf. Benveniste et al. (1990, theorem 24)) and the last equality is due to equation (X.4). By the remark surrounding lemma F.2, $\left\|\check{\xi}_t^{(2)}\right\|_2 = O(\gamma_t^{3/2})$ which proves $\|z_t\|_2^2 = O(\gamma_t^{3/2})$. Hence, $\sum_{t=1}^{T} \|z_t\|_2^2 = O\left(T^{1-3/2\eta}\right)$ so that

$$\frac{1}{T^{b/2}} \sum_{t=1}^{T} z_t = O_p\left(T^{-\eta/4}\right), \tag{A.168}$$

thereby completing the proof of part (1). □

Proof of part (2). Use the definitions of a_t and b_t to write

$$\begin{aligned} a_t^{*2} - b_t^{*2} &= \left(\check{\xi}_t^{(1)2} - \check{e}_t^2\right) + 2\left(\check{\xi}_t^{(1)} - \check{e}_t\right)\left(b_t^* - \check{e}_t\right) + \check{\xi}_t^{(2)}\left(2a_t - \check{\xi}_t^{(2)}\right) \\ &=: A_t^{(1)} + 2A_t^{(2)} + A_t^{(3)}, \end{aligned} \tag{A.169}$$

say. We seek to establish

$$\frac{1}{T^b} \sum_{t=1}^{T} A_t^{(\ell)} = o_p(1) \tag{A.170}$$

by first showing that $\mathrm{E}|T^{-b} \sum_{t=1}^{T} A_t^{(\ell)}| = o(1)$ for $\ell = 1, 2, 3$. If, in addition, $\left\|\sum_{t=1}^{T} A_t^{(\ell)}\right\|_2^2 = o(T^{2b})$, the claim follows from the weak LLN. Making use of Minkowski's inequality,

$$\left\|\sum_{t=1}^{T} A_t^{(\ell)}\right\|_2 \leq \sum_{t=1}^{T} \left\|A_t^{(\ell)}\right\|_2, \tag{A.171}$$

it is seen that a sufficient condition for (A.170) is that both, $\mathrm{E}|A_t^{(\ell)}|$ and $\left\|A_t^{(\ell)}\right\|_2$, are of size $O(\gamma_t^{\alpha})$ for some $\alpha > 1$ and $\ell = 1, 2, 3$.

Begin with $A_t^{(1)}$ and note that by assumption 2,

$$\begin{aligned} \mathrm{E}|A_t^{(1)}| &= c^2 \sum_{s,k=1}^{t} \gamma_{k,t}\gamma_{s,t} \mathrm{E}|(1 - x_k^{*2})(1 - x_s^{*2})(a_{k-1}^* a_{s-1}^* - b_{k-1}^* b_{s-1}^*)| \\ &= c^2 \left(\kappa_x^{(4)}/\kappa_x^{(2)2} - 1\right) \sum_{k=1}^{t} \gamma_{k,t}^2 \left(\left\|a_{k-1}^*\right\|_2^2 - \left\|b_{k-1}^*\right\|_2^2\right). \end{aligned} \tag{A.172}$$

By Benveniste et al. (1990, theorem 24) and equation (X.4), the final display is seen to be of size $O(\gamma_t^2)$. Furthermore,

$$\begin{aligned} \left\|A_t^{(1)}\right\|_2^2 &= c^4 \sum_{k_1,\ldots,k_4=1}^{t} \tilde{\gamma}_{k_1,t} \cdots \tilde{\gamma}_{k_4,t} \Big\| \left(x_{k_1}^{*2} - 1\right) \cdots \left(x_{k_4}^{*2} - 1\right) \\ &\qquad \times \left(a_{k_1-1}^* a_{k_2-1}^* - b_{k_1-1}^* b_{k_2-1}^*\right)\left(a_{k_3-1}^* a_{k_4-1}^* - b_{k_3-1}^* b_{k_4-1}^*\right) \Big\| \\ &= c^4 \left\|x_t^{*2} - 1\right\|_4^4 \sum_{k=1}^{t} \gamma_{k,t}^4 \left\|a_{k-1}^{*2} - b_{k-1}^{*2}\right\|_2^2 \\ &\leq (2c)^4 \left(1 + \kappa_x^{(8)}/\kappa_x^{(2)4}\right) \sum_{k=1}^{t} \gamma_{k,t}^4 \left(\left\|a_{k-1}^*\right\|_4^4 + \left\|b_{k-1}^*\right\|_4^4\right), \end{aligned}$$

where the second equality uses assumption 2 while the inequality is due to the c_r inequality. By part (a) of assumption 3, $\left\|a^*_{k-1}\right\|_4^4 + \left\|b^*_{k-1}\right\|_4^4 < \infty$ so that in conjunction with equation (X.4) it follows that $\left\|A_t^{(1)}\right\|_2^2 = O(\gamma_t^3)$.

Turning to $A_t^{(2)}$, observe first that Cauchy-Schwarz's together with the triangle inequality yield

$$\mathrm{E}|A_t^{(2)}| \le \left\|\check{\xi}_t^{(1)} - \check{e}_t\right\|_2 \|b_t^* - \check{e}_t\|_2 \le \left\|\check{\xi}_t^{(1)} - \check{e}_t\right\|_2 (\|\check{u}_t\|_2 + \|\check{a}_{0t}\|_2)\,. \tag{A.173}$$

Because $\|\check{u}_t\|_2^2 = \tau^2 \sum_{k=1}^t \gamma_{k,t}^2 = O(\gamma_t)$ and $\|\check{a}_{0t}\|_2^2 = O(e^{-t^b})$ (cf. equation (**??**)), it follows from equation (A.167) that $\mathrm{E}|A_t^{(2)}| = O(\gamma_t^{3/2})$. Again, by Cauchy-Schwarz's inequality, $\left\|A_t^{(2)}\right\|_2 \le \left\|\check{\xi}_t^{(1)} - \check{e}_t\right\|_4 \|b_t^* - \check{e}_t\|_4$. Now,

$$\begin{aligned}
\left\|\check{\xi}_t^{(1)} - \check{e}_t\right\|_4^4 &= c^4 \sum_{k_1,\dots,k_4=1}^{t} \tilde{\gamma}_{k_1,t}\cdots\tilde{\gamma}_{k_4,t} \Big\| \left(1 - x_{k_1}^{*2}\right)\cdots\left(1 - x_{k_r}^{*2}\right) \\
&\qquad\qquad \times \left(a^*_{k_1-1} - b^*_{k_1-1}\right)\cdots\left(a^*_{k_4-1} - b^*_{k_4-1}\right)\Big\| \\
&= c^4 \left\|1 - x_t^{*2}\right\|_4^4 \sum_{k=1}^t \gamma_{k,t}^4 \left\|a^*_{k-1} - b^*_{k-1}\right\|_4^4 \\
&\le (2c)^4 (1 + \kappa_x^{(8)}/\kappa_x^{(2)4}) \sum_{k=1}^t \gamma_{k,t}^4 \left(\left\|a^*_{k-1}\right\|_4^4 + \left\|b^*_{k-1}\right\|_4^4\right)
\end{aligned} \tag{A.174}$$

using assumption 2 together with the c_r inequality. Because, by part (a) of assumption 3, $\left\|a^*_{k-1}\right\|_4^4 + \left\|b^*_{k-1}\right\|_4^4 < \infty$ the previous display is seen to be of size $O(\gamma_t^3)$. The claim follows because $\|b_t^* - \check{e}_t\|_4 \le \|\check{a}_{0t}\|_4 + \|\check{u}_t\|_4 = O(\gamma_t^{3/4})$.

Finally, consider $A_t^{(3)}$. By Cauchy-Schwarz's inequality, the definition of a_t and b_t together with lemma F.2, one gets

$$\mathrm{E}|A_t^{(3)}| \le \left\|\check{\xi}_t^{(2)}\right\|_2 \left(2\|a_t\|_2 + \left\|\check{\xi}_t^{(2)}\right\|_2\right) = O(\gamma_t^{3/4}) \tag{A.175}$$

and

$$\left\|A_t^{(3)}\right\|_2 \le \left\|\check{\xi}_t^{(2)}\right\|_4 \left(2\|a_t\|_4 + \left\|\check{\xi}_t^{(2)}\right\|_4\right) = O(\gamma_t^{7/8}). \tag{A.176}$$

This completes the proof of part (2). □

G Proof of remark 2.2

Throughout assumption 2 and 3A are assumed to hold. Further, suppose assumption 1 holds with $\eta = 0$, i.e. $\gamma_t = \gamma$ for all t.

A recursive representation for the simplified recursion (A.1) similar to (A.2) is then readily derived, i.e.

$$b_t^* = (1-\theta)^t b_0^* + \gamma \sum_{i=0}^{t-1} (1-\theta)^i e_{t-i} + \gamma \sum_{i=0}^{t-1} (1-\theta)^i u_{t-i}, \tag{A.177}$$

where $\theta := c\gamma$ while e_t and u_t have been defined in (A.3). Let

$$\tilde{u}_t := \sum_{i=0}^{t-1} (1-\theta)^i u_{t-i} \tag{A.178}$$

$$\tilde{e}_t := \sum_{i=0}^{t-1} (1-\theta)^i e_{t-i}, \tag{A.179}$$

so that (A.177) can be written as

$$b_t^* = (1-\theta)^t b_0^* + \gamma \tilde{u}_t + \gamma \tilde{e}_t. \tag{A.180}$$

Lemma G.1. *If* $\eta = 0$, *then for finite and non-zero constants* C_1 *and* C_2,

$$\begin{aligned} \mathrm{E}\left(b_t^{*2}\right) &\leq C_1 \gamma \\ \mathrm{E}\left(b_t^{*4}\right) &\leq C_2 \gamma^2. \end{aligned} \tag{G.1a}$$

Set $\nu := \lim_{t\to\infty} \mathrm{E}\left(b_t^{*2}\right)$, *then*

$$\nu = \frac{\gamma\tau^2}{2c} + O\left(\gamma^2\right). \tag{G.1b}$$

Proof of lemma G.1. (G.1a) follows from part (a) of lemma 3. Now, by lemma 1 and equation (A.177),

$$\mathrm{E}\left(b_t^{*2}\right) = (1-\theta)^{2t} \kappa_b^{(2)} + \theta^2 \left(\kappa_x^{(4)}/\kappa_x^{(2)2} - 1\right) \sum_{i=0}^{t-1} (1-\theta)^{2i} \mathrm{E}\left(b_{t-1-i}^{*2}\right) + (\gamma\tau)^2 \sum_{i=0}^{t-1} (1-\theta)^{2i}. \tag{A.181}$$

The three terms on the right-hand side of (A.181) will be subsequently investigated: The first term is asymptotically negligible as $\lim_{t\to\infty} (1-\theta)^{2t} = 0$ and $\kappa_b^{(2)} < \infty$. Turning to the second term, note that

$$\sum_{i=0}^{\infty} (1-\theta)^{2i} = \frac{1}{\theta(2-\theta)}. \tag{A.182}$$

By (G.1a), $\mathrm{E}\left(b_t^{*2}\right) = O(\gamma)$ while the definition of θ implies $\theta/(2-\theta) \leq \gamma$. Hence, the second term behaves asymptotically as $O(\gamma^2)$. Finally, consider the third term

$$\lim_{t\to\infty} \gamma^2\tau^2 \sum_{i=0}^{t-1} (1-\theta)^{2i} = \nu + \frac{(\gamma\tau)^2}{2(2-\theta)} = \nu + O(\gamma^2). \tag{A.183}$$

This completes the proof of (G.1b). □

Remark G.2. *Note that part (G.1b) of the previous lemma implies that*

$$\lim_{\gamma T} \frac{1}{\gamma T} \sum_{t=1}^{T} \mathrm{E}\left(b_t^{*2}\right) = \frac{\tau^2}{2c}, \tag{A.184}$$

where the limit is taken as $\gamma \searrow 0$ such that $\gamma T \to \infty$.

Lemma G.3. *Assume α to be known and set $\tilde{x}_t = b^*_{t-1} x_t$. Henceforth, redefine $b_t = b^*_t$ for notational simplicity. Let assumption 1,2 and 3A hold and set*

$$\tilde{\rho}_{t,t+m} := \mathrm{cov}\left(\tilde{x}_t^2, \tilde{x}_{t+m}^2\right).$$

Then

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

where

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

with $\tilde{\psi} := \kappa_x^{(2)} - \theta\left(2\kappa_x^{(4)}/\kappa_x^{(2)} - \theta\kappa_x^{(6)}/\kappa_x^{(2)2}\right)$ and $\psi := 1 - \theta\left(2 - \theta\kappa_x^{(4)}/\kappa_x^{(2)2}\right)$.

Proof of lemma G.3. Recall from lemma 4 that

$$\mathrm{E}\left(\tilde{x}_t^2, \tilde{x}_{t+m}^2\right) = \mathrm{E}\Big[b_{t-1}^2 \mathrm{E}\left(x_t^2 b_{t+m-1}^2 | \mathcal{V}_{t-1}\right)\Big]\kappa_x^{(2)}, \tag{A.185}$$

so that

$$\tilde{\rho}_{t,t+m} = \kappa_x^{(2)}\Big\{\mathrm{E}\Big[b_{t-1}^2 \mathrm{E}\left(x_t^2 b_{t+m-1}^2 | \mathcal{V}_{t-1}\right)\Big] - \mathrm{E}\left(b_{t-1}^2\right)\kappa_x^{(2)} \mathrm{E}\left(b_{t+m-1}^2\right)\Big\}. \tag{A.186}$$

If $\eta = 0$, then

$$\begin{aligned} \mathrm{E}\left(a_{t+m}^2\right) &= \mathrm{E}\left(a_{t+m-1}^2\right)\psi + \gamma^2\tau^2 \\ &= \mathrm{E}(a_t^2)\psi^m + \gamma^2\tau^2 \sum_{i=0}^{m-1} \psi^i, \end{aligned} \tag{A.187}$$

where $\mathrm{E}(a_t^2) = \mathrm{E}\left(a_{t-1}^2\right)\psi + \gamma^2\tau^2$. Similarly,

$$\mathrm{E}\left(x_t^2 a_{t+m}^2 | \mathcal{V}_{t-1}\right) = \mathrm{E}\left(x_t^2 a_t^2 | \mathcal{V}_{t-1}\right)\psi^m + \gamma^2\sigma^2 \sum_{i=0}^{m-1} \psi^i, \tag{A.188}$$

where

$$\mathrm{E}\left(x_t^2 a_t^2 | \mathcal{V}_{t-1}\right) = a_{t-1}^2 \tilde{\psi} + \gamma^2\sigma^2\kappa_x^{(4)}/\kappa_x^{(2)2}. \tag{A.189}$$

Therefore,

$$\mathrm{E}\Big[a_{t-1}^2 \mathrm{E}\left(x_t^2 a_{t+m}^2 | \mathcal{V}_{t-1}\right)\Big] - \mathrm{E}\left(a_{t-1}^2\right)\kappa_x^{(2)} \mathrm{E}\left(a_{t+m}^2\right) = \pi_t \psi^m \tag{A.190}$$

where

$$\begin{aligned}\pi_t =& \mathrm{E}\Big[a_{t-1}^2 \mathrm{E}\left(x_t^2 a_t^2 | \mathcal{V}_{t-1}\right)\Big] - \mathrm{E}\left(a_{t-1}^2\right) \kappa_x^{(2)} \mathrm{E}\left(a_t^2\right) \\ =& \mathrm{E}\left(a_{t-1}^4\right) \tilde{\psi} - \mathrm{E}\left(a_{t-1}^2\right) \left[\kappa_x^{(2)} \mathrm{E}\left(a_{t-1}^2\right) \psi - \sigma^2\gamma^2 \left(\kappa_x^{(4)}/\kappa_x^{(2)2} - 1\right)\right].\end{aligned} \tag{A.191}$$

G.1 Proposition 2.2

Proposition G.1. *Suppose that assumption 1, 2 and 3A hold. Then*

$$(\gamma T)^{1/2}(\widehat{\beta}_0 - \beta) \xrightarrow{d} \mathcal{N}\left(0, 2cb\right).$$

Proof of lemma G.1. Analogous to the proof for the case $\eta > 0$ (cf. appendix B), consider

$$(\gamma T)^{1/2}(\widehat{\beta}_0 - \beta) = \left(\frac{1}{\gamma T}\sum_{t=1}^{T} \tilde{x}_t^2\right)^{-1} \frac{1}{(\gamma T)^{1/2}} \sum_{t=1}^{T} \tilde{x}_t \varepsilon_t. \tag{A.192}$$

The proof proceeds in two steps:

$$\begin{aligned}&(1) \quad \operatorname*{plim}_{\gamma T} \frac{1}{\gamma T}\sum_{t=1}^{T} \tilde{x}_t^2 = \lim_{\gamma T} \frac{1}{\gamma T}\sum_{t=1}^{T} \mathrm{E}\left(\tilde{x}_t^2\right) = \sigma^2/(2c) \\ &(2) \quad \frac{1}{(\gamma T)^{1/2}}\sum_{t=1}^{T} \tilde{x}_t \varepsilon_t \xrightarrow{d} \mathcal{N}\left(0, \sigma^4/(2c)\right),\end{aligned}$$

where the respective limits are taken as $\gamma \searrow 0$ such that $\gamma T \to \infty$.

Step (1). By lemma G.1, it suffices to show that

$$\operatorname{var}\left(\sum_{t=1}^{T} \tilde{x}_t^2\right) = \sum_{t=1}^{T} \operatorname{var}\left(\tilde{x}_t^2\right) + 2\sum_{t=1}^{T-1}\sum_{s=1}^{T-t} \tilde{\rho}_{t,t+s}, \tag{A.193}$$

where, as defined in lemma 4, $\tilde{\rho}_{t,t+m} = \operatorname{cov}\left(\tilde{x}_t^2, \tilde{x}_{t+m}^2\right)$. The first term behaves like $O(T\gamma^2)$ by part (G.1*a*) of lemma G.1. Turning to the second summand, suppose that $\gamma < 2/(\kappa_x^{(4)} c)$. Then $\psi < 1$ and for some suitable $C_1 < \infty$

$$\begin{aligned}\sum_{t=1}^{T-1}\sum_{s=1}^{T-t} \tilde{\rho}_{t,t+s} =& \sum_{t=1}^{T-1} \tilde{\psi}_t \sum_{s=1}^{T-t} \psi^{s-1} \\ \leq& C_1\gamma^2 \sum_{t=1}^{T-1}\sum_{s=1}^{T-t} \psi^{s-1} \\ =& C_1\gamma^2 \frac{\psi^T - \psi^2(T-1) + \psi(T-2)}{(1-\psi)^2\psi} = O(T\gamma^2),\end{aligned} \tag{A.194}$$

which completes step (1).

Step (2). Note that

$$z_t := \tilde{x}_t \varepsilon_t = b_{t-1} x_t \varepsilon_t \tag{A.195}$$

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

$$s_T^2 := \sum_{t=1}^{T} \sigma_t^2 \quad \text{with} \quad \sigma_t^2 := \mathrm{E}\left(z_t^2\right) = \sigma^2 \mathrm{E}\left(a_{t-1}^2\right) \tag{A.196}$$

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

$$\frac{1}{(\gamma T)^{1/2}} \sum_{t=1}^{T} \tilde{x}_t \varepsilon_t = \left(\frac{1}{\gamma T} s_T^2\right)^{1/2} \sum_{t=1}^{T} Z_{tT}. \tag{A.197}$$

Now, taking account of equation (G.1*b*) of lemma G.1, it is seen that

$$\frac{s_T^2/(\gamma T)}{\sigma^4/(2c)} = 1 + o(1). \tag{A.198}$$

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 to conditions hold:

1. $\sum_{t=1}^{T} Z_{tT} \xrightarrow{p} 1$;

2. $\max_{1 \le t \le T} |Z_{tT}| \xrightarrow{p} 0$.

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

G.2 Proposition 2.3

Next, proposition 2.3 will be shown to hold if $\eta = 0$. The recursive representation (A.177) yields with a little rearrangement that

$$\begin{aligned} H_t + b_0^* c(1-\theta)^{t-1} &= \left(u_t - \theta \sum_{i=1}^{t-1} u_i (1-\theta)^{t-1-i}\right) + \left(e_t - \theta \sum_{i=1}^{t-1} e_i (1-\theta)^{t-1-i}\right) \\ &=: H_t^{(u)} + H_t^{(e)} \quad \text{for } t \ge 2, \end{aligned} \tag{A.199}$$

where $H_t^{(u)}$ and $H_t^{(e)}$ denote respectively the two terms in parentheses on the right hand-side of the equality. Let $H_{\mathcal{T}}^{(z)} := T^{-1/2} \sum_{t=2}^{T} H_t^{(z)}$ for $z \in \{e, u\}$, then $H_{\mathcal{T}}$ might be expressed as:

$$\begin{aligned} H_{\mathcal{T}} + o_p(1) &= H_{\mathcal{T}}^{(u)} + H_{\mathcal{T}}^{(e)} + \frac{H_1}{\sqrt{T}} \\ &= H_{\mathcal{T}}^{(u)} + H_{\mathcal{T}}^{(e)} + o_p(1), \end{aligned} \tag{A.200}$$

where the $o_p(1)$ term on the left-hand side follows upon noting that $\kappa_b^{(2)} < \infty$ and $\sum_{t=1}^{\infty}(1-\theta)^{t-1} < \infty$, whereas the last equality follows directly from (A.114) taking into account that u_1, e_1 and a_0 are uncorrelated. The following two steps establish the asymptotic negligibility of the partial sums $H_{\mathcal{T}}^{(u)}$ and $H_{\mathcal{T}}^{(e)}$, both of which are mean-zero.

Negligibility of $H_{\mathcal{T}}^{(u)}$. By the weak LLN, it suffices to show that

$$\operatorname{var}\left(H_{\mathcal{T}}^{(u)}\right) = \frac{1}{T}\sum_{t=2}^{T-1} \mathrm{E}\left[\left(H_t^{(u)}\right)^2\right] + \frac{2}{T}\sum_{t=2}^{T}\sum_{m=1}^{T-t} \mathrm{E}\left(H_t^{(u)} H_{t+m}^{(u)}\right) =: A_T + B_T \tag{A.201}$$

approaches zero as $T \to \infty$. It is not difficult to see that the summands of the first partial sum are given by $\tau^2\left(1+\theta^2\sum_{i=1}^{t-1}(1-\theta)^{2(t-1-i)}\right)$. Since

$$\sum_{i=1}^{t-1}(1-\theta)^{2(t-1-i)} = \frac{1-(1-\theta)^{2(t-1)}}{\theta(2-\theta)} \to \frac{1}{\theta(2-\theta)}, \tag{A.202}$$

it follows from Cèsaro's mean convergence theorem that $A_T \to A := \tau^2/(1-\theta/2)$. Hence, it remains to establish $-B_T \to A$. Now, observe that

$$\begin{aligned}
\mathrm{E}\left(H_t^{(u)} H_{t+m}^{(u)}\right) &= -\theta \sum_{i=1}^{t+m-1} \mathrm{E}(u_t u_i)(1-\theta)^{t+m-1-i} + \theta^2 \sum_{i=1}^{t-1}\sum_{j=1}^{t+m-1} \mathrm{E}(u_i u_j)(1-\theta)^{t-1-i}(1-\theta)^{t+m-1-j} \\
&= -\tau^2\theta(1-\theta)^{m-1} + \tau^2\theta^2\sum_{i=1}^{t-1}(1-\theta)^{2t+m-2-2i} \\
&= \tau^2\theta(1-\theta)^{m-1}\left[\theta\sum_{i=1}^{t-1}(1-\theta)^{2t-2i-1} - 1\right] \\
&= \tau^2\theta(1-\theta)^{m-1}\left[\theta\sum_{i=1}^{t-1}(1-\theta)^{2i-1} - 1\right] \quad \text{for } m > 0,
\end{aligned} \tag{A.203}$$

where the last equality is due to

$$\begin{aligned}
\sum_{i=1}^{t-1}(1-\theta)^{2i-1} &= \frac{\theta^2 - (1-\theta)^{2t} - 2\theta + 1}{\theta^3 - 3\theta^2 + 2\theta} \\
&= \sum_{i=1}^{t-1}(1-\theta)^{2t-2i-1}.
\end{aligned} \tag{A.204}$$

Note that the limit of the partial sum (A.204) is $(1-\theta)/(\theta(2-\theta))$. Taking this into account, B_T is decomposed as

$$B_T = \frac{\theta}{\theta-2}\frac{2}{T}\sum_{t=2}^{T-1}\sum_{m=1}^{T-t}(1-\theta)^{m-1} - \frac{2}{(1-\theta)(2-\theta)}\frac{1}{T}\sum_{t=2}^{T-1}\left[(1-\theta)^{2t} - (1-\theta)^{t+T}\right], \tag{A.205}$$

where it has been used that

$$\theta \sum_{i=1}^{t-1} (1-\theta)^{2i-1} - 1 = \frac{1}{\theta - 2} - \frac{(1-\theta)^{2t}}{(1-\theta)(2-\theta)} \tag{A.206}$$

$$\sum_{m=1}^{T-t} (1-\theta)^{2t+m-1} = \frac{(1-\theta)^{2t} - (1-\theta)^{t+T}}{\theta}. \tag{A.207}$$

As the second term on the right-hand side of (A.205) tends to zero while

$$\frac{1}{T} \sum_{t=2}^{T-1} \sum_{m=1}^{T-t} (1-\theta)^{m-1} = \frac{\theta^2(T-1) - (1-\theta)^T - \theta T + 1}{T\theta^2(\theta - 1)} \to \frac{1}{\theta}, \tag{A.208}$$

it follows that $-B_T \to A$, which proves $H_T^{(u)} = o_p(1)$.

Negligibility of $H_T^{(e)}$. Analogously to the previous part, it will be shown that

$$\operatorname{var}\left(H_T^{(e)}\right) = \frac{1}{T} \sum_{t=2}^{T} \mathrm{E}\left[\left(H_t^{(e)}\right)^2\right] + \frac{2}{T} \sum_{t=2}^{T-1} \sum_{m=1}^{T-t} \mathrm{E}\left(H_t^{(e)} H_{t+m}^{(e)}\right) =: A_T + B_T \tag{A.209}$$

tends to zero as $T \to \infty$. Mirroring the previous arguments, it is readily established that

$$\mathrm{E}\left(H_t^{(e)} H_{t+m}^{(e)}\right) = \begin{cases} \sigma^2(e_t) + \theta^2 \sum_{i=1}^{t-1} \sigma^2(e_i)(1-\theta)^{2(t-1-i)} & \text{if } m = 0 \\ \theta(1-\theta)^{m-1} \left[\theta \sum_{i=1}^{t-1} \sigma^2(e_i)(1-\theta)^{2t-2i-1} - \sigma^2(e_t)\right] & \text{if } m > 0, \end{cases} \tag{A.210}$$

where (A.10) has been used. Consequently, A_T can be rewritten as:

$$A_T = c^2 \left(\kappa_x^{(4)}/\kappa_x^{(2)2} - 1\right) \left[\frac{1}{T} \sum_{t=2}^{T} \mathrm{E}\left(a_{t-1}^{*\,2}\right) + \theta^2 \frac{1}{T} \sum_{t=2}^{T} \sum_{i=1}^{t-1} \mathrm{E}\left(a_{i-1}^{*\,2}\right)(1-\theta)^{2(t-1-i)}\right]. \tag{A.211}$$

It will be shown that

$$A_T \to A := c^2 \left(\kappa_x^{(4)}/\kappa_x^{(2)2} - 1\right) \frac{\nu}{1 - \theta/2}, \tag{A.212}$$

where $\nu := \lim_t \mathrm{E}\left(a_t^{*2}\right)$; cf. lemma 3. In order to see this, take account of (A.202) and let

$$\xi_{it} := \theta(2-\theta)(1-\theta)^{2(t-1-i)}. \tag{A.213}$$

Since $\lim_{t\to\infty} \xi_{it} = 0$ and $\lim_{t\to\infty} \sum_{i=1}^{t-1} \xi_{it} = 1$, Toeplitz's lemma and lemma 3 yield

$$\lim_{t\to\infty} \sum_{i=1}^{t-1} \mathrm{E}\left(a_{i-1}^{*\,2}\right)(1-\theta)^{2(t-1-i)} = \frac{1}{\theta(2-\theta)} \lim_{t\to\infty} \sum_{i=1}^{t-1} \mathrm{E}\left(a_{i-1}^{*\,2}\right) \xi_{it} = \frac{\nu}{\theta(2-\theta)}. \tag{A.214}$$

Equation (A.212) follows upon applying Cèsaro's mean convergence theorem. Turning to B_T, note first that

$$\sum_{m=1}^{T-t} \theta(1-\theta)^{m-1} = 1 - (1-\theta)^{T-t} \tag{A.215}$$

implies

$$
\begin{aligned}
B_T &= \frac{2}{T}\sum_{t=2}^{T-1}\left(\theta\sum_{i=1}^{t-1}\sigma^2(e_i)(1-\theta)^{2t-2i-1}-\sigma^2(e_t)\right)\left(1-(1-\theta)^{T-t}\right) \\
&= \frac{2}{T}\sum_{t=2}^{T-1}\theta\sum_{i=1}^{t-1}\sigma^2(e_i)(1-\theta)^{2t-2i-1} \\
&\quad -(1-\theta)^T\frac{2}{T}\sum_{t=2}^{T-1}\theta\sum_{i=1}^{t-1}\sigma^2(e_i)(1-\theta)^{t-2i-1} \\
&\quad -\frac{2}{T}\sum_{t=2}^{T-1}\sigma^2(e_t) \\
&\quad +\frac{2}{T}\sum_{t=2}^{T-1}\sigma^2(e_t)(1-\theta)^{T-t} =: B_T^i - B_T^{ii} - B_T^{iii} + B_T^{iv}.
\end{aligned} \tag{A.216}
$$

Begin with B_T^i and define $\xi_{it} := (1-\theta)^{2t-2i-2}\theta(2-\theta)$. Taking account of (A.204), it is seen that that $\lim_{t\to\infty}\xi_{it} = 0$ while $\lim_{t\to\infty}\sum_{i=1}^{t-1}\xi_{it} = 1$. Hence, by Toeplitz's lemma and Cèsaro's mean convergence theorem

$$
B_T^i = \left(\frac{1-\theta}{2-\theta}\right)\frac{2}{T}\sum_{t=2}^{T-1}\sum_{i=1}^{t-1}\sigma^2(e_i)\xi_{i,t} \to c^2\left(\kappa_x^{(4)}-1\right)2\nu\left(\frac{1-\theta}{2-\theta}\right). \tag{A.217}
$$

Next, by lemma 3, there exists a finite constant C such that

$$
\begin{aligned}
B_T^{ii} &\leq C(1-\theta)^T\frac{2}{T}\sum_{t=2}^{T-1}\theta\sum_{i=1}^{t-1}(1-\theta)^{t-2i-1} \\
&= C(1-\theta)^T\frac{2}{T}\sum_{t=2}^{T-1}O\left((1-\theta)^{-t}\right) \\
&= C(1-\theta)^T\frac{2}{T}O\left((1-\theta)^{-T}\right) = O(1/T).
\end{aligned} \tag{A.218}
$$

By lemma 3, $B_T^{iii} \to 2c^2\left(\kappa_x^{(4)}-1\right)\nu$ while, similar to B_T^{iii}, $B_T^{iv} = O(1/T)$. Hence, upon collecting terms, it is seen that $-B_T \to A$, which proves that $H_T^{(e)} = o_p(1)$. This completes the proof for the constant gain case. □

G.3 Proposition 2.1

The following equivalent of proposition 2.1 will be established:

Proposition G.2. *Suppose assumption 1 with $\eta = 0$ so that $\gamma^2 T \not\to 0$, assumption 2 and 3A hold. Then*

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

Proof of proposition G.2. The following closely mirrors the proof of proposition 2.1 contained in appendix D. Specifically, consider

$$(\gamma T)^{1/2}(\widehat{\lambda}-\lambda) = \gamma T W_T^{-1}\frac{1}{(\gamma T)^{1/2}}\sum_{t=1}^{T} w_t\varepsilon_t = \frac{\gamma T^2}{\det W_T}\left(\frac{Q_T/T}{(\gamma T)^{1/2}}\sum_{t=1}^{T} w_t\varepsilon_t\right). \tag{A.219}$$

The main idea is to show that

$$\text{(1)} \qquad \operatorname*{plim}_{\gamma T}\frac{\det W_T}{\gamma T^2} = \sigma^2\kappa_x^{(2)}/(2c)$$

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

$$\text{(3)} \qquad \frac{Q_T/T}{(\gamma T)^{1/2}}\sum_{t=1}^{T} w_t\varepsilon_t \xrightarrow{d} \mathcal{N}\left(0, \frac{\kappa_x^{(2)2}\sigma^4}{2c}\begin{bmatrix}1 & -\alpha\\ -\alpha & \alpha^2\end{bmatrix}\right),$$

where the respective limits are taken as $\gamma \searrow 0$ such that $\gamma^2 T \nrightarrow 0$.

Step (1). Analogous to (A.131), consider

$$\frac{\det W_T}{\gamma T^2} = m_T\left[\frac{1}{\gamma T}\sum_{t=1}^{T}\tilde{x}_t^2 - m_T^{-1}\left(\frac{1}{\gamma^{1/2}T}\sum_{t=1}^{T} x_t^2 b_{t-1}^*\right)^2\right]. \tag{A.220}$$

Similarly to (A-D.1), it will first be shown that under the conditions of the proposition

$$\frac{1}{T^{1/2}}\sum_{t=1}^{T} x_t^2 b_{t-1}^* = O_p(1). \tag{A-G.1}$$

Proof of (A-G.1). From (A.180) it follows that

$$\begin{aligned}\frac{1}{T^{1/2}}\sum_{t=1}^{T} x_t^2 b_{t-1}^* &= \frac{b_0^*}{T^{1/2}}\sum_{t=1}^{T} x_t^2(1-\theta)^{t-1} + \frac{\gamma}{T^{1/2}}\sum_{t=1}^{T} x_t^2\tilde{u}_{t-1} + \frac{\gamma}{T^{1/2}}\sum_{t=1}^{T} x_t^2\tilde{e}_{t-1}\\ &=: A_T + B_T + C_T,\end{aligned} \tag{A.221}$$

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

$$\mathrm{E}|A_T| = \frac{\mathrm{E}|b_0^*|\kappa_x^{(2)}}{T^{1/2}}\sum_{t=1}^{T}(1-\theta)^{t-1} = O\left(\frac{1}{\gamma T^{1/2}}\right), \tag{A.222}$$

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

$$\|A_T\|_2^2 = \frac{\kappa_b^{(2)}}{T}\sum_{t,s=1}^{T}\mathrm{E}(x_t^2x_s^2)(1-\theta)^{t-1}(1-\theta)^{s-1} \leq \kappa_b^{(2)}\kappa_x^{(4)}\left(\frac{1}{T^{1/2}}\sum_{t=1}^{T}(1-\theta)^{t-1}\right)^2, \tag{A.223}$$

where Cauchy Schwarz's inequality has been used. The term in brackets on the majorant side of (A.223) has already been shown to be of size $O\left(\gamma^{-1}T^{-1/2}\right)$, thereby proving that $A_T = O_p(1)$ if $\gamma^2 T \nrightarrow 0$. Before turning to B_T and C_T, note that the definitions of (A.178) and (A.179) imply that

$$\mathrm{E}(\tilde{u}_t\tilde{u}_s) = \begin{cases} \tau^2 \sum_{i=0}^{t-1}(1-\theta)^{2i} & \text{if } s=t \\ \tau^2 \sum_{i=0}^{s-1}(1-\theta)^{t-s+2i} & \text{if } s<t \end{cases} \tag{A.224}$$

$$\mathrm{E}(\tilde{e}_t\tilde{e}_s) = \begin{cases} C_0 \sum_{i=0}^{t-1}(1-\theta)^{2i}\mathrm{E}(b^{*2}_{t-1-i}) & \text{if } s=t \\ C_0 \sum_{i=0}^{s-1}(1-\theta)^{t-s+2i}\mathrm{E}(b^{*2}_{s-1-i}) & \text{if } s<t, \end{cases} \tag{A.225}$$

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

$$B_T = \frac{\kappa_x^{(2)}\gamma}{T^{1/2}}\sum_{t=1}^{T}\check{u}_{t-1} + \frac{\kappa_x^{(2)}\gamma}{T^{1/2}}\sum_{t=1}^{T}(x_t^{*2}-1)\check{u}_{t-1} =: B_T^{(1)} + B_T^{(2)}, \tag{A.226}$$

say. Now, from (A.224) one gets

$$\left\|B_T^{(1)}\right\|_2^2 = \frac{(\kappa_x^{(2)}\tau)^2\gamma^2}{T}\sum_{t=1}^{T}\sum_{i=0}^{t-1}(1-\theta)^{2i} + \frac{2(\kappa_x^{(2)}\tau)^2\gamma^2}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{i=0}^{s-1}(1-\theta)^{t-s+2i}. \tag{A.227}$$

Note that

$$\begin{aligned}\sum_{t=1}^{T}\sum_{i=0}^{t-1}(1-\theta)^{2i} &= \frac{T}{(2-\theta)\theta} - \frac{1-(1-\theta)^{2T}+\theta(2-\theta)((\theta-1)^{2T}-1)}{(2-\theta)^2\theta^2} \\ &= O\left(\frac{T}{\gamma}\right) + O\left(\frac{1}{\gamma^2}\right).\end{aligned} \tag{A.228}$$

Similarly,

$$\begin{aligned}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{i=0}^{s-1}(1-\theta)^{t-s+2i} &= \left[\frac{(1-\theta)^{T+1}}{\theta^3} - \frac{(\theta-1)^{2(T+1)}+(1-\theta)^2}{(2-\theta)^2\theta^3}\right] + T\frac{1-\theta}{(2-\theta)\theta^2} - \frac{1-\theta}{(2-\theta)^2\theta} \\ &= O\left(\frac{1}{\gamma^3}\right) + O\left(\frac{T}{\gamma^2}\right) + O\left(\frac{1}{\gamma}\right).\end{aligned} \tag{A.229}$$

Hence it follows that $\left\|B_T^{(1)}\right\|_2^2 = O(1)$ and thus $B_T^{(1)} = O_p(1)$ as $\mathrm{E}(B_T^{(1)}) = 0$. Turning to $B_T^{(2)}$ note that its summands are uncorrelated with mean zero so that

$$\left\|B_T^{(2)}\right\|_2^2 = \frac{\tau^2(\kappa_x^{(4)} - \kappa_x^{(2)2})\gamma^2}{T}\sum_{t=1}^{T}\sum_{i=0}^{t-1}(1-\theta)^{2i} = O(\gamma) + O\left(\frac{1}{T}\right), \tag{A.230}$$

using the same arguments as before. Hence, $B_T = O_p(1)$.

Finally, consider C_T and let

$$C_T = \frac{\kappa_x^{(2)}\gamma}{T^{1/2}}\sum_{t=1}^{T}\check{e}_{t-1} + \frac{\kappa_x^{(2)}\gamma}{T^{1/2}}\sum_{t=1}^{T}(x_t^{*2}-1)\check{e}_{t-1} =: C_T^{(1)} + C_T^{(2)}, \tag{A.231}$$

say. (G.1a) ensures the existence of some finite constants $K_0 > 0$ and $K_1 > 0$ so that

$$\left\|C_T^{(1)}\right\|_2^2 \leq K_0\,\gamma\left\|B_T^{(1)}\right\|_2^2 \text{ and } \left\|C_T^{(2)}\right\|_2^2 \leq K_1\,\gamma\left\|B_T^{(2)}\right\|_2^2 \tag{A.232}$$

what proves (A-G.1). □

Now, the claim follows by step (1) of appendix G.1, (A-G.1) and the almost sure convergence of $m_{\mathcal{T}}$. □

Step (2). Taking the above into account, the proof is analogous to that of step (2) in appendix D.

Step (3). Similarly to (A.148), the entries of the 2×1 vector

$$\frac{Q_T/T}{\sqrt{\gamma T}}\sum_{t=1}^{T} w_t\varepsilon_t \tag{A.233}$$

can be written as

$$\begin{aligned}
\frac{Q_T/T}{\sqrt{\gamma T}}\sum_{t=1}^{T} w_t\varepsilon_t &= \begin{bmatrix}1\\-\alpha\end{bmatrix}\frac{m_{\mathcal{T}}}{\sqrt{\gamma T}}\sum_{t=1}^{T}\tilde{x}_t\varepsilon_t \\
&\quad - \begin{bmatrix}1\\-\alpha\end{bmatrix}\left(\frac{\kappa_x^{(2)}}{\gamma^{1/2}T}\sum_{t=1}^{T}x_t^2 b_{t-1}^*\right)u_{\mathcal{T}} \\
&\quad + \begin{bmatrix}0\\1\end{bmatrix}\left\{\left(\frac{\kappa_x^{(2)}}{\gamma^{1/2}T}\sum_{t=1}^{T}\tilde{x}_t^2\right)u_{\mathcal{T}} - \left(\frac{1}{T}\sum_{t=1}^{T}x_t^2 b_{t-1}^*\right)\frac{1}{\sqrt{\gamma T}}\sum_{t=1}^{T}\tilde{x}_t\varepsilon_t\right\} \\
&= \begin{bmatrix}1\\-\alpha\end{bmatrix}\frac{m_{\mathcal{T}}}{\gamma^{1/2}T}\sum_{t=1}^{T}\tilde{x}_t\varepsilon_t + O_p\left(\frac{1}{\sqrt{\gamma T}}\right) + O_p\left(\sqrt{\gamma}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),
\end{aligned} \tag{A.234}$$

using again that $u_{\mathcal{T}} = O_p(1)$ together with step (1) and (2) of appendix G.1 together with (A-G.1). Hence, it follows that (A.233) is asymptotically equal to

$$\begin{bmatrix}1\\-\alpha\end{bmatrix}\frac{\kappa_x^{(2)}}{\sqrt{\gamma T}}\sum_{t=1}^{T}\tilde{x}_t\varepsilon_t \xrightarrow{d} \mathcal{N}\left(0, \frac{\kappa_x^{(2)2}\sigma^4}{2c}\begin{bmatrix}1 & -\alpha\\ -\alpha & \alpha^2\end{bmatrix}\right). \tag{A.235}$$

The limiting distribution is a direct consequence of step (2) of appendix G.1. Using step (1) of this proof in conjunction with Slutzky's theorem gives the stated result. □

G.4 Corollary 2.1

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

$$\begin{aligned}
T^{1/2}\left(\widehat{\lambda}_\alpha-\alpha\right) &= \frac{(\det W_T/(\gamma T^2))^{-1}}{1-\widehat{\lambda}_\beta}\frac{1}{T^{1/2}}\sum_{t=1}^T b_{t-1}x_t\varepsilon_t\left(\frac{\alpha}{\gamma T}\sum_{t=1}^T x_t^2-\frac{1}{\gamma T}\sum_{t=1}^T x_t^2 b_{t-1}\right) \\
&\quad+\frac{(\det W_T/(\gamma T^2))^{-1}}{1-\widehat{\lambda}_\beta}\frac{1}{T^{1/2}}\sum_{t=1}^T x_t\varepsilon_t\left(\frac{1}{\gamma T}\sum_{t=1}^T(x_t b_{t-1})^2-\frac{\alpha}{\gamma T}\sum_{t=1}^T x_t^2 b_{t-1}\right) \\
&= \frac{u_T}{1-\widehat{\lambda}_\beta}\left(\frac{\kappa_x^{(2)}(\det W_T/(\gamma T^2))^{-1}}{\gamma T}\sum_{t=1}^T\tilde{x}_t^2\right) \\
&\quad-\frac{(\det W_T/(\gamma T^2))^{-1}}{1-\widehat{\lambda}_\beta}\left(\frac{1}{\gamma^{1/2}T}\sum_{t=1}^T x_t^2 b_{t-1}^*\right)\left(\frac{1}{(\gamma T)^{1/2}}\sum_{t=1}^T\tilde{x}_t\varepsilon_t\right). \qquad \text{(A.236)}
\end{aligned}$$

Taking account of step (1) and (2) of proposition G.1 together with (A-G.1) and step (1) of the proceeding proof, it follows from the same arguments as that used in proving corollary 2.1 that

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

□

H Proof of remark 2.3

Assumption 2 M. *The elements of the random vector* $v_t := (x_t', \varepsilon_t)'$ *are mutually independent and identically distributed so that* $\mathrm{E}(\varepsilon_t^2)=\sigma^2\in(0,\infty)$ *and* $\mathrm{E}(x_t x_t'):=Q$ *is positive definite.*

Similar to the auxiliary assumption 3A, it is assumed that the recursion for the regressor second moment matrix is centered at $R_t = Q$ for all t. The agent's recursion (13) for the k dimensional RE equilibrium vector α is thus given by

$$b_t = b_{t-1}+\gamma_t Q^{-1}x_t(y_t-b_{t-1}'x_t), \qquad \text{(A.237)}$$

so that $y_{t|t-1}^e = b_{t-1}'x_t$ represents the agent's forecast of y_t.

Assumption 3 M. *The elements of* v_t *satisfy assumption 3-SG and* $\mathrm{E}(b_0^* b_0^{*\prime})<\infty$.

Lemma H.1. *Let* $\tilde{x}_t := b_{t-1}^{*\prime}x_t$. *Then*

$$\frac{1}{T^b}\sum_{t=1}^T\tilde{x}_t^2\xrightarrow{p}k\sigma^2\frac{\gamma}{2cb}.$$

Proof. To begin with, consider

$$\mathrm{E}(\tilde{x}_t^2) = \mathrm{E}(x_t' b_{t-1}^* b_{t-1}^{*\prime} x_t) = \mathrm{E}\left(\mathrm{tr}\left[x_t x_t' b_{t-1}^* b_{t-1}^{*\prime}\right]\right) = \mathrm{tr}\left[Q\mathrm{E}(b_{t-1}^* b_{t-1}^{*\prime})\right], \tag{A.238}$$

where tr[·] represents the trace of a matrix. Next, note that

$$\frac{1}{T^b}\sum_{t=1}^{T}\mathrm{E}(b_t^* b_t^{*\prime}) \to \sigma^2 Q^{-1}\frac{\gamma}{2cb}. \tag{A.239}$$

In order to see this, observe that analogous to (A.5), one gets

$$\begin{aligned} b_t^* &= b_{t-1}^*(1 - c\gamma_t) + \gamma_t Q^{-1} x_t \varepsilon_t + \gamma_t c(I_k - Q^{-1} x_t x_t') b_{t-1}^* \\ &= b_0^* \Phi_{t,1} + \breve{u}_t + \breve{e}_t, \end{aligned} \tag{A.240}$$

with

$$\breve{z}_t := \sum_{k=1}^{t}\gamma_{k,t} z_k \quad \text{for} \quad z \in \{u,\, e\}, \tag{A.241}$$

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

$$\mathrm{E}(b_t^* b_t^{*\prime}) = \mathrm{E}(b_0^* b_0^{*\prime})\Phi_{t,1}^2 + \mathrm{E}(\breve{u}_t \breve{u}_t') + \mathrm{E}(\breve{e}_t \breve{e}_t'). \tag{A.242}$$

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

$$\mathrm{E}(\breve{u}_t \breve{u}_t') = \sum_{k,\,s=1}^{t}\gamma_{k,t}\gamma_{s,t}\mathrm{E}(u_k u_s') = \sigma^2 Q^{-1}\sum_{k=1}^{t}\gamma_{k,t}^2 = \sigma^2 Q^{-1}\frac{\gamma_t}{2c} + o(\gamma_t), \tag{A.243}$$

and the third term is of order $O(\gamma_t^2)$. This proves (A.239).

Lemma H.2. *Let* $z_t := \tilde{x}_t\varepsilon_t$. *Then*

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

Proof. Note that

$$\mathrm{var}(z_t) = \sigma^2\mathrm{E}(\tilde{x}_t^2) = \sigma^2\mathrm{tr}\Big[Q\,\mathrm{E}\left(b_{t-1}^* b_{t-1}^{*\prime}\right)\Big], \tag{A.244}$$

so that by lemma H.1 above

$$\frac{1}{T^b}\sum_{t=1}^{T}\mathrm{var}(z_t) = \sigma^2\mathrm{tr}\Big[Q\,\frac{1}{T^b}\sum_{t=1}^{T}\mathrm{E}\left(b_{t-1}^* b_{t-1}^{*\prime}\right)\Big] \to k\sigma^4\frac{\gamma}{2cb}. \quad \square \tag{A.245}$$

Define the T dimensional vector $y^e := (y^e_{1|0}, \dots, y^e_{T|T-1})'$, where $y^e_{t|t-1} = b'_{t-1}x_t$ and the $T \times k$ dimensional matrix $X = (x'_1, \dots, x'_T)'$. Furthermore, define the $k+1$ dimension vector $w_t := (y^e_{t|t-1},\, x'_t)'$ so that

$$M_T := \sum_{t=1}^{T} w_t w'_t. \tag{A.246}$$

It follows that

$$M_T^{-1} = \frac{1}{\phi_T}\begin{bmatrix} 1 & -\pi_T \\ -\pi_T & B_T \end{bmatrix}, \tag{A.247}$$

where

$$\pi_T := (X'X)^{-1}X'y^e \tag{A.248}$$

$$B_T := (X'X)^{-1}\Big[I_k\phi_T + X'y^e y^{e\prime}X(X'X)^{-1}\Big] \tag{A.249}$$

$$\phi_T := y^{e\prime}y^e - y^{e\prime}X(X'X)^{-1}X'y^e, \tag{A.250}$$

see, e.g., Abadir and Magnus (2006, exercise 5.15). Note that

$$\frac{U_T}{T^{(1+b)/2}} = O_p\left(\frac{1}{T^{b/2}}\right) \quad \text{where } U_T := \sum_{t=1}^{T} x_t\tilde{x}'_t \tag{A.251}$$

$$\frac{X'X}{T} = Q + o_p(1) \tag{A.252}$$

$$\frac{X'y^e}{T} = \frac{X'X}{T}\alpha + \frac{U_T}{T} = Q\alpha + o_p(1) \tag{A.253}$$

$$\frac{\phi_T}{T^b} = \frac{\tilde{X}'\tilde{X}}{T^b} - \frac{U'_T}{T^{(1+b)/2}}\left(\frac{X'X}{T}\right)^{-1}\frac{U_T}{T^{(1+b)/2}} = k\sigma^2\frac{\gamma}{2cb} \tag{A.254}$$

$$B_T = \left(\frac{X'X}{T}\right)^{-1}\left[I_k(\phi_T/T) + \left(\frac{X'y^e}{T}\right)\left(\frac{X'y^e}{T}\right)'\left(\frac{X'X}{T}\right)^{-1}\right] = \alpha\alpha' + o_p(1), \tag{A.255}$$

where $\tilde{X} := (\tilde{x}_1, \dots, \tilde{x}_T)'$. Taking this into account, it follows that

$$T^b M_T = k\sigma^2\frac{\gamma}{2cb}\ell_\alpha\ell'_\alpha + o_p(1), \tag{A.256}$$

where $\ell_\alpha := (1, -\alpha')'$. Now, consider $\widehat{\lambda} - \lambda = M_T^{-1}W'\varepsilon$, where $\varepsilon := (\varepsilon_1, \dots, \varepsilon_T)'$ and $W = (y^e, X)$ is the $T \times (k+1)$ sample matrix. With a little rearrangement (similar to that of (A.143)), it follows

$$
\begin{aligned}
T^{b/2}(\widehat{\lambda}-\lambda) = & \frac{1}{\phi_T/T^b}\begin{bmatrix}1\\-\alpha\end{bmatrix}\frac{\tilde{X}'\varepsilon}{T^{b/2}} \\
& - \frac{1}{\phi_T/T^b}\begin{bmatrix}1\\-\alpha\end{bmatrix}\frac{U_T'}{T^{(1+b)/2}}\left(\frac{X'X}{T}\right)^{-1}\frac{X'\varepsilon}{T^{1/2}} \\
& + \frac{1}{\phi_T/T^b}\begin{bmatrix}0_k'\\I_k\end{bmatrix}\left[\left(\frac{X'X}{T}\right)^{-1}\frac{\phi_T}{T^{(1+b)/2}} + \left(\frac{X'X}{T}\right)^{-1}\frac{U_T}{T^{(1+b)/2}}\frac{U_T'}{T}\left(\frac{X'X}{T}\right)^{-1}\right]\frac{X'\varepsilon}{T^{1/2}} \\
& + \frac{1}{\phi_T/T^b}\begin{bmatrix}0_k'\\I_k\end{bmatrix}\left(\frac{X'X}{T}\right)^{-1}\frac{U_T}{T^{(1+b)/2}}\frac{\tilde{X}'\varepsilon}{T^{1/2}} \\
= & \frac{2cb}{k\gamma\sigma^2}\begin{bmatrix}1\\-\alpha\end{bmatrix}\frac{\tilde{X}'\varepsilon}{T^{b/2}} + o_p(1) \xrightarrow{d} \mathcal{N}\left(0_{k+1}, V\right) \quad \text{with } V := \frac{2cb}{k\gamma}\ell_\alpha\ell_\alpha'. \qquad \text{(A.257)}
\end{aligned}
$$

Next, similar to the proof of corollary 2.1, it is seen that

$$
\begin{aligned}
T^{1/2}I_\alpha(\widehat{\lambda}-\lambda) = & T^{1/2}(\widehat{\lambda}_\delta - \delta + \alpha(\widehat{\lambda}_\beta - \beta)) \\
= & \left(\frac{X'X}{T}\right)^{-1}\frac{X'\varepsilon}{T^{1/2}} \\
& - \frac{1}{\phi_T/T^b}\left(\frac{X'X}{T}\right)^{-1}\frac{U_T}{T^{1/2}}\frac{\tilde{X}'\varepsilon}{T^b} \\
& + \frac{1}{\phi_T/T^b}\left(\frac{X'X}{T}\right)^{-1}\frac{U_T}{T^b}\frac{U_T'}{T}\left(\frac{X'X}{T}\right)^{-1}\frac{X'\varepsilon}{T^{1/2}} \\
= & Q^{-1}\frac{X'\varepsilon}{T^{1/2}} + o_p(1) \xrightarrow{d} \mathcal{N}(0_k, \sigma^2 Q^{-1}). \qquad \text{(A.258)}
\end{aligned}
$$