Nonlinear time series models and weakly dependent innovations

Robert M. de Jong*

July 11, 2005

Abstract

This paper provides a strict stationarity proof for general nonlinear regression models. In particular, it shows the existence of a strictly stationary solution to first order dynamic nonlinear models when the innovations are only assumed to satisfy a strict stationarity or strong mixing condition. The results of this paper can be applied to show the strict stationarity of threshold unit root models under general conditions that include weak dependence assumptions on the innovations. Previous results of this type were derived using a Markov chain approach, which uses an i.i.d. assumption on the innovations. The results are applied to threshold autoregressive models to yield strict stationarity results for threshold unit root models in the presence of weakly dependent errors.

1 Introduction

This paper establishes strict stationarity results for general nonlinear models of the type

$$x_t = g(x_{t-1}) + \varepsilon_t \tag{1}$$

for $t \in \mathbb{Z}$, where as little as possible is assumed regarding ε_t beyond its strict stationarity, and I will also attempt to make minimal assumptions regarding $g(\cdot)$. There do not currently appear to exist results in the probability or econometrics literature that show weak

*Department of Economics, Ohio State University, 429 Arps Hall, 1945 N. High Street, Columbus, OH 43210, USA, email `dejong@econ.ohio-state.edu`. I thank Benedikt Pötscher, Maxwell Stinchcombe, and Werner Ploberger for helpful discussions.

dependence properties for x_t as generated above for the case of weakly dependent ε_t, and this paper aims to fill this gap in the literature. As a consequence, this paper will provide conditions under which a class of threshold unit root models are strictly stationary. As Seo (2004) points out, "Unfortunately, however, our understanding is not complete as to the stationarity conditions for general TAR processes. When the errors are independent, Chan, Petruccelli, Tong, and Woolford (1985) provide a necessary and sufficient condition...". The results of this paper can be used to show the strict stationarity of a class of threshold autoregressive (TAR) models, thereby filling in a part of this lack of understanding. In addition, this paper aims to develop a new and general methodology for deriving strict stationarity for nonlinear models, and the methodology has the potential for being adapted to showing strict stationarity in different models.

This paper shows the existence of a strictly stationary solution to the model of Equation (1), but the results in this paper have two limitations. First, no uniqueness of the strictly stationary solution is proven; and second, no attempt will be made to establish properties of x_t generated according to the law of motion of Equation (1) when the process is started from an arbitrary starting value x_0. Therefore, the results of this paper can only be used to establish strict stationarity for x_t generated according to Equation (1) under the assumption that x_0 is drawn from the stationary distribution and not for arbitrary starting values x_0. For the case of threshold models and i.i.d. errors, it is known that uniqueness and convergence to the stationary distribution can be shown, but such results are outside the scope of this paper.

Throughout this paper, it is assumed that $g(\cdot)$ is a Borel measurable function and that $|g(x)| < \infty$ for any $x \in \mathbb{R}$. Currently, the literature on the topic of establishing fading memory properties for models such as the above seems to be based on three different approaches. The first approach is based on an assumption of a Lipschitz coefficient for $g(\cdot)$ that is strictly less than 1. This approach seems to have been pioneered by Bierens (1981); Pötscher and Prucha (1997) contains more results of this type, and these authors discuss generalization of this result towards the case of multiple lags occurring in the response function $g(\cdot)$. The condition of a Lipschitz coefficient that is strictly less than 1 excludes a number of interesting examples, and in particular it rules out threshold unit root models, such as for example the model

$$y_t = y_{t-1} + \phi y_{t-1} I(|y_{t-1}| > c) + \varepsilon_t \tag{2}$$

where $-2 < \phi < 0$, or the case

$$y_t = 3\Phi(y_{t-1}) + \varepsilon_t \tag{3}$$

where $\Phi(\cdot)$ denotes the standardnormal distribution function.

A second type of results is based on a Markov chain approach. This approach uses the i.i.d. condition on ε_t and employs advanced techniques from Markov chain theory. Tong (1990, Appendix 1) and Meyn and Tweedie (1993) provide treatments of this subject. Using this assumption of i.i.d. ε_t, paragraph 4 of Ango Nze and Doukhan (2004) provides an overview of fading memory properties that are available. However, it is not easy to see how one could possibly extend these results based on Markov chain results to a setting where ε_t satisfies a condition weaker than the i.i.d. condition.

A third approach is based on linearity assumptions and a vector-valued x_t that satisfies $x_t = A_t x_{t-1} + B_t$ for an (A_t, B_t) that satisfies an i.i.d. or strict stationarity assumption; see Bougerol and Picard (1992) for such an approach. The results of Bougerol and Picard can for example be used to derive the strict stationarity of the GARCH(p,q) model in the presence of strictly stationary innovations. However, the linearity assumption is obviously restrictive.

This paper aims to carefully investigate the strict stationarity of x_t defined in Equation (1), under as general as possible conditions on $g(\cdot)$, while avoiding the i.i.d. assumption on ε_t. I pursue a novel and primitive approach to establishing my results and I do not use a Markov chain approach. The use of this primitive approach highlights the essential aspects of the strict stationarity proof while avoiding conditions that may be needed in the Markov chain approach, but are not essential for establishing strict stationarity only. In addition, it is my hope that the new methodology of proof that I develop in this paper can be easily adapted and extended to other stationarity issues. The results are applied to derive the strict stationarity of a class of threshold unit root models.

Section 2 outlines the main results of this paper for the special and simpler case of nondecreasing response functions. In Section 3, I derive results for more general response functions that satisfy a bounded variation assumption, at the expense of conditions that are stronger than those of Section 2. Section 4 considers threshold autoregressive models. All proofs have been gathered in the Appendix.

2 Main results: nondecreasing response functions

Below, I consider the model $z_t = g(z_{t-1} + \varepsilon_t)$ for strictly stationary ε_t and nondecreasing functions $g(\cdot)$, and I construct a strictly stationary solution y_t for this model. Such a result also implies strict stationarity of $x_t = g(x_{t-1}) + u_t$. To see this, assume that $y_t = f(u_{t-1}, u_{t-2}, \ldots)$ is a strictly stationary solution of the model $z_t = g(z_{t-1} + u_{t-1})$ and assume u_t is strictly stationary. Then $y_t + u_t$ is obviously strictly stationary as well and forms a solution to the model $x_t = g(x_{t-1}) + u_t$. Therefore, showing the existence of a strictly stationary solution for the model $z_t = g(z_{t-1} + u_{t-1})$ implies the existence of a strictly stationarity solution for

$x_t = g(x_{t-1}) + u_t$.

The reason for needing the assumption that $g(\cdot)$ is nondecreasing is that in the proof, lower and nondecreasing approximations $g^K(\cdot)$ to $g(\cdot)$ are used. In spite of my attempts, I was unable to construct an analogue of the proof of this section for increasing approximations or for the case of nonincreasing functions.

The analysis in this paper is based on the assumption of strict stationarity of innovations:

Assumption 1 *ε_t is a strictly stationary sequence of random variables.*

I will need the following continuity assumption for the function $g(\cdot)$.

Assumption 2 *Either*

1. *$g(\cdot)$ is continuous on $\mathbb{R}$; or*

2. *$g(\cdot) = c(x) + \sum_{j=1}^{N} a_j I(b_j < x \leq b_{j+1})$ for some function $c(\cdot)$ that is continuous on $\mathbb{R}$ and constants a_j and b_j such that $b_j \leq b_{j+1}$, and the distribution of $\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}$ is continuous on $\mathbb{R}$.*

Assumption 2 provides a tradeoff between a continuity assumption on the one hand that allows for an arbitrary distribution for ε_t and an assumption that allows for a finite number of jumps in the response function at the expense of a continuity assumption on the conditional distribution of $\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}$.

In this section, the following assumption will be used, for which alternatives will be developed later:

Assumption 3.1 *$g(\cdot)$ is nondecreasing on $\mathbb{R}$.*

For my first result for cases where $g(\cdot)$ has a lower bound, we will also need the following assumption:

Assumption 4.1 *For all $x \in \mathbb{R}$,*

$$E \prod_{l=0}^{m} (1 - I(\varepsilon_{t-l} > x)) \to 0 \quad as \quad m \to \infty. \tag{4}$$

The next assumption contains the lower bound assumption, and it also allows for the derivation of a strictly stationary upper bound for y_t:

Assumption 5.1 *$g(\cdot) \geq B$ and $g(x) \leq a + b\max(0, x)$ for constants $a > 0$ and $b \in [0, 1)$. In addition, either $b = 0$, or for some $\eta > 0$, $E|\max(0, \varepsilon_t)|^\eta < \infty$.*

The first stationarity result of this paper, for $g(\cdot)$ that are nondecreasing and possess a lower bound, is as follows:

Theorem 1 *Assume that Assumptions 1, 2, 3.1, 4.1 and 5.1 hold. Then there exists a strictly stationary solution y_t to the model $z_t = g(z_{t-1} + \varepsilon_t)$, and for any bounded function $D(\cdot)$ that is continuous on $\mathbb{R}$,*

$$\lim_{m\to\infty} \| D(y_t) - E(D(y_t)|\varepsilon_t, \dots, \varepsilon_{t-m}) \|_2 = 0. \quad (5)$$

To verify the key weak dependence condition of the above theorem under a more primitive condition, we can use the following alternatives to Assumptions 4.1:

Assumption 4.1' *ε_t is strong mixing and $P(\varepsilon_t > x) > 0$ for all $x \in \mathbb{R}$.*

A different type of primitive assumption can also be shown to imply Assumption 4.1:

Assumption 4.1" *For some function $H(\cdot)$, $P(\varepsilon_t > x|\varepsilon_{t-1}, \varepsilon_{t-2}, \dots) \geq H(x) > 0$ for all $x \in \mathbb{R}$.*

The formal proof that Assumptions 4.1' and 4.1" suffice can be stated as follows:

Lemma 1 *If Assumption 4.1' or 4.1" holds, then Assumption 4.1 holds.*

Assumptions 4.1' and 4.1" use a type of unbounded support assumption. It is hard to see how an assumption of this type can be avoided for the verification of Assumption 4.1 from primitive conditions. However, in the case of a simple linear AR(1) model, it is clear that the unbounded support assumption is not needed for strict stationarity, which suggests that perhaps in some cases, the unbounded support assumption can be relaxed.

The conclusion that y_t is strictly stationary and satisfies

$$\lim_{m\to\infty} \| D(y_t) - E(D(y_t)|\varepsilon_t, \dots, \varepsilon_{t-m}) \|_2 = 0 \quad (6)$$

for any continuous bounded function implies that functions of y_t satisfy weak laws of large numbers if ε_t is strong mixing. This condition is known in the econometrics literature as the condition of L_2-near epoch dependence of $D(y_t)$; see Andrews (1988) and Pötscher and Prucha (1997) for more information on the L_2-near epoch dependence concept. The formal result is the following:

Theorem 2 *Let y_t be a sequence of random variables that is strictly stationary and satisfies*

$$\lim_{m\to\infty} \| D(y_t) - E(D(y_t)|\varepsilon_t, \ldots, \varepsilon_{t-m}) \|_2 = 0 \tag{7}$$

for any continuous and bounded function $D(\cdot)$ and a sequence of random variables ε_t that is strong mixing. If $f(\cdot)$ is a Borel measurable function and $E|f(y_t)| < \infty$, then

$$n^{-1} \sum_{t=1}^{n} (f(y_t) - Ef(y_t)) \xrightarrow{p} 0. \tag{8}$$

Note that Theorem 1 asserts that a solution to the model $z_t = g(z_{t-1} + \varepsilon_t)$ exists, but it does not assert that this solution is unique. In the linear case considered in Bougerol and Picard (1992), the authors were able to assert uniqueness of the solution, but the nonlinear setting considered here complicates the situation substantially.

For the case when no upper or lower bound for $g(\cdot)$ is available, some stronger assumptions are needed. I needed the following alternative assumption to Assumption 4.1:

Assumption 4.2 *ε_t is strong mixing, for some function $H(\cdot)$,*

$$P(\varepsilon_t > x|\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) \geq H(x) > 0, \tag{9}$$

and either $E|\varepsilon_t|^2 < \infty$, or for some $\eta > 0$, $E|\varepsilon_t|^\eta < \infty$ and the distribution of $\varepsilon_t|\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$ is continuous on $\mathbb{R}$.

The following assumption, that does not require a lower bound for $g(\cdot)$, can be used instead of Assumption 5.1:

Assumption 5.2 *$|g(x)| \leq a + b|x|$ for constants $a < 0$ and $b \in [0, 1)$. In addition, for some $\eta > 0$, $E|\varepsilon_t|^\eta < \infty$.*

A general result for nondecreasing response functions that does not require the assumption of boundedness from below is the following:

Theorem 3 *Assume that Assumptions 1, 2, 3.1, 4.2 and 5.2 hold. Then there exists a strictly stationary solution y_t to the model $z_t = g(z_{t-1} + \varepsilon_t)$, and for any bounded function $D(\cdot)$ that is continuous on $\mathbb{R}$,*

$$\lim_{m\to\infty} \| D(y_t) - E(D(y_t)|\varepsilon_t, \ldots, \varepsilon_{t-m}) \|_2 = 0. \tag{10}$$

3 Main results: functions of bounded variation

It is possible to relax the nondecreasing assumption that was made earlier. In the proofs for this section, we will use either one of the following assumptions. Remember that a function is called monotone if it is either nondecreasing or nonincreasing.

Assumption 3.2 *$g(\cdot) = g_1(\cdot) - g_2(\cdot)$, where $g_1(\cdot)$ and $g_2(\cdot)$ are nondecreasing, and $g(\cdot)$ is monotone for $x \geq C$.*

Assumption 3.3 *$g(\cdot) = g_1(\cdot) - g_2(\cdot)$, where $g_1(\cdot)$ and $g_2(\cdot)$ are nondecreasing, and $g(\cdot)$ is monotone for $x \leq -C$.*

The added generality of dropping the nondecreasing assumption will need to come at the price of either one of the assumptions below, which are slightly stronger than Assumption 4.1:

Assumption 4.3 *For all $x \in \mathbb{R}$,*

$$E \prod_{l=0}^{m} (1 - I(\varepsilon_{t-l-1} > x) I(\varepsilon_{t-l} \leq -x)) \to 0 \quad as \quad m \to \infty. \tag{11}$$

Assumption 4.4 *For all $x \in \mathbb{R}$,*

$$E \prod_{l=0}^{m} (1 - I(\varepsilon_{t-l-1} \leq -x) I(\varepsilon_{t-l} > x)) \to 0 \quad as \quad m \to \infty. \tag{12}$$

In establishing an analogue of Theorem 1 and/or Theorem 3, I was unable to use a one-sided boundedness assumption. Instead, a total boundedness assumption is needed:

Assumption 5.3 *$|g(\cdot)| \leq B$.*

My first stationarity result for functions of bounded variation is the following:

Theorem 4 *Assume that Assumptions 1, 2, 5.3, and either 3.2 and 4.3, or 3.3 and 4.4 hold. Then there exists a strictly stationary solution y_t to the model $z_t = g(z_{t-1} + \varepsilon_t)$, and for any bounded function $D(\cdot)$ that is continuous on $\mathbb{R}$,*

$$\lim_{m \to \infty} \| D(y_t) - E(D(y_t) | \varepsilon_t, \ldots, \varepsilon_{t-m}) \|_2 = 0. \tag{13}$$

Similarly to the results of the previous section, we can use more primitive conditions to verify Assumptions 4.3 and 4.4:

Assumption 4.3' *ε_t is strong mixing and $P(\varepsilon_{t-1} > x, \varepsilon_t \leq -x) > 0$ for all $x \in \mathbb{R}$.*

Assumption 4.4' *ε_t is strong mixing and $P(\varepsilon_{t-1} \leq -x, \varepsilon_t > x) > 0$ for all $x \in \mathbb{R}$.*

Assumption 4.3" *For some function $H(\cdot)$, $P(\varepsilon_{t-1} > x, \varepsilon_t \leq -x | \varepsilon_{t-2}, \varepsilon_{t-3}, \ldots) \geq H(x) > 0$ for all $x \in \mathbb{R}$.*

Assumption 4.4" *For some function $H(\cdot)$, $P(\varepsilon_{t-1} \leq -x, \varepsilon_{t-1} > x | \varepsilon_{t-2}, \varepsilon_{t-3}, \ldots) \geq H(x) > 0$ for all $x \in \mathbb{R}$.*

The statement that Assumptions 4.3', 4.3", 4.4' and 4.4" suffice is the following. The proof is analogous to that of Lemma 1 and is therefore omitted.

Lemma 2 *If Assumption 4.3' or 4.3" holds, then Assumption 4.3 holds. If Assumption 4.4' or 4.4" holds, then Assumption 4.4 holds.*

It is also possible to relax the boundedness assumption of Assumption 5.3, and use Assumption 5.2 instead, at the expense of further regularity conditions:

Assumption 4.5 *ε_t is strong mixing, for some function $H(\cdot)$,*

$$P(\varepsilon_{t-1} > y, \varepsilon_t + |\varepsilon_{t-1}| \leq x | \varepsilon_{t-2}, \varepsilon_{t-3}, \ldots) = H(x, y) > 0, \tag{14}$$

and either $E|\varepsilon_t|^2 < \infty$, or for some $\eta > 0$, $E|\varepsilon_t|^\eta < \infty$ and the distribution of $\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$ is continuous on $\mathbb{R}$.

Assumption 4.6 *ε_t is strong mixing, for some function $H(\cdot)$,*

$$P(\varepsilon_{t-1} \leq -y, \varepsilon_t + |\varepsilon_{t-1}| > x | \varepsilon_{t-2}, \varepsilon_{t-3}, \ldots) = H(x, y) > 0, \tag{15}$$

and either $E|\varepsilon_t|^2 < \infty$, or for some $\eta > 0$, $E|\varepsilon_t|^\eta < \infty$ and the distribution of $\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$ is continuous on $\mathbb{R}$.

Theorem 5 *Assume that Assumptions 1, 2, 5.2, and either 3.2 and 4.5, or 3.3 and 4.6 hold. Then there exists a strictly stationary solution y_t to the model $z_t = g(z_{t-1} + \varepsilon_t)$, and for any bounded function $D(\cdot)$ that is continuous on $\mathbb{R}$,*

$$\lim_{m \to \infty} \| D(y_t) - E(D(y_t) | \varepsilon_t, \ldots, \varepsilon_{t-m}) \|_2 = 0. \tag{16}$$

4 Threshold autoregressive models

In this section, I apply the results of the previous section to the three-regime threshold unit root model

$$z_t = (\alpha_1 + \rho_1 z_{t-1})I(z_{t-1} \leq \tau_1) + (\alpha_2 + \rho_2 z_{t-1})I(z_{t-1} > \tau_2) + z_{t-1}I(\tau_1 \leq z_{t-1} \leq \tau_2) + \varepsilon_t \quad (17)$$

where $\tau_2 > \tau_1$ for the case of weakly dependent ε_t. The case $\alpha_1 + \rho_1\tau_1 = \tau_1$ and $\alpha_2 + \rho_2\tau_2 = \tau_2$ corresponds to that of a continuous response function. Seo (2004) used a model of this type in a unit root testing context, and made the explicit assumption that ε_t is a possibly weak dependent process. For the case of i.i.d. ε_t, Chan, Petrucelli , Tong and Woolford (1985) analyzed stationarity of the above model and they showed that stationarity of the above model can be achieved for a set of parameter values that is larger than that for which simply $|\rho_1| < 1$ and $|\rho_2| < 1$. Berben and van Dijk (1999) and Enders and Granger (1998) considered two-regime TAR models with a potentially nonstationary switching variable. Bec, Guay and Guerre (2002) consider the above model for the case $\alpha_1 = -\alpha_2$ and $\rho_1 = -\rho_2$ in the setting of unit root testing. Caner and Hansen (2001) considered a model similar to the above, but the regime switching variable is stationary in their setting.

The first corollary to Theorem 3 concerns the case of three-regime threshold unit root models with a continuous and nondecreasing response function:

Corollary 1 *Assume that Assumptions 1 and 4.2 hold. Also assume that* $0 \leq \rho_1 < 1$ *and* $0 \leq \rho_2 < 1$*, and assume that* $\alpha_1 + \rho_1\tau_1 = \tau_1$ *and* $\alpha_2 + \rho_2\tau_2 = \tau_2$*. Then there exists a strictly stationary solution* y_t *to the model of Equation (17), and for any bounded function* $D(\cdot)$ *that is continuous on* $\mathbb{R}$*,*

$$\lim_{m\to\infty} \| D(y_t) - E(D(y_t)|\varepsilon_t, \ldots, \varepsilon_{t-m}) \|_2 = 0. \quad (18)$$

The conditions $0 \leq \rho_1 < 1$, $0 \leq \rho_2 < 1$, and $\alpha_1 + \rho_1\tau_1 = \tau_1$ and $\alpha_2 + \rho_2\tau_2 = \tau_2$ ensure that the response function is increasing and continuous, which will make that Theorem 3 directly applies. For the more general case where the response function is not necessarily continuous and nondecreasing, the following corollary to Theorem 5 applies:

Corollary 2 *Assume that Assumptions 1 and either 3.2 and 4.5, or 3.3 and 4.6 hold. Also assume that* $|\rho_1| < 1$ *and* $|\rho_2| < 1$*, and assume that either* $\alpha_1 + \rho_1\tau_1 = \tau_1$ *and* $\alpha_2 + \rho_2\tau_2 = \tau_2$*, or that the distribution of* $\varepsilon_t|\varepsilon_{t-1}, \ldots$ *is continuous. Then there exists a strictly stationary solution* y_t *to the model of Equation (17), and for any bounded function* $D(\cdot)$ *that is continuous on* $\mathbb{R}$*,*

$$\lim_{m\to\infty} \| D(y_t) - E(D(y_t)|\varepsilon_t, \ldots, \varepsilon_{t-m}) \|_2 = 0. \quad (19)$$

References

Andrews, D.W.K. (1988), Laws of large numbers for dependent non-identically distributed random variables, *Econometric Theory* 4, 458-467.

Ango Nze, P. and Doukhan (2004), Weak dependence: models and applications to econometrics, *Econometric Theory* 20, 995-1045.

Bec, F., Guay, A. and E. Guerre (2002), Adaptive consistent unit Root tests based on autoregressive threshold model, working paper.

Berben, R. P. and D. van Dijk, Unit root tests and asymmetric adjustment: a reassessment, Research Report EI-9902/A, Econometric Institute 1999.

Bierens, H.J. (1981) *Robust methods and asymptotic theory in nonlinear econometrics.* New York: Springer-Verlag.

Bougerol, P. and N. Picard (1992), Strict stationarity of generalized autoregressive processes, *Annals of Probability* 20, 1714-1730.

Chan, K.S., J.D. Petrucelli, H.Tong, and S. Woolford (1985), A Multiple-threshold AR(1) Model, *Journal of Applied Probability* 22, 267-279.

Davidson, J. (1994) *Stochastic Limit Theory.* Oxford: Oxford University Press.

Enders, W. and C.W.J. Granger (1998), Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates, *Journal of Business and Economic Statistics*, 16, 30411.

Kapetanios, G. and Y. Shin (2001), Unit root tests in three-regime SETAR models, mimeo, Queen Mary, University of London.

Meyn, S.P. and R.L. Tweedie (1993) *Markov chains and stochastic stability.* London: Springer-Verlag.

Tong, H. (1990) *Non-linear time series: a dynamical system approach.* Oxford: Oxford University Press.

Seo, M. (2004), *Unit root test in a threshold autoregression: asymptotic theory and residual-based block bootstrap*, Working paper, University of Wisconsin-Madison.

Pötscher, B.M. and I.R. Prucha (1997) *Dynamic nonlinear econometric models.* Berlin: Springer-Verlag.

Mathematical Appendix

The following lemma serves as a basis for the derivation of my results:

Lemma 3 *If y_t^K is a strictly stationary sequence of random variables for all $K > 0$, and if $y_t^K \xrightarrow{as} y_t$ as $K \to \infty$, then y_t is a strictly stationary sequence of random variables.*

Proof of Lemma 3:

By assumption, because y_t^K is strictly stationary,

$$E\exp(i\sum_{j=0}^{m}\xi_j y_{t-j}) = \lim_{K\to\infty} E\exp(i\sum_{j=0}^{m}\xi_j y_{t-j}^K) = \psi_K(\xi_1,\ldots,\xi_m),$$

and because the latter characteristic function does not depend on t and the pointwise limit as $K \to \infty$ of $\psi_K(\cdot)$ is well-defined as the characteristic function of $(y_t,\ldots,y_{t-m})$, it follows that the distribution of $(y_t,\ldots,y_{t-m})$ does not depend on t. □

Below, the strategy of proof will be to find strictly stationary approximate solutions to the model $z_t = g(z_{t-1}+\varepsilon_t)$. For all $K > 0$ consider the model $z_t^K = g^K(z_{t-1}^K + \varepsilon_t)$ where

$$g^K(x) = g(x)I(x \le K) + g(K)I(x > K).$$

The idea of the proof is to show that these models have strictly stationary solutions y_t^K, and that $\lim_{K\to\infty} y_t^K$ is a strictly stationary solution of the model $z_t = g(z_{t-1}+\varepsilon_t)$.

The existence of strictly stationary solutions y_t^K that satisfy a limit property is proven in the following lemma.

Lemma 4 *Assume that Assumptions 1, 3.1, 4.1 and 5.1 hold. Then there exists a strictly stationary solution y_t^K to the model $z_t^K = g^K(z_{t-1}^K+\varepsilon_t)$, and $\lim_{K\to\infty} y_t^K$ is a well-defined strictly stationary sequence of random variables.*

Proof:

I consider the approximation $\hat{y}_t^{K,m}$ that would result from ignoring the ε_{t-m-j} for all $j \ge 1$ in a recursive substitution definition. Define $g_{0t}^K(x) = g^K(\varepsilon_t + x)$, $g_{1t}^K(x) = g^K(\varepsilon_t + g^K(\varepsilon_{t-1}+x))$, etc., and define $\hat{y}_t^{K,m} = g_{mt}^K(0)$; this definition implies that $\hat{y}_t^{K,m} = g^K(\hat{y}_{t-1}^{K,m-1} + \varepsilon_t)$ for $m \ge 1$. Then $\hat{y}_t^{K,m} \ge B = h_t^L$ and under Assumption 5.1,

$$\hat{y}_t^{K,m} \le \sum_{j=0}^{m} ab^j + b\sum_{j=0}^{m}\max(0,\varepsilon_{t-j})b^j \le a/(1-b) + b\sum_{j=0}^{\infty}\max(0,\varepsilon_{t-j})b^j = h_t^U,$$

and note that $\sum_{j=0}^{\infty} \max(0, \varepsilon_{t-j}) b^j$ is well-defined by the Cauchy criterion because for $\eta \in (0, 1]$, by Loève's c_η inequality (see for example Davidson (1994, p. 140)),

$$E \max_{n \geq m} | \sum_{j=0}^{n} \max(0, \varepsilon_{t-j}) b^j - \sum_{j=0}^{m} \max(0, \varepsilon_{t-j}) b^j |^\eta$$

$$\leq E \sum_{j=m}^{\infty} | \max(0, \varepsilon_{t-j}) b^j |^\eta \leq E | \max(0, \varepsilon_{t-j}) |^\eta \sum_{j=m}^{\infty} b^{\eta j} < \infty \to 0$$

as $m \to \infty$. The case $b = 0$ gives a trivial upper bound without moment condition on ε_t. Note that for $m \geq 0$, $\hat{y}_t^{K,m} = \hat{y}_t^{K,m}(\varepsilon_t, \ldots, \varepsilon_{t-m})$, and also note that $\hat{y}_t^{K,m} = \hat{y}_t^{K,n}$ for all $n \geq m$ if for some $j \in \{0, 1, \ldots, m\}$, $\varepsilon_{t-j} + h_{t-j-1}^L > K$. Therefore,

$$P(\exists n \geq m : \hat{y}_t^{K,m} \neq \hat{y}_t^{K,n}) \leq P(\forall j \in \{0, 1, \ldots, m\} : \varepsilon_{t-j} + h_{t-j}^L \leq K) \to 0$$

as $m \to \infty$ by assumption because of Assumption 4.1 and because $h_{t-j}^L \geq B$, implying that $\hat{y}_t^{K,m}$ converges a.s. by the Cauchy criterion as $m \to \infty$. Define $y_t^K = \lim_{m \to \infty} \hat{y}_t^{K,m}$. For every m and K, $\hat{y}_t^{K,m}$ is a function of $(\varepsilon_t, \ldots, \varepsilon_{t-m})$ and is therefore strictly stationary. Therefore by Lemma 3, y_t^K is strictly stationary as well. Because a solution to the model $z_t^K = g^K(z_{t-1}^K + \varepsilon_t)$ is given by $\hat{y}_t^{K,m}$ for some large enough value of m, it follows that y_t^K is a strictly stationary solution to the model $z_t^K = g^K(z_{t-1}^K + \varepsilon_t)$.

To show that $\lim_{K \to \infty} y_t^K$ exists, I first show that $\hat{y}_t^{K,m}$ is nondecreasing in K for every m, from which it follows that y_t^K is nondecreasing in K as well. This can be seen as follows. Because $g_{mt}^K(x)$ is nondecreasing in K for every $x \in \mathbb{R}$ and m, $\hat{y}_t^{K,m}$ satisfies, for all $n \geq 0$,

$$\hat{y}_t^{K,m} = g_{mt}^K(0) \leq g_{mt}^{K+n}(0) = \hat{y}_t^{K+n,m}.$$

Because $y_t^K \leq h_t^U$ and y_t^K is nondecreasing, y_t^K must possess an almost sure limit as $K \to \infty$; by Lemma 3, $\lim_{K \to \infty} y_t^K$ is a strictly stationary sequence of random variables. □

The following lemma proves the existence of a solution to the model $z_t = g(z_{t-1} + \varepsilon_t)$ under some conditions:

Lemma 5 *Under Assumptions 1, 2, 3.1, 4.1 and 5.1,*

1. $y_t = \lim_{K \to \infty} y_t^K$ *is a strictly stationary solution to the model* $z_t = g(z_{t-1} + \varepsilon_t)$*;*

2. *For any bounded function* $D(\cdot)$ *that is continuous on* $\mathbb{R}$*,*

$$\lim_{m \to \infty} \| D(y_t) - E(D(y_t) | \varepsilon_t, \ldots, \varepsilon_{t-m}) \|_2 = 0.$$

Proof:

By Lemma 4, $\lim_{K\to\infty} y_t^K = y_t$ is a well-defined random variable and therefore,

$$\Phi(y_t) = \lim_{K\to\infty} \Phi(y_t^K) = \lim_{K\to\infty} \Phi(g^K(y_{t-1}^K + \varepsilon_t)),$$

where $\Phi(\cdot)$ denoted the standardnormal distribution function. Also, because $\Phi(g(y_{t-1}^K + \varepsilon_t))$ is nondecreasing in K and bounded, $\lim_{K\to\infty} \Phi(g(y_{t-1}^K + \varepsilon_t))$ is well-defined. By dominated convergence and the above relationship, we also have

$$\begin{aligned}
& E|\Phi(y_t) - \lim_{K\to\infty} \Phi(g(y_{t-1}^K + \varepsilon_t))| \\
& \leq E \lim_{K\to\infty} |\Phi(g^K(y_{t-1}^K + \varepsilon_t)) - \Phi(g(y_{t-1}^K + \varepsilon_t))| \\
& = \lim_{K\to\infty} E|\Phi(g^K(y_{t-1}^K + \varepsilon_t)) - \Phi(g(y_{t-1}^K + \varepsilon_t))| \\
& \leq \lim_{K\to\infty} P(y_{t-1}^K + \varepsilon_t > K) \\
& \leq \lim_{K\to\infty} P(h_{t-1}^U + \varepsilon_t > K) = 0,
\end{aligned}$$

where h_t^U is as in Lemma 4. By continuity of $\Phi(\cdot)$ and the above result, if we can show that

$$\lim_{K\to\infty} E|\Phi(g(y_{t-1}^K + \varepsilon_t)) - \Phi(g(\lim_{K\to\infty} y_{t-1}^K + \varepsilon_t))| = 0, \tag{20}$$

we have

$$\begin{aligned}
\Phi(y_t) &= \lim_{K\to\infty} \Phi(g(y_{t-1}^K + \varepsilon_t)) \\
&= \Phi(g(\lim_{K\to\infty} y_{t-1}^K + \varepsilon_t)) = \Phi(g(y_{t-1} + \varepsilon_t)),
\end{aligned}$$

implying that y_t is a strictly stationary solution to the model $z_t = g(z_{t-1} + \varepsilon_t)$ because $h_t^L \leq y_t \leq h_t^U$ for a well-defined random variable (h_t^L, h_t^U). Therefore, the proof is complete with a demonstration of the result of Equation (20). By Assumption 2 we can write

$$g(x) = c(x) + \sum_{j=1}^{N} a_j I(b_j \leq x \leq b_{j+1});$$

therefore

$$\lim_{K\to\infty} E|\Phi(g(y_{t-1}^K + \varepsilon_t)) - \Phi(g(y_{t-1} + \varepsilon_t))|$$

$$\leq \lim_{K\to\infty} \sum_{j=1}^{N} E|I(b_j < y_{t-1}^K + \varepsilon_t \leq b_{j+1})\Phi(c(y_{t-1}^K + \varepsilon_t) + a_j)$$

$$-I(b_j < \lim_{K\to\infty} y_{t-1}^K + \varepsilon_t \leq b_{j+1})\Phi(c(\lim_{K\to\infty} y_{t-1}^K + \varepsilon_t) + a_j)|$$

$$\leq \lim_{K\to\infty} \sum_{j=1}^{N} E|\Phi(c(y_{t-1}^K + \varepsilon_t) + a_j) - \Phi(c(\lim_{K\to\infty} y_{t-1}^K + \varepsilon_t) + a_j)|$$

$$+ \lim_{K\to\infty} \sum_{j=1}^{N} E|I(b_j < y_{t-1}^K + \varepsilon_t \leq b_{j+1}) - I(b_j < \lim_{K\to\infty} y_{t-1}^K + \varepsilon_t \leq b_{j+1})|,$$

and the first term is obviously 0 by continuity of $c(\cdot)$ and $\Phi(\cdot)$ and dominated convergence. Under the first part of Assumption 2, the second term is zero because we can set $g(\cdot) = c(\cdot)$; this shows that the first part of Assumption 2 suffices for the result. To show that the second part of Assumption 2 also suffices, note that because $y_t^K \leq \lim_{K\to\infty} y_t^K$, the second term equals zero as well because by dominated convergence,

$$\lim_{K\to\infty} \sum_{j=1}^{N} E|I(b_j < y_{t-1}^K + \varepsilon_t \leq b_{j+1}) - I(b_j < \lim_{K\to\infty} y_{t-1}^K + \varepsilon_t \leq b_{j+1})|$$

$$\leq E \lim_{K\to\infty} \sum_{j=1}^{N} F_{\varepsilon_t|\varepsilon_{t-1},\ldots}(b_{j+1} - y_{t-1}^K) - F_{\varepsilon_t|\varepsilon_{t-1},\ldots}(b_{j+1} - \lim_{K\to\infty} y_{t-1}^K)$$

$$+E \lim_{K\to\infty} \sum_{j=1}^{N} F_{\varepsilon_t|\varepsilon_{t-1},\ldots}(b_j - y_{t-1}^K) - F_{\varepsilon_t|\varepsilon_{t-1},\ldots}(b_j - \lim_{K\to\infty} y_{t-1}^K).$$

By the presumed continuity of $F_{\varepsilon_t|\varepsilon_{t-1},\ldots}(y)$, the result now follows.

To show the second assertion of the theorem, note that because the conditional expectation is the best possible L_2-approximation,

$$\| D(y_t) - E(D(y_t)|\varepsilon_t, \ldots, \varepsilon_{t-m}) \|_2$$
$$\leq \| D(y_t) - E(D(y_t^K)|\varepsilon_t, \ldots, \varepsilon_{t-m}) \|_2$$
$$\leq 2 \| D(y_t) - D(y_t^K) \|_2 + \| D(y_t^K) - E(D(y_t^K)|\varepsilon_t, \ldots, \varepsilon_{t-m}) \|_2$$
$$\leq 2 \| D(y_t) - D(y_t^K) \|_2 + \| D(y_t^K) - D(\hat{y}_t^{K,m}) \|_2 .$$

Because $y_t^K \xrightarrow{as} y_t$, $\hat{y}_t^{K,m} \xrightarrow{as} y_t^K$ as $m \to \infty$ and boundedness and continuity of $D(\cdot)$, the result now follows from the dominated convergence theorem by taking the limits first as $m \to \infty$ and then as $K \to \infty$. □

Proof of Theorem 1:

This theorem follows directly from Lemma 4 and Lemma 5. □

Proof of Lemma 1:

I will first show that Assumption 4.1' suffices for Assumption 4.1 to hold. Let r_m and b_m integer-valued sequences such that $r_m \to \infty$, $b_m \to \infty$, and $r_m b_m \leq m$. Then because $|E(X_t X_{t+m}) - E(X_t)E(X_{t+m})| \leq \alpha(m)$ if X_t is strong mixing and $|X_t| \leq 1$ and $|X_{t+m}| \leq 1$,

$$\lim_{m\to\infty} E \prod_{l=0}^{m} (1 - I(\varepsilon_{t-l} > x))$$

$$\leq \lim_{m\to\infty} E \prod_{k=1}^{r_m} \prod_{i=(k-1)b_m+1}^{kb_m} (1 - I(\varepsilon_{t-i} > x))$$

$$\leq \lim_{m\to\infty} E \prod_{k=1}^{r_m} (1 - I(\varepsilon_{t-kb_m} > x))$$

$$\leq \lim_{m\to\infty} (\alpha(b_m) + (1 - P(\varepsilon_t > x))E \prod_{k=2}^{r_m} (1 - I(\varepsilon_{t-kb_m} > x))),$$

and by repeating this reasoning we find the upper bound

$$\lim_{m\to\infty} \alpha(b_m)/P(\varepsilon_t > x),$$

and because $\alpha(m) \to 0$ as $m \to \infty$ and $P(\varepsilon_t > x) > 0$ for all $x \in \mathbb{R}$ by assumption, it follows that the last expression equals 0. □

Proof of Theorem 2:

For any Borel measurable function $f(\cdot)$ there exists for any $\delta > 0$ a bounded and continuous function $f^\delta(\cdot)$ such that

$$E|f(y_t) - f^\delta(y_t)| < \delta.$$

Observing that

$$n^{-1} \sum_{t=1}^{n} (f(y_t) - Ef(y_t)) = n^{-1} \sum_{t=1}^{n} (f^\delta(y_t) - Ef^\delta(y_t))$$

$$+n^{-1}\sum_{t=1}^{n}(f(y_t)-f^{\delta}(y_t))+n^{-1}\sum_{t=1}^{n}(Ef^{\delta}(y_t)-Ef(y_t)),$$

obviously the absolute expectation of the last two terms can be bounded by 2δ. Therefore, it remains to show that the weak law of large numbers holds for the first term. Since $f^{\delta}(\cdot)$ is bounded and continuous, by assumption

$$\lim_{m\to\infty} \parallel f^{\delta}(y_t)-E(f^{\delta}(y_t)|\varepsilon_t,\ldots,\varepsilon_{t-m}) \parallel_2 = 0,$$

implying that $f^{\delta}(y_t)$ is L_2-near epoch dependent with respect to ε_t; see Andrews (1988) for definitions of L_2-near epoch dependent processes, L_2-mixingales, mixingale magnitude indices and mixingale numbers. Because $f^{\delta}(\cdot)$ is bounded, $f^{\delta}(y_t)$ is also an L_2-mixingale with uniformly bounded magnitude indices and mixingale numbers

$$\alpha(m)+ \parallel f^{\delta}(y_t)-E(f^{\delta}(y_t)|\varepsilon_t,\ldots,\varepsilon_{t-m}) \parallel_2 .$$

By the law of large numbers of Andrews (1988), the weak law of large numbers for the first term follows.

Proof of Theorem 3:

The proofs of Lemmas 4 and 5 can be completely copied for this case, except that we need to show that

$$E\prod_{l=0}^{m-1}(1-I(\varepsilon_{t-l}+h^L_{t-l-1}>K))\to 0,$$

because we do not assume the existence of a lower bound B for $g(\cdot)$ in Theorem 3. To identify a different lower bound h^L_t, observe that under Assumption 5.2, if $g(\cdot)$ is nondecreasing for $x>C$, then for $K>C$,

$$y^K_t \geq g^C(y^K_{t-1}+\varepsilon_t)\geq -|g(C)|-a-b|\varepsilon_t|-b|y_{t-1}|$$

$$\geq\ldots\geq -((a+|g(C)|)/(1-b)-\sum_{j=0}^{\infty}b^j|\varepsilon_{t-j}|=c_2-\sum_{j=0}^{\infty}b^j|\varepsilon_{t-j}|=h^L_t,$$

where similarly to the proof of Lemma 4, it can be shown that h^L_t is well-defined by Loéve's c_η inequality. Therefore,

$$E\prod_{l=0}^{m-1}(1-I(\varepsilon_{t-l}+h^L_{t-l-1}>K))$$

$$= E\exp(\sum_{l=0}^{m-1}\log(1 - I(\varepsilon_{t-l} + h^L_{t-l-1} > K)))$$

$$\leq \exp(-m) + P(m^{-1}\sum_{l=0}^{m-1}\log(1 - I(\varepsilon_{t-l} + h^L_{t-l-1} > K)) > -1).$$

I will show that the last probability converges to 0 as $m \to \infty$ for any $K > 0$. To do this, define

$$I_{t-l,K} = I(\varepsilon_{t-l} + h^L_{t-l-1} > K)$$

and

$$I^1_{t-l,K} = H(\varepsilon_{t-l} + h^L_{t-l-1} - K)$$

for

$$H(x) = xI(0 \leq x \leq 1) + I(x > 1).$$

Note that $I_{t-l,K} \geq I^1_{t-l,K} \geq I_{t-l,K-1}$. Then for all $\delta \in (0, 1)$,

$$m^{-1}\sum_{l=0}^{m-1}\log(1 - I(\varepsilon_{t-l} + h^L_{t-l-1} > K))$$

$$= m^{-1}\sum_{l=0}^{m-1}\log(1 - I_{t-l,K})$$

$$\leq m^{-1}\sum_{l=0}^{m-1} I_{t-l,K}\log(\delta).$$

I will first consider the case of $E|\varepsilon_t|^2 < \infty$; the case of a continuous conditional distribution and $E|\varepsilon_t|^\eta < \infty$ will be tackled later. For the first case, we can bound the last expression by

$$m^{-1}\sum_{l=0}^{m-1}(I^1_{t-l,K} - EI^1_{t-l,K})\log(\delta) + EI^1_{t-l,K}\log(\delta).$$

Now

$$P(m^{-1}\sum_{l=0}^{m-1}(I^1_{t-l,K} - EI^1_{t-l,K})\log(\delta) + EI^1_{t-l,K}\log(\delta) > -1)$$

$$\leq P(|m^{-1}\sum_{l=0}^{m-1}(I^1_{t-l,K} - EI^1_{t-l,K})\log(\delta)| > 1/2) + I(EI^1_{t-l,K}\log(\delta) > -1/2). \quad (21)$$

The second term converges to 0 as $\delta \to 0$ because for any $K > 0$,

$$EI^1_{t-l,K} \geq EI_{t-l,K-1} = EI(\varepsilon_{t-l} + h^L_{t-l-1} > K-1) > 0. \quad (22)$$

The first term in Equation (21) now converges to zero by the weak law of large numbers, which will complete the proof. To show the result of Equation (22), note that

$$\varepsilon_t + h^L_{t-1} = \varepsilon_t + c_2 - \sum_{j=0}^{\infty} b^j|\varepsilon_{t-1-j}|,$$

and therefore it follows from our assumption that for some function $H(.,\ldots,.)$,

$$P(\varepsilon_t + h^L_{t-1} > y|\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) \geq H(y - h^L_{t-1}) > 0.$$

Therefore,

$$EI(\varepsilon_t + h^L_{t-1} > K-1)$$
$$= EP(\varepsilon_t > K - 1 - h^L_{t-1}|\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots)$$
$$\geq EH(K - 1 - h^L_{t-1}).$$

Now because $H(.) > 0$ and h^L_{t-1} is strictly stationary, it follows that for all $K > 0$, $EI^1_{t-l,K} > 0$. To show that the first term in Equation (21) satisfies a weak law of large numbers, note that

$$I^1_{t-l,K} = H(\varepsilon_{t-l} + h^L_{t-l-1} - K) = H(\varepsilon_{t-l} + c_2 - \sum_{i=0}^{\infty} b^i|\varepsilon_{t-l-1-i}| - K)$$

and for all t,

$$w_{t-l} = \varepsilon_{t-l} + c_2 - \sum_{i=0}^{\infty} b^i|\varepsilon_{t-l-1-i}| - K$$

is strictly stationary (with respect to l) and L_2-near epoch dependent on ε_{t-l}, and that $\nu(M)$ decays exponentially. This is because for $M \geq 1$, and all t, l and K,

$$\| w_{t-l} - E(w_{t-l}|\varepsilon_{t-l-M}, \ldots, \varepsilon_{t-l}) \|_2$$

$$\leq \sum_{i=M+1}^{\infty} b^i \parallel \varepsilon_{t-l-i} \parallel_2 = \parallel \varepsilon_t \parallel_2 \sum_{i=M+1}^{\infty} b^i,$$

and the last expression converges to 0 as $M \to \infty$ at exponential rate. Therefore, because $H(\cdot)$ is Lipschitz-continuous with a Lipschitz coefficient of 1,

$$H(\varepsilon_{t-l} + c_2 - \sum_{i=0}^{\infty} b^i |\varepsilon_{t-l-1-i}| - K)$$

is also L_2-near epoch dependent on ε_{t-l} with an exponentially decreasing $\nu(\cdot)$ sequence. See Pötscher and Prucha (1997) for more information about these manipulations with near epoch dependent processes. The result of this lemma then follows from the weak law of large numbers for L_1-mixingales of Andrews (1988), because the mixingale numbers $\psi(M)$ for

$$H(\varepsilon_{t-l} + c_2 - \sum_{i=0}^{\infty} b^i |\varepsilon_{t-l-1-i}| - K)$$

can be chosen to satisfy

$$\psi(M) \leq \sum_{i=M+1}^{\infty} b^i + \parallel \varepsilon_t \parallel_2 \alpha(M)^{1/2}$$

and the last expression converges to 0 as $M \to \infty$.

In the second case of a continuous conditional distribution and $E|\varepsilon|_t|^\eta < \infty$, I will show that a weak law of large holds for

$$m^{-1} \sum_{l=0}^{m-1} I_{t-l,K} \log(\delta),$$

and noting that it has been shown earlier that $EI_{t-l,K} > 0$, this also suffices to complete the proof. To show this weak law of large numbers, I verify that $EI_{t-l,K}$ is L_2-near epoch dependent. Note that because conditional expectations are the best L_2-approximations,

$$\parallel I_{t-l,K} - E(I_{t-l,K}|\varepsilon_{t-l-M}, \ldots, \varepsilon_{t-l}) \parallel_2$$

$$= \parallel I_{tK} - E(I_{t,K}|\varepsilon_{t-M}, \ldots, \varepsilon_t) \parallel_2$$

$$\leq \parallel I(\varepsilon_t + c_2 - \sum_{j=1}^{\infty} b^j |\varepsilon_{t-j}| > K) - I(\varepsilon_t + c_2 - \sum_{j=1}^{M} b^j |\varepsilon_{t-j}| > K) \parallel_2$$

$$\leq \| I(K + \sum_{j=1}^{\infty} b^j |\varepsilon_{t-j}| - \sum_{j=M+1}^{\infty} b^j |\varepsilon_{t-j}| \leq \varepsilon_t \leq K + \sum_{j=1}^{\infty} b^j |\varepsilon_{t-j}| + \sum_{j=M+1}^{\infty} b^j |\varepsilon_{t-j}|) \|_2$$

$$\leq E|F_{\varepsilon_t|\varepsilon_{t-1},\ldots}(K + \sum_{j=1}^{\infty} b^j |\varepsilon_{t-j}| - \sum_{j=M+1}^{\infty} b^j |\varepsilon_{t-j}|) - F_{\varepsilon_t|\varepsilon_{t-1},\ldots}(K + \sum_{j=1}^{\infty} b^j |\varepsilon_{t-j}| + \sum_{j=M+1}^{\infty} b^j |\varepsilon_{t-j}|)|,$$

and by dominated convergence, continuity of $F_{\varepsilon_t|\varepsilon_{t-1},\ldots}$ and because $\sum_{j=M+1}^{\infty} b^j |\varepsilon_{t-j}| \xrightarrow{as} 0$ as $M \to 0$, it now follows that $I_{t-l,K}$ is L_2-near epoch dependent. Therefore, the weak law of large numbers for L_1-mixingales of Andrews (1988) applies, which completes the proof. □

Proof of Theorem 4:

For the case of response functions $g(\cdot)$ that are of bounded variation, I need a construction that is slightly more complex than the one used earlier for nondecreasing $g(\cdot)$. The strategy of proof now will be to find two strictly stationary approximate solutions to the model $z_t = g(z_{t-1} + \varepsilon_t)$. First, for all $K > 0$ define

$$g_1^K(x) = g_1(x)I(x \leq K) + g_1(K)I(x > K),$$

$$g_2^K(x) = g_2(x)I(x \leq K) + g_2(K)I(x > K),$$

$$g_{1K}(x) = g_1(x)I(x > -K) + g_1(-K)I(x < -K),$$

and

$$g_{2K}(x) = g_1(x)I(x > -K) + g_2(-K)I(x < -K).$$

For all $K > 0$ and $M > 0$, consider the two-equation model

$$z_t^{KM} = g_1^K(z_{t-1}^{KM} + \varepsilon_t) - g_{2M}(z_{t-1,KM} + \varepsilon_t)$$

$$z_{t,KM} = g_{1K}(z_{t-1,KM} + \varepsilon_t) - g_2^M(z_{t-1}^{KM} + \varepsilon_t).$$

If $g(\cdot)$ is nondecreasing for $x > C$, we can presume that $g_2(x) = g_2(C)$ for $x > C$. This observation is the basis for the derivation of the strict stationarity property of the solutions that are constructed for the above models. In this proof, I will only consider the case where $g(\cdot)$ is nondecreasing for $x > C$; the other cases are proven analogously. Also, observe that for all $n > 0$, $g_1^K(\cdot) \leq g_1^{K+n}(\cdot)$, $g_2^K(\cdot) \leq g_2^{K+n}(\cdot)$, $g_{1K}(\cdot) \geq g_{1,K+n}(\cdot)$, and $g_{2K}(\cdot) \geq g_{2,K+n}(\cdot)$. Similarly to the earlier proof for nondecreasing $g(\cdot)$, the idea of the proof is here to show that the above two-equation models have strictly stationary solutions $(y_t^{KM}, y_{t,KM})$, and that a limit solution as K and M approach infinity provides a strictly stationary solution of the model $z_t = g(z_{t-1} + \varepsilon_t)$.

Consider the approximations $(\hat{y}_t^{KM,m}, \hat{y}_{t,KM}^m)$ that would result from ignoring the ε_{t-m-j} for all $j \geq 1$ in a recursive substitution definition, similarly to Lemma 4. If $\varepsilon_{t-j-1} + h_{t-j-2}^L > \max(C, K)$, then $\hat{y}_{t-j-1}^{KM,m-1} = g_1(K) - g_2(C)$ for $m \geq 2$. If in addition, $\varepsilon_{t-j} + h_{t-j-1}^U \leq -\max(M, K)$, then

$$\hat{y}_{t-j}^{KM,m} = g_1^K(\varepsilon_{t-j} + g_1(K) - g_2(C)) - g_2(-M)$$

$$\hat{y}_{t-j,KM}^m = g_1(-K) - g_2^M(\varepsilon_{t-j} + g_1(K) - g_2(C)).$$

Using the same reasoning as in Lemma 4, this reasoning implies that $\lim_{m\to\infty}(\hat{y}_t^{KM,m}, \hat{y}_{t,KM}^m) = (y_t^{KM}, y_{t,KM})$ exists if for all $K, M > 0$,

$$P(\exists 0 \leq j \leq m : \varepsilon_{t-j-1} + h_{t-j-2}^L > \max(C, K) \quad \text{and} \quad \varepsilon_{t-j} + h_{t-j-1}^U \leq -\max(K, M)) \to 1$$

as $m \to \infty$, which holds by assumption, given the boundness assumption $|g(\cdot)| \leq B$. Because $(\hat{y}_t^{KM,m}, \hat{y}_{t,KM}^m)$ is strictly stationary for all $m \geq 0$, it follows from Lemma 3 that $(y_t^{KM}, y_{t,KM})$ is a strictly stationary solution of the model.

To show that $\lim_{K\to\infty} \lim_{M\to\infty} y_t^{KM}$ exists, note that y_t^{KM} is nondecreasing in K and nondecreasing in M because $\hat{y}_t^{KM,m}$ is nondecreasing in K and M by construction. Under Assumption 5.4, $y_t^{KM} \leq h_t^U$ for some strictly stationary random sequence h_t^U. Therefore, $\lim_{M\to\infty} y_t^{KM}$ exists in the almost sure sense, and is nondecreasing in K. By Lemma 3, $\lim_{M\to\infty} y_t^{KM}$ is also a strictly stationary sequence of random variables. Therefore, again by Lemma 3, $\lim_{K\to\infty} \lim_{M\to\infty} y_t^{KM}$ exists in the almost sure sense and is also a strictly stationary sequence of random variables.

To show that $\lim_{K\to\infty} \lim_{M\to\infty} y_t^{KM}$ is a solution to the model $z_t = g(z_{t-1} + \varepsilon_t)$, we can follow exactly the reasoning of Lemma 5, which completes the proof. □

Proof of Theorem 5:

I will only consider the case of $g(\cdot)$ that is monotone for $x > C$ of Assumptions 3.2 and 4.5; the other case is analogous. The proof of Theorem 3 can be completely copied for this case, except that we need to show that

$$E \prod_{l=0}^{m-1} (1 - I(\varepsilon_{t-l-1} + h_{t-l-2}^L > K) I(\varepsilon_{t-l} + h_{t-l-1}^U \leq -K)) \to 0,$$

because we do not assume uniform boundedness of $g(\cdot)$ in this theorem. Note that if $g(\cdot)$ is monotone for $x > C$, then

$$|g^K(x)| \leq a + b|x| + |g(C)|,$$

implying that we can use

$$h_t^U = -h_t^L = (a + |g(C)|)/(1-b) + \sum_{j=0}^{\infty} b^j |\varepsilon_{t-j}|$$

which is well-defined because $E|\varepsilon_t|^\eta < \infty$. Similarly to the proof of Theorem 3, define

$$I_{t-l,K} = I(\varepsilon_{t-l} + h_{t-l-1}^L > K),$$

$$\tilde{I}_{t-l,K} = I(\varepsilon_{t-l} + h_{t-l-1}^U \leq -K),$$

and

$$I_{t-l,K}^1 = H(\varepsilon_{t-l} + h_{t-l-1}^L - K)$$

$$\tilde{I}_{t-l,K}^1 = H(\varepsilon_{t-l} + h_{t-l-1}^U + K)$$

for

$$H(x) = xI(0 \leq x \leq 1) + I(x > 1).$$

As before, first the case where $E|\varepsilon_t|^2 < \infty$ will be considered. Note that

$$E \prod_{l=0}^{m-1} (1 - I(\varepsilon_{t-l-1} + h_{t-l-2}^L > K) I(\varepsilon_{t-l} + h_{t-l-1}^U \leq K))$$

$$\leq \exp(-m) + P(m^{-1} \sum_{l=0}^{m-1} \log(1 - I_{t-1-l,K} \tilde{I}_{tlK}) > -1)$$

$$\leq P(|m^{-1} \sum_{l=0}^{m-1} (I_{t-1-l,K}^1 \tilde{I}_{tlK}^1 - EI_{t-1-l,K}^1 \tilde{I}_{t-l,K}^1) \log(\delta)| > 1/2) + I(EI_{t-1-l,K}^1 \tilde{I}_{t-l,K}^1 \log(\delta) > -1/2).$$

Also,

$$EI_{t-1-l,K}^1 \tilde{I}_{t-l,K}^1 \geq EI_{t-1-l,K-1} \tilde{I}_{t-l,K-1} > 0$$

because by assumption,

$$P[\varepsilon_{t-1} + h_{t-2}^L > y, \varepsilon_t + h_{t-1}^U \leq -y]$$

$$= P[\varepsilon_{t-1} + h_{t-2}^L > y, \varepsilon_t + |\varepsilon_{t-1}| \leq -y - (a + |g(C)|)/(1-b) - \sum_{j=1}^{\infty} b^j |\varepsilon_{t-j-1}|]$$

$$= EH(y - h_{t-2}^L, -y - (a + |g(C)|)/(1-b) - \sum_{j=1}^{\infty} b^j |\varepsilon_{t-j-1}|) > 0$$

by assumption. Finally, by the reasoning of the proof of Theorem 3, it follows that $I_{t-l-1,K}^1$ and $\tilde{I}_{t-l,K}^1$ are L_2-near epoch dependent with exponentially declining $\nu(\cdot)$ sequence, implying that $I_{t-l-1,K}^1 \tilde{I}_{t-l,K}^1$ is also L_2-near epoch dependent with exponentially declining $\nu(\cdot)$ sequence. Therefore, the reasoning of the proof of Theorem 3 for showing the law of large numbers applies to this case. For the case of showing a weak law of large numbers directly for $m^{-1} \sum_{l=0}^{m} I_{t-1-l,K} \tilde{I}_{t-l,K}$, note that for all K,

$$EI_{t-1-l,K} \tilde{I}_{t-l,K} > 0$$

as explained before, and therefore it only remains to show the L_2-near epoch dependence property for $I_{t-1-l,K} \tilde{I}_{t-l,K}$. To show this, note that

$$\| I_{t-1-l,K} \tilde{I}_{t-l,K} - E(I_{t-1-l,K} \tilde{I}_{t-l,K} | \varepsilon_{t-l-M}, \ldots, \varepsilon_{t-l}) \|_2$$

$$\leq \| I_{t-1-l,K} \tilde{I}_{t-l,K} - E(I_{t-1-l,K} | \varepsilon_{t-l-M}, \ldots, \varepsilon_{t-1-l}) E(\tilde{I}_{t-l,K} | \varepsilon_{t-l-M}, \ldots, \varepsilon_{t-l}) \|_2$$

$$\leq \| I_{t-1-l,K} - E(I_{t-1-l,K} | \varepsilon_{t-l-M}, \ldots, \varepsilon_{t-l-1}) \|_2 + \| \tilde{I}_{t-l,K} - E(\tilde{I}_{t-l,K} | \varepsilon_{t-l-M}, \ldots, \varepsilon_{t-l}) \|_2,$$

and by the earlier argument, the last expression converges to 0 as $M \to \infty$. □