THE LOCAL TO UNITY DYNAMIC TOBIT MODEL

ANNA BYKHOVSKAYA AND JAMES A. DUFFY

Abstract. This paper extends local to unity asymptotics to the non-linear setting of the dynamic Tobit model, motivated by the application of this model to highly persistent censored time series. We show that the standardised process converges weakly to a non-standard limiting process that is constrained (regulated) to be positive, and derive the limiting distributions of the OLS estimates of the model parameters. This allows inferences to be drawn on the overall persistence of a process (as measured by the sum of the autoregressive coefficients), and for the null of a unit root to be tested in the presence of censoring. Our simulations illustrate that the conventional ADF test substantially over-rejects when the data is generated by a dynamic Tobit with a unit root.

Keywords: non-negative time series, dynamic Tobit, local unit root, unit root test.

1. Introduction

Since the 1950s nonlinear models have played an increasingly prominent role in the analysis and prediction of time series data. In many cases, as was noted in early work by Moran (1953), linear models are unable to adequately match the features of observed time series. The efforts to develop models that enjoy the flexibility afforded by nonlinearities, while retaining the tractability of linear models, have subsequently engendered an enormous literature (see e.g. Fan and Yao, 2003; Gao, 2007; Chan, 2009; and Terasvirta *et al.*, 2010).

An important instance of non-linearity arises when data is bounded by, truncated at, or censored below some threshold, since such phenomena cannot be adequately captured – even approximately – by a linear model. Many observed series are bounded below by construction, and may spend lengthy periods at or near their lower boundary, such as unemployment rates, prices, sectoral trade flows, and nominal interest rates. The non-negativity of interest rates, and the resulting constraints that this may impose on the efficacy of monetary policy, has received particular attention in recent years, as central bank policy rates have remained at or near the zero lower bound for a significant portion of the past two decades, across many economies (see e.g. Mavroeidis, 2021, and the works cited therein).

A tractable model for such series, which generates both their characteristic serial dependence and censoring, is the dynamic Tobit model. In its static formulation, the model originates with Tobin (1958). In its dynamic formulation, the model typically comes in one of two varieties, which we refer to as the *latent* and *censored* models. In the latent dynamic Tobit, an unobserved process $\{y_t^*\}$ follows a linear autoregression, with $y_t = \max\{y_t^*, 0\}$ being observed; whereas in the censored dynamic Tobit, $\{y_t\}$ is modelled as the positive part of a linear function of its own lags, and an additive error (see e.g. Maddala 1983, p. 186, or Wei 1999, p. 419). In both models the right hand side may be augmented with other explanatory variables. Relative to the latent model, the censored model has the advantage of being Markovian, which greatly facilitates its use in forecasting. It has been successfully applied to a range of censored series, in both purely time series and panel data settings, including: the open market operations of the Federal Reserve (Demiralp and Jordà, 2002; de Jong and Herrera, 2011); household commodity purchases (Dong *et al.*, 2012); loan charge-off rates (Liu *et al.*, 2019); credit default and overdue loan repayments (Brezigar-Masten *et al.*, 2021); and sectoral bilateral trade flows (Bykhovskaya, 2021). Recently, Mavroeidis (2021) proposed the censored and kinked structural VAR model to describe the operation of monetary policy during periods when the zero lower bound may occasionally bind on the policy rate. If only the actual interest rate (rather than some 'shadow rate') affects agents' decision making, as assumed in closely related work by Aruoba *et al.* (2021), then the univariate counterpart of this model is exactly the censored dynamic Tobit.

The present work is concerned with the censored, rather than the latent, dynamic Tobit model. In the latent model the dynamics are simply those of the latent autoregression, and so are easily understood using existing results; whereas in the censored model, the censoring affects the dynamics of $\{y_t\}$ in a non-trivial manner, making the analysis rather more challenging. Indeed, establishing the stationarity or weak dependence of the censored dynamic Tobit is far from trivial, as can be seen from the works by Hahn and Kuersteiner (2010), de Jong and Herrera (2011), Michel and de Jong (2018), and Bykhovskaya (2021). Henceforth, all references to the 'dynamic Tobit' are to the censored version of the model.

Motivated in part by recent work on modelling nominal interest rates near the zero lower bound, our concern is with the application of this model to series that are highly persistent, so that above the censoring point they exhibit the random wandering that is characteristic of integrated processes. The appropriate configuration of the dynamic Tobit model for such series, in which the autoregressive polynomial has a root local to unity, has not been considered in the literature to date – apart from the special case of a first-order model with an exact unit root, as in Cavaliere (2004) and Bykhovskaya (2021). Our results are thus entirely new to the literature.

Our principal technical contribution, within this setting, is to derive the limiting distributions of both the standardised regressor process, and the ordinary least squares (OLS) estimates of the parameters of the dynamic Tobit, when that model has an autoregressive root local to unity. In this setting, OLS is consistent and we obtain a usable limit theory for the estimated sum of the autoregressive coefficients, which conventionally provides a measure of the overall persistence of a process (cf. Andrews and Chen, 1994; Mikusheva, 2007). Our asymptotics provide the basis for practical unit root tests for highly persistent, censored time series. The associated t statistic coincides with the (constant only) augmented Dickey–Fuller (ADF) t statistic, but employs critical values modified to reflect the censoring present in the data generating process. We show, via Monte Carlo simulations, that as our critical values are larger than the conventional ADF critical values, their use eliminates the significant over-rejection that may result from the naive application of the latter to censored data.

Our work may be construed, more broadly, as extending the analysis of highly persistent time series, and the associated machinery of unit root testing, from a linear setting to a nonlinear setting appropriate to time series that are subject to a lower bound. In doing so, we complement the seminal work of Cavaliere (2005), which similarly sought to extend this machinery to the setting of bounded time series. Our contribution here is to effect this extension within a class of nonlinear autoregressive models that have been widely applied to censored time series (as can be seen from works cited above), and which fall outside his framework.

On a technical level, the most closely related works to our own are those of Cavaliere (2004, 2005) and Cavaliere and Xu (2014), who develop the asymptotics of what they term 'limited autoregressive processes' with a near-unit root, which are (one- or two-sided) censored processes constructed by the addition of regulators to a latent linear autoregression. While their (one-sided) model has a superficial resemblance to the dynamic Tobit, there are important, but subtle differences between the two: see Section 2.4 for a discussion. Perhaps the most striking similarity is that both models, in the case of an exact unit root, give rise to regressor processes that converge weakly to regulated Brownian motions – but whose limits differ from each other when roots are merely local to unity. Similar asymptotics, in the exact unit root case, are also shared by threshold autoregressive models with a unit root regime and a stationary regime, as considered by Liu *et al.* (2011) and Gao *et al.* (2013) in the first-order case.

The remainder of this paper is organised as follows. Section 2 discusses the model and our assumptions. Asymptotic results and corresponding tests are derived in Section 3, while supporting Monte Carlo simulations are shown in Section 4. Section 5 applies our framework to the exchange rate between the Swiss franc and the euro during a period when this was subject to a lower bound. Finally, Section 6 concludes. All proofs appear in the appendices.

Notation. C, C', C'', etc. denote generic constants that may take different values in different parts of this paper. All limits are taken as $T \to \infty$, unless otherwise specified. $\xrightarrow{p}$ and $\xrightarrow{d}$ respectively denote convergence in probability and distribution (weak convergence). We write '$X_T(r) \xrightarrow{d} X(r)$ on $D[0,1]$' to denote that $\{X_T\}$ converges weakly to X, where these are considered as random elements of $D[0,1]$, the space of cadlag functions on $[0,1]$, equipped with the uniform topology. For $p \geq 1$ and X a random variable, let $\|X\|_p := (\mathbb{E}|X|^p)^{1/p}$.

2. The dynamic Tobit model with a near-unit root

2.1. Model and assumptions. Consider a time series $\{y_t\}$ generated by the dynamic Tobit model of order $k \geq 1$, written in augmented Dickey–Fuller (ADF) form,[1]

$$y_t = \left[\alpha + \beta y_{t-1} + \sum_{i=1}^{k-1} \phi_i \Delta y_{t-i} + u_t\right]_+, \quad t = 1, \ldots, T, \tag{2.1}$$

where $\Delta y_t := y_t - y_{t-1}$, and $[x]_+ := \max\{x, 0\}$ denotes the positive part of $x \in \mathbb{R}$. We impose the following on the data generating process (2.1).

Assumption A1. *$\{y_t\}$ is initialised by (possibly) random variables $\{y_{-k+1}, \ldots, y_0\}$. Moreover, $T^{-1/2} y_0 \xrightarrow{p} b_0$ for some $b_0 \geq 0$.*

Assumption A2. *$\{y_t\}$ is generated according to (2.1), where:*

1. *$\{u_t\}_{t\in\mathbb{Z}}$ is independently and identically distributed (i.i.d.) with $\mathbb{E}u_t = 0$ and $\mathbb{E}u_t^2 = \sigma^2$.*
2. *$\alpha = \alpha_T := T^{-1/2}a$ and $\beta = \beta_T = \exp(c/T)$ for some $a, c \in \mathbb{R}$.*
3. *All roots of $\phi(z) := 1 - \sum_{i=1}^{k-1} \phi_i z^i$ lie strictly outside the unit circle.*

Assumption A3. *There exist $\delta_u > 0$ and $C < \infty$ such that:*

1. *$\mathbb{E}|u_t|^{2+\delta_u} < C$.*
2. *$\mathbb{E}|T^{-1/2}y_0|^{2+\delta_u} < C$, and $\mathbb{E}|\Delta y_i|^{2+\delta_u} < C$ for $i \in \{-k+2, \ldots, 0\}$.*

Figure 2.1 displays a typical sample path for the dynamic Tobit (2.1), under the preceding assumptions.

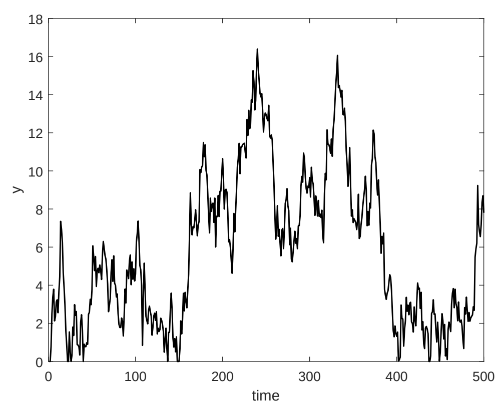


FIGURE 2.1. $y_t = \left[\frac{a}{T^{1/2}} + \left(1 + \frac{c}{T}\right) y_{t-1} + u_t\right]_+$, $a = 1$, $c = -5$, $y_0 = 0$, $u_t \sim$ i.i.d. $\mathcal{N}(0, 1)$, $T = 500$.

[1]The autoregressive form is $y_t = [\alpha + \sum_{i=1}^{k} \beta_i y_{t-i} + u_t]_+$, where $\beta_1 = \beta + \phi_1$, $\beta_k = -\phi_{k-1}$, and $\beta_i = \phi_i - \phi_{i-1}$ for $i \in \{2, \ldots, k-1\}$. In particular $\beta = \sum_{i=1}^{k} \beta_i$ corresponds to the sum of the autoregressive coefficients.

The main consequences of our assumptions may be summarised as follows.

(i) By the functional central limit theorem, A2.1 implies $T^{-1/2}\sum_{t=1}^{\lfloor rT\rfloor} u_t \xrightarrow{d} \sigma W(r)$ on $D[0,1]$, where $W(\cdot)$ is a standard Brownian motion. This convergence alone is sufficient to determine the asymptotics of $T^{-1/2}y_{\lfloor rT\rfloor}$, and of the OLS estimators, when $k=1$ (Theorems 3.1 and 3.3 below), but extending these results to $k\geq 2$ necessitates the slightly stronger conditions on $\{u_t\}$ provided by Assumption A3.

(ii) In the absence of censoring, A2.2 would entail that y_t has an autoregressive root within a $O(T^{-1})$ neighbourhood of real unity. Just as in that case, we shall show that in the present setting $T^{-1/2}y_{\lfloor rT\rfloor}$ converges weakly to a continuous process, albeit one that differs importantly from the diffusion process limit familiar from the uncensored case.

(iii) We require $\alpha = O(T^{-1/2})$ in A2.2, to ensure that the drift in $\{y_t\}$ is of no larger order than the stochastic trend component. If this assumption were relaxed, so that e.g. α were now a non-zero constant, the large-sample behaviour of $\{y_t\}$ and the asymptotics of the OLS estimators would be quite different from those developed here. A fixed positive α would generate an increasing linear trend, driving y_t ever further away from origin and making the censoring ultimately irrelevant; whereas a fixed negative α would lead to $\{y_t\}$ being stationary (see, e.g., Bykhovskaya (2021, Theorem 3)).

(iv) The specific parametrisations in A2.2 are chosen merely for convenience: all of our results also hold when α_T and β_T more generally satisfy $T^{1/2}\alpha_T \to a$ and $T(\beta_T - 1)\to c$. For ease of notation, we shall routinely suppress the T subscripts on α_T and β_T throughout the following.

(v) Assumption A2.3 ensures that $B(z) \coloneqq 1-\beta z-(1-z)\sum_{i=1}^{k-1}\phi_i z^i$ has only one root that is local to unity.

(vi) Assumptions A1 and A3 imply that $T^{-1/2}y_i \xrightarrow{p} b_0$ for $i\in\{-k+1,\dots,0\}$.

2.2. Non-zero lower bound. Our machinery extends straightforwardly to the case where y_t is censored at some $\mathbf{L}\neq 0$. Suppose (2.1) is modified to

$$y_t = \max\left\{\mathbf{L},\ \alpha+\beta y_{t-1}+\sum_{i=1}^{k-1}\phi_i\Delta y_{t-i}+u_t\right\}, \tag{2.2}$$

and take $\mathbf{L} = T^{1/2}\ell$ for some $\ell\in\mathbb{R}$, to allow the censoring point to be of the same order of magnitude as $\{y_t\}$. Defining $\tilde{y}_t \coloneqq y_t - \mathbf{L}$ and subtracting $\mathbf{L}$ from both sides of (2.2), it may be verified that

$$\tilde{y}_t = \left[\tilde{\alpha}+\beta\tilde{y}_{t-1}+\sum_{i=1}^{k-1}\phi_i\Delta\tilde{y}_{t-i}+u_t\right]_+$$

where $\tilde{\alpha} \coloneqq \alpha + (\beta - 1)\mathbf{L}$. Thus, $\{\tilde{y}_t\}$ follows a dynamic Tobit with censoring at zero, with drift

$$T^{1/2}\tilde{\alpha} = T^{1/2}[\alpha + (\beta - 1)T^{1/2}\ell] \to a + c\ell =: \tilde{a}$$

and initialisation

$$\tilde{b}_0 = \operatorname*{plim}_{T\to\infty} T^{-1/2}\tilde{y}_0 = \operatorname*{plim}_{T\to\infty} T^{-1/2}(y_0 - \mathbf{L}) = b_0 - \ell.$$

All our results hold in the setting of (2.2), with appropriate modifications. For simplicity, we work with $\mathbf{L} = 0$ throughout the rest of the paper, except where indicated otherwise.

2.3. Alternative representation. It will be occasionally useful to rewrite (2.1) in a form that helps to clarify the connections between the dynamic Tobit and the linear autoregressive model. We can do this by defining

$$y_t^- \coloneqq \left[\alpha + \beta y_{t-1} + \sum_{i=1}^{k-1} \phi_i \Delta y_{t-i} + u_t\right]_- , \tag{2.3}$$

where $[x]_- \coloneqq \min\{x, 0\}$. That is, when $y_t = 0$, y_t^- records the value that y_t would have taken had it not been censored at zero.

Since $[x]_+ = x - [x]_-$, we may then rewrite (2.1) as

$$y_t = \alpha + \beta y_{t-1} + \sum_{i=1}^{k-1} \phi_i \Delta y_{t-i} + u_t - y_t^-, \tag{2.4}$$

or equivalently, letting L denote the lag operator, as

$$B(L)y_t = \alpha + u_t - y_t^-. \tag{2.5}$$

Thus, if we view y_t^- as an additional noise term, (2.5) takes the form of a linear autoregression. The main challenge is that y_t^- is itself a complicated non-linear object. Nonetheless, we will be able to show that it is, roughly speaking, negligible, at least so far as the OLS estimates are concerned.

2.4. Connections with limited autoregressive processes. The representation (2.5) allows us to draw out the connections between our model and the limited autoregressive processes developed by Cavaliere (2004, 2005) and Cavaliere and Xu (2014). To put their model – for the special case of a process constrained to lie in $[0, \infty)$ – in a form comparable to ours, consider a latent process $\{x_t^*\}$,

$$x_t^* = \rho_T x_{t-1}^* + \varepsilon_t, \quad \rho_T = 1 + c/T, \tag{2.6}$$

where $\{\varepsilon_t\}$ is stationary. Define an observed process $\{x_t\}$, whose increments are related to those of $\{x_t^*\}$ via

$$\Delta x_t = \Delta x_t^* + \underline{\xi}_t \tag{2.7}$$

where $\underline{\xi}_t > 0$ if and only if $x_{t-1} + \Delta x_t^* < 0$, so as to ensure that $x_t \geq 0$ for all t. In particular, if we set

$$\underline{\xi}_t = -x_t^- := -[x_{t-1} + \Delta x_t^*]_- \tag{2.8}$$

then $\{x_t\}$ will be censored at zero. When $c = 0$, by combining (2.6)–(2.8) we obtain

$$x_t = x_{t-1} + \varepsilon_t - x_t^- \tag{2.9}$$

as a valid representation of a limited autoregressive process censored at zero.

While (2.9) and (2.5) both describe censored processes, they also have some subtle, and yet important differences. To clarify these, suppose $\alpha = 0$ and $\beta = 1$, so that $B(L) = (1 - L)\phi(L)$ in (2.5), where $\phi(z)$ has all its roots outside the unit circle; and also suppose that $\varepsilon_t = \phi(L)^{-1}u_t$ in (2.6). Then the dynamic Tobit (2.5) can be rewritten as

$$y_t = y_{t-1} + \varepsilon_t - \phi(L)^{-1}y_t^-. \tag{2.10}$$

Comparing (2.9) and (2.10), we can see that the censoring affects the dynamics of $\{y_t\}$ and $\{x_t\}$ in different ways. Lagged values of y_t^- have a direct effect on future y_t (via $\phi(L)^{-1}y_t^- = \sum_{i=0}^{\infty} c_i y_{t-s}^-$), whereas lagged x_t^- have no such effect on x_t. Indeed, the only case in which $\{y_t\}$ and $\{x_t\}$ will coincide is if $\phi(L) = 1$ (for a further discussion of which case, see Cavaliere (2005, Remark 2.3)).

3. Asymptotic results

In this section we derive the weak limits of the standardised process $T^{-1/2}y_{\lfloor rT \rfloor}$ and the ordinary least squares (OLS) estimators of the parameters of the dynamic Tobit. The latter provides the basis for a unit root test for non-negative time series.

3.1. Limiting distribution of the regressor process. Let $\theta := (a, b_0, c)$, define the process

$$K_\theta(r) := b_0 + a \int_0^r \mathrm{e}^{-cs}\,\mathrm{d}s + \sigma \int_0^r e^{-cs}\,\mathrm{d}W(s), \tag{3.1}$$

and denote its 'regulated' counterpart by

$$J_\theta(r) := \mathrm{e}^{cr}\left\{K_\theta(r) + \sup_{r' \leq r}[-K_\theta(r')]_+\right\}. \tag{3.2}$$

We first provide a result for the case $k = 1$, under which the model (2.1) reduces to

$$y_t = [\alpha + \beta y_{t-1} + u_t]_+. \tag{3.3}$$

Theorem 3.1. *Suppose Assumptions* A1 *and* A2 *hold with $k = 1$ in* (2.1)*. Then on $D[0, 1]$,*

$$T^{-1/2}y_{\lfloor rT \rfloor} \xrightarrow{d} J_\theta(r). \tag{3.4}$$

The preceding is a new result, which relates to some of the previous literature as follows.

(i) The supremum in the definition of $J_\theta(\cdot)$ guarantees non-negativity of the limiting process: if $K_\theta(r)$ is negative, $[-K_\theta(r)]_+ = -K_\theta(r) > 0$, so that $K_\theta(r) + \sup_{r' \leq r}[-K_\theta(r')]_+ \geq 0$. The appearance of the supremum is in line with the solution to the Skorokhod reflection problem (Revuz and Yor (1999, p. 239)).

(ii) Suppose that $a = b_0 = 0$. Then $\mathrm{e}^{cr} K_\theta(r) = S_c(r) = \sigma \int_0^r \mathrm{e}^{c(r-s)}\, \mathrm{d}W(s)$, an Ornstein–Uhlenbeck process with autoregressive parameter c (e.g., Chan and Wei (1987); Phillips (1987b)), and (3.4) specialises to

$$T^{-1/2} y_{\lfloor rT \rfloor} \xrightarrow{d} S_c(r) + \sup_{r' \leq r}[-\mathrm{e}^{c(r-r')} S_c(r')]_+.$$

on $D[0,1]$. Taking $\varepsilon_t = u_t$, the limited autoregressive process (2.6)–(2.8) satisfies

$$T^{-1/2} x_{\lfloor rT \rfloor} \xrightarrow{d} S_c(r) + \sup_{r' \leq r}[-S_c(r')]_+$$

on $D[0,1]$. Comparing the two preceding limits, we observe a subtle but crucial difference, due to the presence of the factor $\mathrm{e}^{c(r-r')}$.

(iii) When $a = b_0 = c = 0$, $J_\theta(r)$ coincides with a Brownian motion regulated from below at zero, which has the same distribution as a Brownian motion reflected at the origin, $|W(\cdot)|$, see e.g. Karatzas and Shreve (2012, p. 97). Another model that generates a process with this asymptotic distribution (upon rescaling by $T^{-1/2}$) is a first-order threshold autoregression with 'unit root' and 'stationary' regimes, as studied by Liu *et al.* (2011) and Gao *et al.* (2013). A special case of their model posits

$$x_t = \beta(x_{t-1}) x_{t-1} + u_t,$$

where $\beta(x) = 1$ if $x \geq 0$, and $\beta(x) = 0$ otherwise. It follows that $x_t = [x_{t-1}]_+ + u_t$, and so $[x_t]_+ = [[x_{t-1}]_+ + u_t]_+$, which corresponds to our setting (2.1) with $\alpha = 0$, $\beta = 1$, $k = 1$, and $y_t = [x_t]_+$. It is thus not surprising that, in this case, our Theorem 3.1 agrees exactly with the corresponding Theorem 3.1 of Liu *et al.* (2011).

When $k > 1$, the other roots of the lag polynomial $B(L)$ affect the behaviour of $\{y_t\}$, and we need a further condition to ensure that the first differences $\{\Delta y_t\}$ are well behaved. Let

$$F_\delta := \begin{bmatrix} \phi_1\delta & \phi_2 & \cdots & \phi_{k-2} & \phi_{k-1} \\ \delta & 0 & \cdots & 0 & 0 \\ 0 & 1 & & & \\ & & \ddots & & \\ & & & 1 & 0 \end{bmatrix}. \tag{3.5}$$

Under an appropriate condition on the matrices $\{F_\delta \mid \delta \in [0,1]\}$, we can ensure $\{\Delta y_t\}$ is stochastically bounded. To state that, we need the following (cf. Jungers (2009), Defn. 1.1):

Definition. The *joint spectral radius* (JSR) of a bounded collection $\mathcal{A}$ of square matrices is

$$\lambda_{\mathrm{JSR}}(\mathcal{A}) \coloneqq \limsup_{m\to\infty} \sup_{B\in\mathcal{A}^m} \lambda(B)^{1/m}$$

where $\lambda(B)$ denotes the spectral radius of B, and $\mathcal{A}^m \coloneqq \{\prod_{i=1}^m A_i \mid A_i \in \mathcal{A}\}$.

Control over the JSR has been previously used to ensure the stationarity of regime-switching autoregressive models (e.g., Liebscher (2005); Saikkonen (2008)), and we shall utilise it in a similar manner here.

Assumption A4. $\lambda_{\mathrm{JSR}}(\{F_0, F_1\}) < 1$.

The counterpart of this condition, in a linear autoregression, would be a requirement that the roots of $\phi(z)$ are strictly bounded away from the unit circle (Assumption A2.3). Such a condition is indeed necessary, though not sufficient for A4, which accordingly implies that $\phi(1) > 0$.[2]

Theorem 3.2. *Suppose Assumptions* A1–A4 *hold. Then*

$$T^{-1/2} y_{\lfloor rT \rfloor} \xrightarrow{d} \phi(1)^{-1} J_{\theta_\phi}(r) =: Y_{\theta_\phi}(r) \tag{3.6}$$

on $D[0,1]$*, where* $\theta_\phi \coloneqq [a, \phi(1)b_0, \phi^{-1}(1)c]$.

The principal difference between Theorems 3.1 and 3.2 is that when $k > 1$, the stationary roots appear in the limit via the factor $\phi(1)$. Notably, the local autoregressive parameter c is replaced by $\phi(1)^{-1}c$ – exactly as it would be if $\{y_t\}$ were generated by a linear autoregression with a root local to unity (cf. Hansen, 1999, p. 599). Indeed, $\phi(1) = 1$ when $k = 1$, so in this case the two results coincide.

Corollary 3.1. *For the Tobit model* (2.2) *with censoring point* $\mathbf{L}$*,* $T^{-1/2}\tilde{y}_{\lfloor rT \rfloor} \xrightarrow{d} Y_{\tilde{\theta}_\phi}(r)$*, where* $\tilde{\theta} \coloneqq [\tilde{a}, \phi(1)\tilde{b}_0, \phi^{-1}(1)c]$.

3.2. OLS estimates. We first consider the case where $k = 1$, as in the model (3.3), to develop intuition for our results.

When estimating (3.3) by OLS, we need to decide which deterministic terms should be included in the regression. In the absence of censoring, i.e. if the data generating process were simply a linear autoregression, the inclusion of a constant and a linear trend would render the distribution of the OLS estimator of β free of any nuisance parameters related to

[2]That this condition is not sufficient for A4 can be seen e.g. by taking $k-1 = 2$, $\phi_1 = 1.3$ and $\phi_2 = -0.8$. Then all the roots of $\phi(z)$ are indeed outside the unit circle, but the largest eigenvalue (in modulus) of $F_1F_1F_0$ is -1.04, and so $\lambda_{\mathrm{JSR}}(\{F_0, F_1\}) \geq |-1.04|^{1/3} > 1$. On the other hand, simulations indicate that $\{\Delta y_t\}$ appears stochastically bounded in this case, so it remains an open question as to whether a condition merely on the roots of $\phi(z)$ is sufficient for Theorem 3.2.

the deterministic components.[3] Unfortunately, the nonlinearity introduced by the censoring entails that α – or rather, the local parameter a – will show up in the limiting distribution of $\hat{\beta}_T$, *irrespective* of which deterministics are included in the regression. To permit inferences to also be drawn on a, if required, we consider the OLS regression of y_t on a constant and y_{t-1}, i.e.

$$\begin{bmatrix}\hat{\alpha}_T \\ \hat{\beta}_T\end{bmatrix} := \left(\sum_{t=1}^{T}\begin{bmatrix}1 & y_{t-1} \\ y_{t-1} & y_{t-1}^2\end{bmatrix}\right)^{-1}\sum_{t=1}^{T}\begin{bmatrix}1 \\ y_{t-1}\end{bmatrix} y_t =: \mathcal{M}_T^{-1} m_T. \tag{3.7}$$

In the stationary dynamic Tobit model, OLS is inconsistent (see e.g., Bykhovskaya, 2021, Supplementary Material, Lemma B.1). However, as the following shows, when β is local to unity, consistency is restored. The reason is that observations in the vicinity of zero accumulate only at rate $T^{1/2}$, so that a vanishingly small fraction of the sample is affected by the censoring.

Theorem 3.3. *Suppose Assumptions* A1 *and* A2 *hold, with $k = 1$ in* (2.1). *Then*

$$\begin{aligned}\begin{bmatrix}T^{1/2}(\hat{\alpha}_T - \alpha) \\ T(\hat{\beta}_T - \beta)\end{bmatrix} &\xrightarrow{d} \begin{bmatrix}1 & \int_0^1 J_\theta(r)\,\mathrm{d}r \\ \int_0^1 J_\theta(r)\,\mathrm{d}r & \int_0^1 J_\theta^2(r)\,\mathrm{d}r\end{bmatrix}^{-1}\begin{bmatrix}J_\theta(1) - c\int_0^1 J_\theta(r)\,\mathrm{d}r - b_0 - a \\ \sigma\int_0^1 J_\theta(r)\,\mathrm{d}W(r)\end{bmatrix} \\ &=: \mathcal{J}_\theta^{-1}\mathcal{U}_\theta =: \begin{bmatrix}\mathfrak{a}_\theta \\ \mathfrak{b}_\theta\end{bmatrix}.\end{aligned} \tag{3.8}$$

Remark 3.1. Letting $J_\theta^\mu(r) := J_\theta(r) - \int_0^1 J_\theta(r)\,\mathrm{d}r$, an alternative expression for the limiting distribution of $\hat{\beta}_T$ is given by

$$\mathfrak{b}_\theta = \frac{J_\theta^\mu(1)^2 - J_\theta^\mu(0)^2 - \sigma^2}{2\int (J_\theta^\mu(r))^2\,\mathrm{d}r} - c. \tag{3.9}$$

This agrees with the limiting distribution that would be obtained in the linear autoregressive model, except with $J_\theta(\cdot)$ taking the place of the usual Ornstein–Uhlenbeck process. (See Appendix A.2 for details.)

---

[3]Strictly speaking, this is true only if the autoregressive model is formulated in 'unobserved components' form (see e.g. Andrews and Chen (1994, Section 2.1) as

$$y_t = \mu + \delta t + y_t^* \qquad\qquad y_t^* = \beta y_{t-1}^* + u_t$$

so that the presence (or absence) of a linear drift in y_t is independent of the value of β, and so can always be removed by deterministic detrending. By contrast, if the model is formulated 'directly' as

$$y_t = \alpha + \beta y_{t-1} + u_t,$$

then the linear trend that is present when $\beta = 1$ becomes an exponential trend when β is local to unity. In the present (censored) setting, we may note that (3.3) is *not* equivalent to

$$y_t = [\mu + \delta t + y_t^*]_+ \qquad\qquad y_t^* = \beta y_{t-1}^* + u_t.$$

(This model is in fact the *latent* dynamic Tobit referred to in Section 1.)

For the case of general $k \geq 1$, let $\boldsymbol{\phi} := (\phi_1, \ldots, \phi_{k-1})^\mathsf{T}$, and

$$(\hat{\alpha}_T, \hat{\beta}_T, \hat{\phi}_{1,T}, \ldots, \hat{\phi}_{k-1,T}) := \underset{(a,b,f_1,\ldots,f_{k-1})}{\operatorname{argmin}} \sum_{t=1}^{T} \left(y_t - a - by_{t-1} - \sum_{i=1}^{k-1} f_i \Delta y_{t-i} \right)^2 \tag{3.10}$$

denote the OLS estimators of the parameters of (2.1). Since, as the next results shows, the limiting distributions of $(\hat{\alpha}_T, \hat{\beta}_T)$ depend on $\phi(1)$, a consistent estimate of that quantity is needed to compute valid critical values for test statistics based on these estimators. The following also guarantees the consistency $\hat{\phi}(1) := 1 - \sum_{i=1}^{k-1} \hat{\phi}_{i,T}$.

Theorem 3.4. *Suppose Assumptions* A1–A4 *hold. Then* $\hat{\boldsymbol{\phi}}_T \xrightarrow{p} \boldsymbol{\phi}$, *and*

$$\begin{bmatrix} T^{1/2}(\hat{\alpha}_T - \alpha) \\ T(\hat{\beta}_T - \beta) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} 1 & \int Y_{\theta_\phi}(r)\,\mathrm{d}r \\ \int Y_{\theta_\phi}(r)\,\mathrm{d}r & \int Y_{\theta_\phi}^2(r)\,\mathrm{d}r \end{bmatrix}^{-1} \begin{bmatrix} \phi(1)[Y_{\theta_\phi}(1) - b_0 - c_\phi \int Y_{\theta_\phi}(r)\,\mathrm{d}r] - a \\ \sigma \int Y_{\theta_\phi}(r)\,\mathrm{d}W(r) \end{bmatrix} \tag{3.11}$$

Remark 3.2. The parameters of the Tobit model (2.2) with censoring point $\mathbf{L}$ can be estimated as in (3.10), with $\tilde{y}_t$ in place of y_t and with $\tilde{\theta}_\phi = (\tilde{a}, \phi(1)\tilde{b}_0, \phi^{-1}(1)c)$ in (3.11).

3.3. Unit root tests. The preceding results allow us to conduct asymptotically valid hypothesis tests on key parameters of the dynamic Tobit: in particular, to test the hypothesis of a unit root in this setting. In contrast to the linear model, where the random wandering typical of a unit root occurs whenever the sum of the autoregressive coefficients is unity, in the dynamic Tobit the value of the intercept also matters. In particular, we need to exclude the possibility of a negative intercept ($\alpha < 0$) that would otherwise continually push the process back towards the censoring point, thereby rendering it stationary (Bykhovskaya, 2021, Theorem 3). Thus, when testing the null of a 'unit root' in this model, it is more appropriate to regard this as a test of the null that $\alpha = 0$ and $\beta = 1$, as opposed to merely the restriction that $\beta = 1$, with it being desirable to reject this null in favour of a stationary alternative, when either $\beta < 1$ (exactly as in a linear model), or when $\beta = 1$ but $\alpha < 0$.

To construct our test statistics, we need an estimate of the error variance σ^2. We use $\hat{\sigma}_T^2 := \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^2$, where

$$\hat{u}_t := y_t - \hat{\alpha}_T - \hat{\beta}_T y_{t-1} - \sum_{i=1}^{k-1} \hat{\phi}_{i,T} \Delta y_{t-i}. \tag{3.12}$$

That is, $\{\hat{u}_t\}$ are the OLS residuals, computed as if y_t were not subject to censoring. Let $\mathcal{M}_T := \sum_{t=1}^{T} \boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}^\mathsf{T}$, where $\boldsymbol{x}_t := (1, y_{t-1}, \Delta y_{t-1}, \ldots, \Delta y_{t-k+1})^\mathsf{T}$.

Corollary 3.2. *Suppose Assumptions* A1 *and* A2 *hold. If either:* $k = 1$ *in* (2.1)*; or* $k > 1$, A3 *and* A4 *holds, then* $\hat{\sigma}_T^2 \xrightarrow{p} \sigma^2$ *and*

$$t_{\alpha,T} := \frac{\hat{\alpha}_T - \alpha}{\hat{\sigma}\sqrt{\mathcal{M}_T^{-1}(1,1)}} \xrightarrow{d} \frac{\mathfrak{a}_{\theta_\phi}}{\sigma\sqrt{\mathcal{J}_{\theta_\phi}^{-1}(1,1)}} \qquad t_{\beta,T} := \frac{\hat{\beta}_T - \beta}{\hat{\sigma}\sqrt{\mathcal{M}_T^{-1}(2,2)}} \xrightarrow{d} \frac{\mathfrak{b}_{\theta_\phi}}{\sigma\sqrt{\mathcal{J}_{\theta_\phi}^{-1}(2,2)}}, \tag{3.13}$$

where $M^{-1}(i,j)$ denotes the (i,j) element of M^{-1}.

This result allows us to conduct a one-sided test of a unit root versus a stationary alternative, which rejects when $t_{\beta,T} \leq c$, where c is drawn from an appropriate quantile of the asymptotic distribution of t_β. As the simulations in the following section illustrate, such a test indeed has the desirable properties outlined above, in the sense of tending to reject both when either $\beta < 1$, or when $(\beta = 1,\ \alpha < 0)$, i.e. it has power to reject the null whenever $\{y_t\}$ is stationary.

4. Simulations

We now illustrate how the values of the initial condition b_0 and the localising parameters a and c affect the distribution of t_β, and compare the performance of a test based on critical values derived from Corollary 3.2 with one based on the conventional ADF critical values (Dickey and Fuller (1979)), when the data is subject to censoring.

4.1. Effect of b_0. Figure 4.1 depicts how a change in b_0 shifts the density of t_β. As b_0 moves further above zero, the density shifts progressively to the right, as the probability that any trajectory of K_θ (initialised at b_0) will reach zero, and so be subject to censoring, correspondingly declines. Indeed, once b_0 is sufficiently large to make this probability negligible, the density becomes indistinguishable from that generated by a linear model (the solid green line in Figure 4.1), which is invariant to b_0. (Under the parametrisation used in the figure, this occurs when $b_0 = 2$; in general, this will depend on the magnitude of $\phi(1)b_0/\sigma$, in accordance with Theorem 3.2.)

For the remainder of this section, all simulations are conducted with $y_0 = b_0 = 0$.

4.2. Effects of a and c. Figure 4.2 shows how a change in local intercept a (left panel) and local slope coefficient c (right panel) affects the density of t_β. The means of these distributions across a range of values for a and c are also reported in Table 4.1. We can see that as a or c fall further below zero, the distribution of t_β (both its mean and its entire probability mass) shifts leftward – with the opposite effect being observed when these parameters are progressively raised above zero. This has desirable consequences for power, as discussed below.

4.3. Power. The preceding illustrates how changes in a and/or c may shift the distribution of t_β in either direction, and so will affect the ability of the test to reject the null of a unit root (i.e. $H_0 : \alpha = 0, \beta = 1$). Power envelopes (rejection probabilities for a nominal 5 per cent, one-sided test), are displayed in Figure 4.3. These show that, on the one side, more negative values of a and/or c make it easier for the test to reject the null in favour of a stationary alternative. The power eventually reaches 100 per cent, indicating the consistency of the test against fixed alternatives in this region. This tendency to reject the null, as a

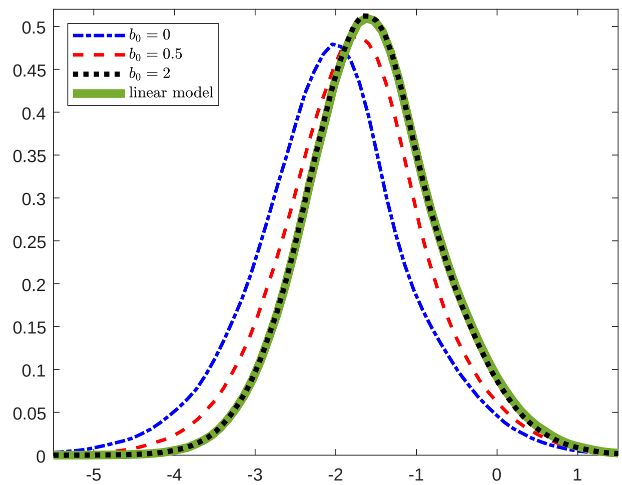
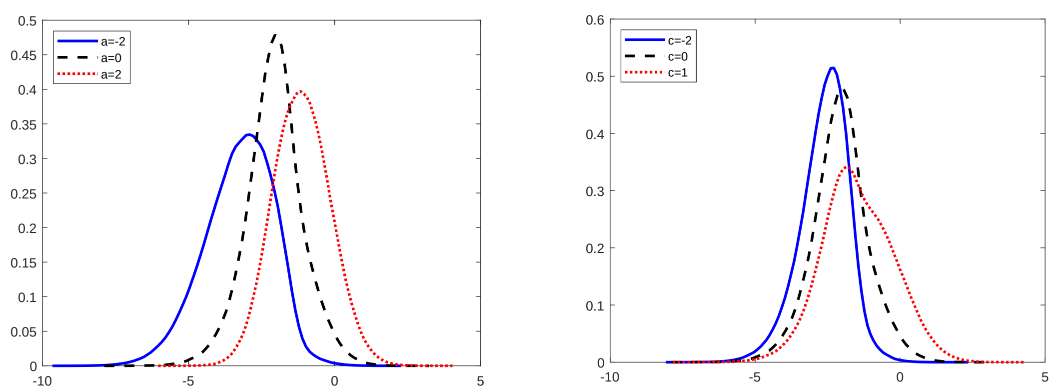


FIGURE 4.1. Densities of t_β under the Tobit and linear models. Data generating process is $y_t = [y_{t-1} + u_t]_+$, $y_0 = b_0\sqrt{T}$ for Tobit model and $y_t^\ell = y_{t-1}^\ell + u_t$, $y_0^\ell = 0$ for a linear model, $u_t \sim$ i.i.d. $\mathcal{N}(0,1)$. Data is obtained from 10^6 samples of length $T = 1000$.



(A) Effect of a change in a when $c = 0$. (B) Effect of a change in c when $a = 0$.

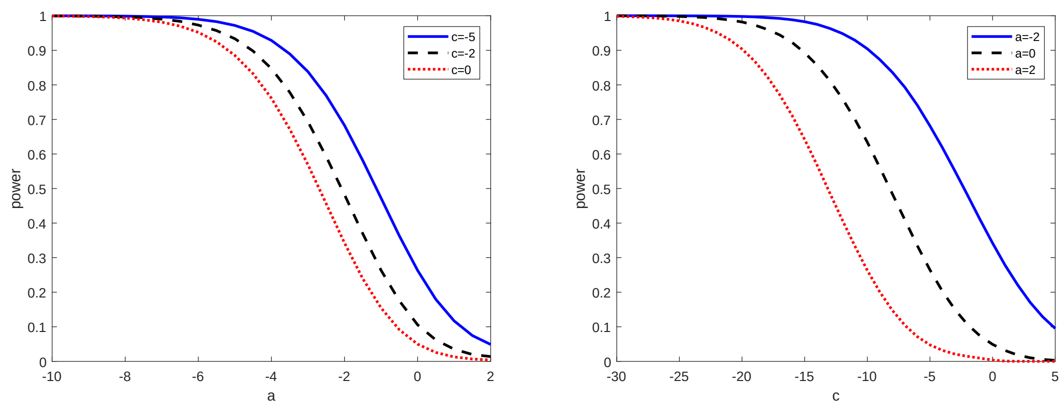
FIGURE 4.2. Densities of t-ratio t_β for various values of a, c. Data generating process is $y_t = \left[\frac{a}{\sqrt{T}} + \left(1 + \frac{c}{T}\right) y_{t-1} + u_t\right]_+$, $y_0 = 0$, $u_t \sim$ i.i.d. $\mathcal{N}(0,1)$. Data is obtained from 10^6 samples of time series of length $T = 1000$.

falls below zero, is in fact a desirable property of the test in this setting, since having $\alpha < 0$ in the dynamic Tobit implies that $\{y_t\}$ is stationary, even when $\beta = 1$ (Bykhovskaya, 2021, Theorem 3).

On the other side, positive values of c move $\{y_t\}$ into the explosive region, and our tendency to not reject in these cases is entirely consistent with the use of a one-sided test, as it is in a linear model (Figure 4.3b). Although we also fail to reject for sufficiently positive values of

a \ c	-5	-2	-1	0	1	2	5
-5	-6.09	-5.70	-5.56	-5.42	-5.26	-5.09	-4.37
-2	-4.24	-3.70	-3.49	-3.27	-3.00	-2.62	8.93
-1	-3.69	-3.10	-2.88	-2.62	-2.26	-1.51	23.05
0	-3.20	-2.60	-2.37	-2.06	-1.46	0.03	43.23
1	-2.82	-2.26	-2.00	-1.56	-0.57	1.85	68.26
2	-2.58	-2.06	-1.76	-1.13	0.28	3.68	96.73
5	-2.52	-2.05	-1.57	-0.48	2.18	9.00	193.14

TABLE 4.1. Mean of t-ratio t_β for various values of a, c. Data generating process is $y_t = \left[\frac{a}{\sqrt{T}} + \left(1 + \frac{c}{T}\right) y_{t-1} + u_t\right]_+$, $y_0 = 0$, $u_t \sim$ i.i.d. $\mathcal{N}(0, 1)$. Data is obtained from 10^6 samples of time series of length $T = 1000$.

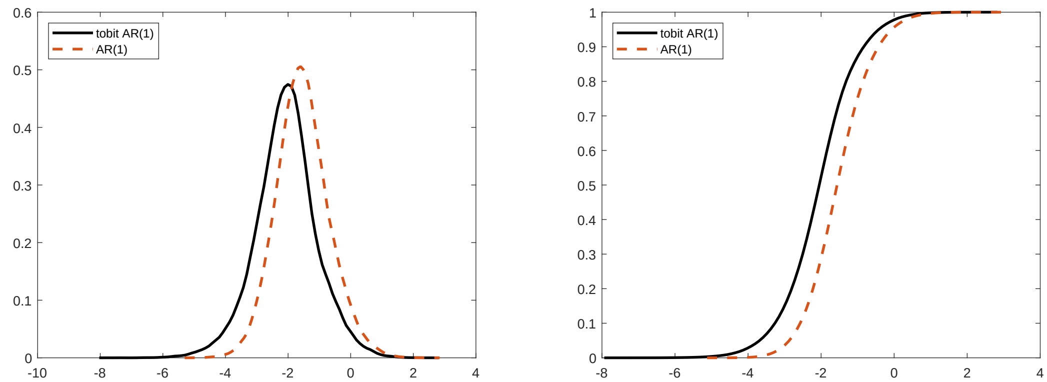


(A) Power envelopes with respect to a. (B) Power envelopes with respect to c.

FIGURE 4.3. Power envelopes with respect to a and c. Data generating process $y_t = \left[\frac{a}{\sqrt{T}} + \left(1 + \frac{c}{T}\right) y_{t-1} + u_t\right]_+$, $y_0 = 0$, $u_t \sim$ i.i.d. $\mathcal{N}(0, 1)$. Data is obtained from 10^5 samples of time series of length $T = 1000$.

a (Figure 4.3a), in such cases an upward trend in $\{y_t\}$ would become discernable, and would carry the process away from the censoring point. Since $\{y_t\}$ would then make few (if any) visits to the censoring point, if one were interested in testing the null of a unit root against a *trend* stationary alternative, in this case, a conventional ADF test with intercept and trend could be employed for this purpose.

4.4. Autoregression vs. dynamic Tobit. Figure 4.4 shows the cumulative distributions (CDF) and probability densities (PDF) for t_β under dynamic Tobit and linear autoregressive data generating processes, with the distribution for the former lying to the left of that for the latter. Thus when data is generated by a dynamic Tobit with a unit root, the conventional ADF test (i.e. that which compares t_β to critical values derived from the linear model) will



(A) Probability density functions. (B) Cumulative distribution functions.

FIGURE 4.4. CDFs and PDFs of t-ratio t_β under linear data generating process with a unit root $y_t = y_{t-1} + u_t$ and dynamic Tobit with a unit root $y_t = [y_{t-1} + u_t]_+$ with $y_0 = 0$, $u_t \sim$ i.i.d. $\mathcal{N}(0, 1)$. Data is obtained from 100,000 samples of time series of length $T = 1000$.

Size	1%	5%	10%
ADF	-3.44	-2.87	-2.57
Tobit ADF ($b_0 = 0$)	-4.53	-3.67	-3.25

TABLE 4.2. Critical values for the ADF test and for Tobit ADF unit root test (based on 10^7 Monte Carlo simulations, $T = 1000$).

tend to over-reject: we find that a nominal 5 per cent test will in fact reject 18 per cent of the time. The intuition for this is that the censoring causes the trajectories of $\{y_t\}$ to appear stationary, masking the presence of a unit root.

Table 4.2 summarises the preceding observations by reporting critical values from the conventional ADF distribution, and from those based on the dynamic Tobit model, as per Corollary 3.2. The former critical values are uniformly larger than the latter, which again indicates that the conventional ADF test is more likely to reject the null of a unit root.

5. Empirical application

In this section we illustrate the use of our methods through an application to testing for unit roots in nominal exchange rates, when these are subject to a one-sided bound. Unit roots are routinely detected in these series by conventional tests in empirical work (see e.g. Baillie and Bollerslev, 1989; Sarno and Valente, 2006, p. 3156; Hong and Phillips, 2010, p. 107). Their presence is also manifested in the robust performance of exchange rate forecasts based on random walks, which more elaborate models have struggled to beat consistently (Rossi, 2013). Here we examine how censoring, as introduced by the deliberate action of a central bank to

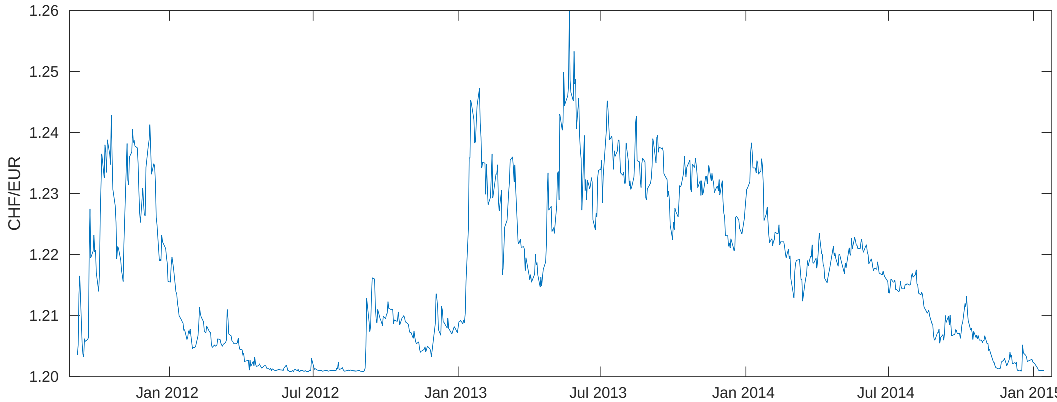


FIGURE 5.1. CHF/EUR exchange rate (Source: ECB).

keep exchange rates above (or below) a nominated threshold, may alter our assessment of the evidence for or against a unit root, depending on whether that censoring is accounted for (cf. Cavaliere, 2005, Sec. 6.1).

In September 2011, in response to the ongoing appreciation of the Swiss franc, and with the policy rate effectively at the zero lower bound, the Swiss National Bank (SNB) instituted a floor on the euro–Swiss franc exchange rate of 1.20 francs per euro (Jordan, 2016; Hertrich, 2022). With the exchange rate well below the floor on the previous day, the immediate intervention of the SNB was required to make the floor effective upon its introduction on 6 September. The floor remained in effect until the end of 15 January 2015. As can be seen from Figure 5.1, during that period, the exchange rate spent most of its time well above the the floor (reaching a peak of 1.26 in May 2013), with two notable exceptions: the periods of April to August 2012, and from November 2014 until the end of the policy, both of which triggered action by the SNB to prevent the floor from being violated (see Figure 9 in Hertrich, 2022). (For a further discussion of the floor and its aftermath, see also von Schweinitz *et al.*, 2021.) The observed trajectory of the exchange rate, during these episodes, is thus more plausibly consistent with a dynamic Tobit than it is with a linear autoregression.

Our data is drawn from the European Central Bank's (ECB) daily reference exchange rate series (code: EXR.D.CHF.EUR.SP00.A) from the period during which the floor was operational (6 Sep 2011 to 15 Jan 2014), transformed by taking logarithms. To select the lag order k used to compute t_β, we evaluated autoregressive models with $k \in \{1, \dots, 15\}$ using the Akaike and Bayesian information criteria; both selected a model with only one lag. For this model, we obtained an ADF statistic slightly above -2.87: so that if the censoring were ignored, and this statistic referred to the conventional ADF critical values (Table 4.2), there would be just sufficient evidence to reject the null of a unit root at the five per cent level. On the other hand, the unit root is not rejected at conventional significance levels under the dynamic Tobit. By simulating the asymptotic distribution of t_β, given in Corollary 3.2, we compute the p value appropriate to the dynamic Tobit as either 0.16 (with $b_0 = 0$ imposed) or 0.18 (using $\hat{b}_0 = T^{-1/2}(y_0 - \mathbf{L})$ with $\mathbf{L} = \log(1.20)$); similar results also obtain when $k = 2, 3$. This places $t_{\beta,T}$ much further from the critical region, thereby lending support to the hypothesis that the data is consistent with a dynamic Tobit model with a unit root.

6. Conclusion

This paper extends local to unity asymptotics to a dynamic Tobit model. Censoring fundamentally changes the analysis and requires new tools to derive the asymptotics. We obtain novel limit theorems for convergence to regulated processes, that is, to processes constrained to lie above a threshold. The effect of that censoring on the limiting distribution of our test statistics varies according to the proximity of the initialisation to the censoring point, with the distributions associated with the linear model re-emerging as that initialisation moves sufficiently far from the censoring point.

Our results underpin the development of a unit root test appropriate to censored data generated by a dynamic Tobit model. In contrast to the setting of a linear model, here the null of a unit root implies restrictions on both α and β, since $\alpha < 0$ (with $\beta = 1$) is consistent with stationarity. Nonetheless, a test of this null can still be effected by the usual t_β statistic, using adjusted critical values. We provide an empirical illustration of our methods to testing for a unit root in nominal exchange rates, when these are subject to a one-sided bound.

The results of this paper could be developed further in a number of directions. One possibility would be to extend our results beyond the local to unity setting, by allowing for moderate deviations from a unit root (Giraitis and Phillips, 2006), with the aim of establishing (uniformly) valid confidence intervals for β, as per Mikusheva (2007, 2012) in the linear model. As we depart from the linear model, new types of asymptotics emerge, which may involve c in different ways (cf. Bykhovskaya and Phillips (2018) where c is no longer a constant but varies with time).

The analysis of $c \to \pm\infty$ for the censored model may be undertaken in conjunction with the derivation of the asymptotics of least absolute deviation (LAD) regression or maximum likelihood estimators of the model, which would enjoy consistency across a wider domain than does OLS. For such extensions, the results of Section 3.1, regarding the asymptotics of $\{y_t\}$, are likely to be of fundamental importance.

For another direction in which our analysis could be extended, recall from the introduction that the dynamic Tobit provides the kernel of more elaborate, multivariate models that allow for the possibility of censoring and/or some other threshold-related nonlinearity. The analysis of (exact) unit roots and cointegration, in the setting of such a model, is developed by Duffy *et al.* (2022), with the aid of our results.

Appendix A. Proofs of results for $k = 1$.

A.1. Limiting distribution of $T^{-1/2} y_{\lfloor nr \rfloor}$.

Lemma A.1. *Suppose that $x_t = [x_{t-1} + v_t]_+$, for $t = 1, \ldots, T$, and $T^{-1/2}(x_0 + \sum_{s=1}^{\lfloor rT \rfloor} v_s) \xrightarrow{d} V(r)$ on $D[0,1]$. Then on $D[0,1]$,*

$$T^{-1/2} x_{\lfloor rT \rfloor} \xrightarrow{d} V(r) + \sup_{r' \le r} [-V(r')]_+.$$

Proof. By Bykhovskaya (2021, Supplementary Material, Lemma D.8) and Cavaliere (2004, Lemma 1),

$$x_t = [x_{t-1} + v_t]_+ = x_0 + \sum_{s=1}^{t} v_s + \sup_{t' \in \{0,\ldots,t\}} \left[-x_0 - \sum_{s=1}^{t'} v_s \right]_+ .$$

Defining $V_T(r) \coloneqq T^{-1/2}(x_0 + \sum_{s=1}^{\lfloor rT \rfloor} v_s)$, we have that $V_T(r) \xrightarrow{d} V(r)$ on $D[0,1]$ by the hypotheses of the lemma. Since

$$T^{-1/2} x_{\lfloor rT \rfloor} = V_T(r) + \sup_{r' \in [0,r]} [-V_T(r')]_+$$

and the supremum on the r.h.s. is a continuous functional of $V_T(\cdot)$, the result follows by the continuous mapping theorem (CMT). □

Proof of Theorem 3.1. Multiplying (3.3) by β^{-t}, and defining $x_t \coloneqq \beta^{-t} y_t$, we have

$$x_t = \beta^{-t} y_t = \beta^{-t}[\beta y_{t-1} + \alpha + u_t]_+ = [\beta^{-(t-1)} y_{t-1} + \beta^{-t}(\alpha + u_t)]_+ = [x_{t-1} + v_t]_+,$$

where $v_t \coloneqq \beta^{-t}(\alpha + u_t)$. Under A1, $T^{-1/2} x_0 = T^{-1/2} y_0 \to b_0$, and so

$$\frac{1}{T^{1/2}} \left(x_0 + \sum_{s=1}^{\lfloor rT \rfloor} v_s \right) = \frac{x_0}{T^{1/2}} + \frac{a}{T} \sum_{s=1}^{\lfloor rT \rfloor} e^{-cs/T} + \frac{1}{T^{1/2}} \sum_{s=1}^{\lfloor rT \rfloor} e^{-cs/T} u_s$$

$$\xrightarrow{d} b_0 + a \int_0^r e^{-cs} \, \mathrm{d}s + \sigma \int_0^r e^{-cs} \, \mathrm{d}W(s) = K_\theta(r) \tag{A.1}$$

on $D[0,1]$, where K_θ is as defined in (3.1), $\theta = (a, b_0, c)$, and the weak convergence follows by the martingale central limit theorem. It follows by Lemma A.1 that on $D[0,1]$,

$$T^{-1/2} y_{\lfloor rT \rfloor} = \beta^{\lfloor rT \rfloor} T^{-1/2} x_{\lfloor rT \rfloor} = e^{c \lfloor rT \rfloor / T} T^{-1/2} x_{\lfloor rT \rfloor} \xrightarrow{d} e^{cr} \left[K_\theta(r) + \sup_{r' \le r} [-K_\theta(r')] \right]. \quad □$$

A.2. OLS asymptotics.

Recall from (2.3) that, in the AR(1) model, $y_t^- = [\alpha + \beta y_{t-1} + u_t]_-$, and that in this case (2.4) specialises to

$$y_t = [\alpha + \beta y_{t-1} + u_t]_+ = \alpha + \beta y_{t-1} + u_t - y_t^- . \tag{A.2}$$

Lemma A.2. *Suppose Assumptions* A1 *and* A2 *hold with $k = 1$. Then*

(i) $T^{-1/2} \sum_{t=1}^{T} (u_t - y_t^-) \xrightarrow{d} J_\theta(1) - b_0 - c \int_0^1 J_\theta(r) \, \mathrm{d}r - a$;

(ii) $\sum_{t=1}^{T}(y_t^-)^2 = o_p(T)$; *and*
(iii) $T^{-1}\sum_{t=1}^{T}(\Delta y_t)^2 \xrightarrow{p} \sigma^2$.

Proof. **(i).** We first note from (A.2) that

$$\sum_{t=1}^{T}(u_t - y_t^-) = \sum_{t=1}^{T}(y_t - \alpha - \beta y_{t-1}) = \sum_{t=1}^{T}(y_t - y_{t-1}) - (\beta - 1)\sum_{t=1}^{T} y_{t-1} - T\alpha$$
$$= y_T - y_0 - (\beta - 1)\sum_{t=1}^{T} y_{t-1} - T^{1/2}a.$$

Since $T(\beta - 1) = c + o(1)$, it follows from Theorem 3.1 and the CMT that, under A1,

$$\frac{1}{T^{1/2}}\sum_{t=1}^{T}(u_t - y_t^-) = \frac{1}{T^{1/2}}(y_T - y_0) - \frac{c + o(1)}{T^{3/2}}\sum_{t=1}^{T} y_{t-1} - a$$
$$\xrightarrow{d} J_\theta(1) - b_0 - c\int_0^1 J_\theta(r)\,\mathrm{d}r - a.$$

(ii). Since $\beta \geq 0$ and $y_{t-1} \geq 0$,

$$0 \geq y_t^- = (\alpha + \beta y_{t-1} + u_t)\mathbf{1}\{\alpha + \beta y_{t-1} + u_t \leq 0\} \geq v_t \mathbf{1}\{v_t \leq -\beta y_{t-1}\},$$

where $v_t := \alpha + u_t$. Hence

$$\max_{1\leq t\leq T}|y_t^-| \leq \max_{1\leq t\leq T}|v_t| \leq \frac{|a|}{T^{1/2}} + \max_{1\leq t\leq T}|u_t| = o_p(T^{1/2})$$

where the final equality holds since $\{u_t\}$ is i.i.d. with finite variance, under Assumption A2.1. Further, by the result of part (i),

$$\sum_{t=1}^{T} y_t^- = \sum_{t=1}^{T} u_t + O_p(T^{1/2}) = O_p(T^{1/2}).$$

Hence

$$\sum_{t=1}^{T}(y_t^-)^2 \leq \max_{1\leq t\leq T}|y_t^-|\sum_{t=1}^{T}|y_t^-| = -\max_{1\leq t\leq T}|y_t^-|\sum_{t=1}^{T} y_t^- = o_p(T^{1/2})O_p(T^{1/2}) = o_p(T).$$

(iii). Since

$$\sum_{t=1}^{T}\alpha^2 = a^2 = O(1), \qquad \sum_{t=1}^{T}\alpha(\beta - 1)y_{t-1} = O_p(1), \qquad \sum_{t=1}^{T}[(\beta - 1)y_{t-1}]^2 = O_p(1)$$

by Theorem 3.1 and the CMT;

$$\sum_{t=1}^{T}\alpha(u_t - y_t^-) = O_p(1),$$

by Lemma A.2(i); and

$$\frac{1}{T}\sum_{t=1}^{T}(u_t - y_t^-)^2 = \frac{1}{T}\sum_{t=1}^{T}u_t^2 - \frac{2}{T}\sum_{t=1}^{T}y_t^- u_t + \frac{1}{T}\sum_{t=1}^{T}(y_t^-)^2 \xrightarrow{p} \sigma^2, \tag{A.3}$$

by the law of large numbers, the Cauchy–Schwarz (CS) inequality, and the result of part (ii); and

$$\left|\sum_{t=1}^{T}(\beta-1)y_{t-1}(u_t - y_t^-)\right| \le \sqrt{\sum_{t=1}^{T}[(\beta-1)y_{t-1}]^2 \sum_{t=1}^{T}(u_t - y_t^-)^2} = O_p(\sqrt{T}),$$

by the preceding; it follows that

$$\frac{1}{T}\sum_{t=1}^{T}(\Delta y_t)^2 = \frac{1}{T}\sum_{t=1}^{T}(\alpha + (\beta-1)y_{t-1} + u_t - y_t^-)^2 = \frac{1}{T}\sum_{t=1}^{T}(u_t - y_t^-)^2 + o_p(1) \xrightarrow{p} \sigma^2. \qquad \square$$

Proof of Theorem 3.3. We have from (A.2) that

$$\begin{bmatrix} \hat{\alpha}_T - \alpha \\ \hat{\beta}_T - \beta \end{bmatrix} = \left(\sum_{t=1}^{T}\begin{bmatrix} 1 & y_{t-1} \\ y_{t-1} & y_{t-1}^2 \end{bmatrix}\right)^{-1} \sum_{t=1}^{T}\begin{bmatrix} 1 \\ y_{t-1} \end{bmatrix}(u_t - y_t^-) \tag{A.4}$$

where

$$\frac{1}{T^{1/2}}\sum_{t=1}^{T}(u_t - y_t^-) \xrightarrow{d} J_\theta(1) - b_0 - c\int_0^1 J_\theta(r)\,\mathrm{d}r - a \tag{A.5}$$

by Lemma A.2(i). To obtain the weak limit of $\sum_{t=1}^{T} y_{t-1}(u_t - y_t^-)$, we note that since only one of y_t and y_t^- can be nonzero, $y_t y_t^- = 0$, and hence $y_t^- y_{t-1} = -y_t^- \Delta y_t$. Thus by the CS inequality and Lemma A.2(ii)–(iii),

$$\left|\sum_{t=1}^{T} y_t^- y_{t-1}\right| = \left|\sum_{t=1}^{T} y_t^- \Delta y_t\right| \le \left[\sum_{t=1}^{T}(y_t^-)^2 \sum_{t=1}^{T}(\Delta y_t)^2\right]^{1/2} = o_p(T).$$

It follows by the preceding and Liang *et al.* (2016, Theorem 2.1) that

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}(u_t - y_t^-) = \frac{1}{T}\sum_{t=1}^{T} y_{t-1}u_t + o_p(1) \xrightarrow{d} \sigma\int_0^1 J_\theta(r)\,\mathrm{d}W(r). \tag{A.6}$$

In view of (A.4)–(A.6), a final appeal to Theorem 3.1 and the CMT yields

$$\begin{aligned}\begin{bmatrix} T^{1/2}(\hat{\alpha}_T - \alpha) \\ T(\hat{\beta}_T - \beta) \end{bmatrix} &= \left(\sum_{t=1}^{T}\begin{bmatrix} T^{-1} & T^{-3/2}y_{t-1} \\ T^{-3/2}y_{t-1} & T^{-2}y_{t-1}^2 \end{bmatrix}\right)^{-1} \sum_{t=1}^{T}\begin{bmatrix} T^{-1/2} \\ T^{-1}y_{t-1} \end{bmatrix}(u_t - y_t^-) \\ &\xrightarrow{d} \begin{bmatrix} 1 & \int J_\theta(r)\,\mathrm{d}r \\ \int J_\theta(r)\,\mathrm{d}r & \int J_\theta^2(r)\,\mathrm{d}r \end{bmatrix}^{-1} \begin{bmatrix} J_\theta(1) - b_0 - c\int J_\theta(r)\,\mathrm{d}r - a \\ \sigma\int J_\theta(r)\,\mathrm{d}W(r) \end{bmatrix}. \qquad \square\end{aligned}$$

Verification of (3.9). By the Frisch-Waugh-Lovell theorem, and using partial summation (as in the proof of Theorem 3.1 in Phillips (1987a)) we have

$$\begin{aligned}\hat{\beta}_T - 1 &= \frac{\sum_{t=1}^T y_{t-1}^\mu y_t^\mu}{\sum_{t=1}^T (y_{t-1}^\mu)^2} - 1 = \frac{\sum_{t=1}^T y_{t-1}^\mu \Delta y_t^\mu}{\sum_{t=1}^T (y_{t-1}^\mu)^2} \\ &= \frac{\sum_{t=1}^T \left((y_{t-1}^\mu + \Delta y_t^\mu)^2 - (y_{t-1}^\mu)^2 - (\Delta y_t^\mu)^2\right)}{2\sum_{t=1}^T (y_{t-1}^\mu)^2} = \frac{(y_T^\mu)^2 - (y_0^\mu)^2 - \sum_{t=1}^T (\Delta y_t^\mu)^2}{2\sum_{t=1}^T (y_{t-1}^\mu)^2}\end{aligned} \tag{A.7}$$

where $y_t^\mu \coloneqq y_t - \frac{1}{T}\sum_{t=1}^T y_{t-1}$. (3.9) then follows (noting the centring here is around 1, rather than β) by Theorem 3.1, Lemma A.2(iii), and the CMT. □

Proof of Corollary 3.2 ($k = 1$). Once we have shown that $\hat{\sigma}_T^2 \xrightarrow{p} \sigma^2$, the limiting distributions of the t statistics will follow from Theorem 3.1 and the CMT. Noting from (A.2) that

$$\hat{u}_t = y_t - \hat{\alpha}_T - \hat{\beta}_T y_{t-1} = (\alpha - \hat{\alpha}_T) + (\beta - \hat{\beta}_T) y_{t-1} + (u_t - y_t^-),$$

and that $\sum_{t=1}^T \hat{u}_t = 0$ and $\sum_{t=1}^T y_{t-1}\hat{u}_t = 0$, we have

$$\begin{aligned}\sum_{t=1}^T [\hat{u}_t^2 - (u_t - y_t^-)^2] &= \sum_{t=1}^T [\hat{u}_t - (u_t - y_t^-)][\hat{u}_t + (u_t - y_t^-)] \\ &= \sum_{t=1}^T [(\alpha - \hat{\alpha}_T) + (\beta - \hat{\beta}_T) y_{t-1}][\hat{u}_t + (u_t - y_t^-)] \\ &= (\alpha - \hat{\alpha}_T)\sum_{t=1}^T (u_t - y_t^-) + (\beta - \hat{\beta}_T)\sum_{t=1}^T y_{t-1}(u_t - y_t^-) \\ &= O_p(T^{-1/2})O_p(T^{1/2}) + O_P(T^{-1})O_p(T) = O_p(1)\end{aligned}$$

where the orders of $\sum_{t=1}^T (u_t - y_t^-)$ and $\sum_{t=1}^T y_{t-1}(u_t - y_t^-)$ follow from (A.5) and (A.6) above, and the rates of convergence of $\hat{\alpha}_T$ and $\hat{\beta}_T$ from Theorem 3.3. Hence, by (A.3),

$$\hat{\sigma}_T^2 = \frac{1}{T}\sum_{t=1}^T \hat{u}_t^2 = \frac{1}{T}\sum_{t=1}^T (u_t - y_t^-)^2 + O_p(T^{-1}) \xrightarrow{p} \sigma^2.$$ □

Appendix B. Proofs of results for general k

B.1. Limiting distribution of $T^{-1/2} y_{\lfloor nr \rfloor}$: AR($k$) case. Let ρ denote the *inverse* of the root of $B(z)$ closest to real unity, which for T sufficiently large must be real because A2 permits $B(z)$ to have only one root local to unity. Thus $B(z)$ factorises as

$$B(z) = (1 - \beta)z + \phi(z)(1 - z) = \psi(z)(1 - \rho z) \tag{B.1}$$

for $z \in \mathbb{C}$ where $\psi(z) = 1 - \sum_{i=1}^{k-1} \psi_i z^i$. Under A2.2, $\beta = \beta_T \to 1$ and, thus, $\rho = \rho_T \to 1$, from which it follows that $\psi_i \to \phi_i$ for $i \in \{1, \ldots, k-1\}$, as $T \to \infty$.[4] Thus for T sufficiently large, ρ is real and positive (as we shall maintain throughout the following), and such a condition as A4 also holds when each ϕ_i in (3.5) is replaced by ψ_i. Moreover, taking $z = \rho_T^{-1}$ in the preceding, it follows that

$$\begin{aligned} 0 &= B(\rho_T^{-1}) = (1-\beta_T)\rho_T^{-1} + \phi(\rho_T^{-1})(1-\rho_T^{-1}) \\ &\Leftrightarrow T(\rho_T - 1) = \phi(\rho_T^{-1})^{-1} T(\beta_T - 1) \to \phi(1)^{-1} c =: c_\phi. \end{aligned} \tag{B.2}$$

The factorisation (B.1) also permits us to rewrite the model (2.5) for $\{y_t\}$ in terms of the quasi-differences $\Delta_\rho y_t$,

$$\psi(L)\Delta_\rho y_t = \alpha + u_t - y_t^-, \tag{B.3}$$

where $\Delta_\rho \coloneqq 1 - \rho L$. With the aid of this representation we establish the following preliminary lemmas.

Lemma B.1. *Suppose Assumption* A2 *holds, and define*

$$x_t \coloneqq \psi(\rho^{-1})\rho^{-t} y_t \qquad \xi_t \coloneqq \sum_{s=1}^{t} \rho^{-s}(\alpha + u_s) - \gamma_\rho(L)[\rho^{-t}\Delta_\rho y_t - \Delta_\rho y_0] \tag{B.4}$$

for $t \in \{1, \ldots, T\}$, *where* $\gamma_\rho(L)$ *is the* $k-2$ *order polynomial such that* $\psi(L) = \psi(\rho^{-1}) + \gamma_\rho(L)\Delta_\rho$ *(with* $\psi(L) = 1$ *and* $\gamma_\rho(L) = 0$ *when* $k = 1$*). Then for* $t \in \{1, \ldots, T\}$,

$$x_t = [x_{t-1} + \Delta\xi_t]_+. \tag{B.5}$$

Proof. Observe that for any series $\{\eta_s\}$,

$$\sum_{s=1}^{t} \rho^{t-s}\Delta_\rho \eta_s = \sum_{s=1}^{t} \rho^{t-s}(\eta_s - \rho\eta_{s-1}) = \sum_{s=1}^{t} \rho^{t-s}\eta_s - \sum_{s=1}^{t} \rho^{t-(s-1)}\eta_{s-1} = \eta_t - \rho^t \eta_0. \tag{B.6}$$

Applying this to $\psi(L)\Delta_\rho y_s = \alpha + u_s - y_s^-$ (from (B.3) above), we obtain

$$\psi(L)y_t - \rho^t \psi(L) y_0 = \sum_{s=1}^{t} \rho^{t-s}(\alpha + u_s) - \sum_{s=1}^{t} \rho^{t-s} y_s^-.$$

Using $\psi(L) = \psi(\rho^{-1}) + \gamma_\rho(L)\Delta_\rho$, and rearranging yields

$$\psi(\rho^{-1})y_t + \sum_{s=1}^{t} \rho^{t-s} y_s^- = \sum_{s=1}^{t} \rho^{t-s}(\alpha + u_s) - \gamma_\rho(L)(\Delta_\rho y_t - \rho^t \Delta_\rho y_0) + \rho^t \psi(\rho^{-1}) y_0.$$

[4]Formally, one should write $\psi_{i,T}$, since the coefficients of the polynomial $\psi(z)$ depend on β, which in turn varies with T. We omit this dependence on T for ease of notation.

Finally, multiplying by ρ^{-t} and recalling the definitions of (x_t, ξ_t) from (B.4), we have

$$x_t + \sum_{s=1}^{t} \rho^{-s} y_s^- = x_0 + \xi_t. \tag{B.7}$$

To proceed from (B.7) to show that (x_t, ξ_t) satisfy (B.5), we note that for $t \geq 1$,

$$\begin{aligned}\xi_{t+1} = x_{t+1} - x_0 + \sum_{s=1}^{t+1} \rho^{-s} y_s^- &= x_{t+1} - x_0 + \rho^{-(t+1)} y_{t+1}^- + \sum_{s=1}^{t} \rho^{-s} y_s^- \\ &= x_{t+1} + \rho^{-(t+1)} y_{t+1}^- + \xi_t - x_t\end{aligned}$$

where the first and last equalities follow from (B.7). Hence

$$x_{t+1} + \rho^{-(t+1)} y_{t+1}^- = x_t + \Delta\xi_{t+1}.$$

From (2.1) and (2.3), at most one of y_{t+1} and y_{t+1}^- can be nonzero, and must have opposite signs. Since $\psi(\rho^{-1}) \to \phi(1) > 0$ (due to stationarity), we must have $\psi(\rho^{-1})\rho^{-t} > 0$ for all T sufficiently large. The same must also be true for $x_{t+1} = \psi(\rho^{-1})\rho^{-(t+1)} y_{t+1}$ and $\rho^{-(t+1)} y_{t+1}^-$. Hence,

$$x_{t+1} = [x_t + \Delta\xi_{t+1}]_+$$

for $t \geq 1$. Plugging $\xi_0 = 0$ into (B.7) when $t = 1$, we have

$$x_1 + \rho^{-1} y_1^- = x_0 + \xi_1 = x_0 + \Delta\xi_1$$

and thus $x_1 = [x_0 + \Delta\xi_1]_+$, by the same argument. □

Lemma B.2. *Suppose Assumptions* A1–A4 *hold. Then there exists a $C < \infty$ such that*

$$\max_{-k+2 \leq t \leq T} (\|\Delta_\rho y_t\|_{2+\delta_u} + \|\Delta y_t\|_{2+\delta_u}) < C.$$

Proof. We have from (B.3) that

$$\Delta_\rho y_t + y_t^- = \sum_{i=1}^{k-1} \psi_i \Delta_\rho y_{t-i} + \alpha + u_t =: w_t, \tag{B.8}$$

for $t \in \{1, \ldots, T\}$. Since, as noted in the proof of Lemma B.1, only one of y_t and y_t^- can be nonzero, and have opposite signs,

$$\Delta_\rho y_t > 0 \implies y_t > \rho y_{t-1} \geq 0 \implies y_t^- = 0.$$

It follows that either $w_t > 0$, in which case $\Delta_\rho y_t = w_t$; or $w_t \leq 0$, in which case $\Delta_\rho y_t \in [w_t, 0]$. Hence there exists a $\delta_t \in [0, 1]$ such that $\Delta_\rho y_t = \delta_t w_t$ for $t \in \{1, \ldots, T\}$. Taking $w_0 = \Delta_\rho y_0$ and $\delta_0 = 1$, (B.8) is equivalent to a dynamical system defined by

$$w_t = \psi_1 \delta_{t-1} w_{t-1} + \sum_{i=2}^{k-1} \psi_i \Delta_\rho y_{t-i} + \alpha + u_t, \tag{B.9}$$

$$\Delta_\rho y_{t-1} = \delta_{t-1} w_{t-1} \tag{B.10}$$

for $t \in \{1, \dots, T\}$, for an appropriate sequence $\{\delta_t\} \subset [0, 1]$. Defining

$$\boldsymbol{w}_t := \begin{bmatrix} w_t \\ \Delta_\rho y_{t-1} \\ \vdots \\ \Delta_\rho y_{t-k+2} \end{bmatrix} \qquad \boldsymbol{v}_t := \begin{bmatrix} \alpha + u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad F_\delta(\boldsymbol{\psi}) := \begin{bmatrix} \psi_1\delta & \psi_2 & \cdots & \psi_{k-2} & \psi_{k-1} \\ \delta & 0 & \cdots & 0 & 0 \\ 0 & 1 & & & \\ & & \ddots & & \\ & & & 1 & 0 \end{bmatrix}$$

where $\boldsymbol{\psi} := (\psi_1, \dots, \psi_{k-1})$, we can write the companion form of (B.9)–(B.10) as

$$\boldsymbol{w}_t = F_{\delta_{t-1}}(\boldsymbol{\psi})\boldsymbol{w}_{t-1} + \boldsymbol{v}_t$$

for $t \in \{1, \dots, T\}$, with the initialisation $\boldsymbol{w}_0 := (\Delta_\rho y_0, \Delta_\rho y_{-1}, \dots, \Delta_\rho y_{-k+2})^\mathsf{T}$.

Since $\boldsymbol{\psi} \to \boldsymbol{\phi} = (\phi_1, \dots, \phi_{k-1})^\mathsf{T}$ as $T \to \infty$, $F_\delta(\boldsymbol{\psi}) \to F_\delta(\boldsymbol{\phi}) = F_\delta$, where F_δ is as defined in (3.5). By Proposition 1.8 in Jungers (2009) and the continuity of the JSR,

$$\lambda_{\mathrm{JSR}}(\{F_\delta(\boldsymbol{\psi}) \mid \delta \in [0, 1]\}) = \lambda_{\mathrm{JSR}}(\{F_0(\boldsymbol{\psi}), F_1(\boldsymbol{\psi})\}) \to \lambda_{\mathrm{JSR}}(\{F_0, F_1\}) < 1$$

under A4. It follows that there exists a norm $\|\cdot\|_*$ and a $\gamma \in [0, 1)$ such that (for all T sufficiently large),

$$\|\boldsymbol{w}_t\|_* \le \|F_{\delta_{t-1}}\|_* \|\boldsymbol{w}_{t-1}\|_* + \|\boldsymbol{v}_t\|_* \le \gamma \|\boldsymbol{w}_{t-1}\|_* + \|\boldsymbol{v}_t\|_*$$

whence by backward substitution,

$$\|\boldsymbol{w}_t\|_* \le \sum_{s=0}^{t-1} \gamma^s \|\boldsymbol{v}_{t-s}\|_* + \gamma^{t-1} \|\boldsymbol{w}_0\|_*.$$

By the equivalence of norms on finite-dimensional spaces, there exists a $C < \infty$ such that

$$|\Delta_\rho y_{t-1}| \le C \left[\sum_{s=0}^{t-1} \gamma^s (|\alpha| + |u_{t-s}|) + \gamma^{t-1} \sum_{i=-k+2}^{0} |\Delta_\rho y_i| \right].$$

Deduce that for any $p \ge 1$,

$$\begin{aligned} \|\Delta_\rho y_{t-1}\|_p &\le \frac{C|\alpha|}{1-\gamma} + C \sum_{s=0}^{t-1} \gamma^s \|u_{t-s}\|_p + C \sum_{i=-k+2}^{0} \|\Delta_\rho y_i\|_p \\ &\le \frac{C|\alpha|}{1-\gamma} + \frac{C}{1-\gamma} \max_{1 \le s \le t} \|u_s\|_p + C \sum_{i=-k+2}^{0} \|\Delta_\rho y_i\|_p. \end{aligned} \tag{B.11}$$

Now for each $i \in \{-k+2, \ldots, 0\}$, $\Delta_\rho y_i = (1-\rho)[y_0 - \sum_{j=i+1}^{0} \Delta y_j] + \rho \Delta y_i$, and hence there exists a $C' < \infty$ such that

$$\|\Delta_\rho y_i\|_{2+\delta_u} \leq C' \left[\frac{1}{T^{1/2}} \|T^{-1/2} y_0\|_{2+\delta_u} + \frac{1}{T} \sum_{j=i+1}^{0} \|\Delta y_j\|_{2+\delta_u} + \|\Delta y_i\|_{2+\delta_u} \right],$$

where we have used that $1 - \rho = O(T^{-1})$. Taking $p = 2 + \delta_u$ in (B.11), it follows under A3 that $\max_{-k+2 \leq t \leq T} \|\Delta_\rho y_t\|_{2+\delta_u}$ is bounded uniformly in T. To obtain the corresponding result for Δy_t, we use (B.6) with $\eta_t = y_t$ to write

$$\Delta y_t = \Delta_\rho y_t + (\rho - 1) y_{t-1} = \Delta_\rho y_t + (\rho - 1) \left[\sum_{s=1}^{t-1} \rho^{t-1-s} \Delta_\rho y_s + \rho^{t-1} y_0 \right]$$

whence there exists a $C'' < \infty$ such that

$$\|\Delta y_t\|_{2+\delta_u} \leq \|\Delta_\rho y_t\|_{2+\delta_u} + C'' \left[T^{-1} \max_{1 \leq s \leq t} \|\Delta_\rho y_s\|_{2+\delta_u} + T^{-1/2} \|T^{-1/2} y_0\|_{2+\delta_u} \right],$$

from which, under A3, the result follows. □

Proof of Theorem 3.2. When $k = 1$, the result follows by Theorem 3.1; we therefore suppose $k \geq 2$. We first note that by (B.2) above, $\rho^{\lfloor rT \rfloor} \to e^{c_\phi r}$ uniformly in $r \in [0, 1]$. Hence for (x_t, ξ_t) as in Lemma B.1,

$$T^{-1/2} \left[x_0 + \sum_{s=1}^{\lfloor rT \rfloor} \Delta \xi_s \right] = T^{-1/2} x_0 + T^{-1/2} \xi_{\lfloor rT \rfloor}$$

$$=_{(1)} \psi(\rho^{-1}) T^{-1/2} y_0 + T^{-1/2} \sum_{s=1}^{\lfloor rT \rfloor} \rho^{-s} (\alpha + u_s) + o_p(1) \xrightarrow{d}_{(2)} K_{\theta_\phi}(r), \tag{B.12}$$

on $D[0, 1]$, where $\theta_\phi = (a, \phi(1) b_0, \phi(1)^{-1} c)$, $=_{(1)}$ holds since Lemma B.2 implies that $\max_{0 \leq t \leq T} |\Delta_\rho y_t| = o_p(T^{1/2})$, and $\xrightarrow{d}_{(2)}$ holds by the same arguments as which yielded (A.1) above and by recalling that $\psi(\rho^{-1}) \to \phi(1)$. Hence by Lemma B.1, x_t and $v_t := \Delta \xi_t$ satisfy the requirements of Lemma A.1. We thus have

$$\begin{aligned} T^{-1/2} y_{\lfloor rT \rfloor} &= \psi(\rho^{-1})^{-1} \rho^{\lfloor rT \rfloor} T^{-1/2} x_{\lfloor rT \rfloor} \\ &\xrightarrow{d} \phi(1)^{-1} \mathrm{e}^{c_\phi r} \left\{ K_{\theta_\phi}(r) + \sup_{r' \leq r} [-K_{\theta_\phi}(r')] \right\} = \phi(1)^{-1} J_{\theta_\phi}(r) \end{aligned}$$

on $D[0, 1]$. □

B.2. OLS asymptotics: AR(k) case. Since, by Assumption A2.3, all the roots of $\phi(z) = 1 - \sum_{i=1}^{k-1} \phi_i z^i$ lie strictly outside the unit circle, there exists a sequence $\{\varphi_i\}_{i=0}^{\infty}$ with $\varphi_0 = 1$ and $\sum_{0=1}^{\infty} |\varphi_i| < \infty$ such that $\phi^{-1}(z) := \sum_{i=0}^{\infty} \varphi_i z^i$ satisfies $\phi^{-1}(z)\phi(z) = 1$ for all $|z| \leq 1$.

Moreover, there exists a $C < \infty$ and a $\gamma_\phi \in (0,1)$ such that $|\varphi_i| < C\gamma_\phi^i$ for all $i \geq 0$. (See e.g., Brockwell and Davis (1991, Section 3.3).)

Lemma B.3. *Let $\phi_m^{-1}(z) \coloneqq \sum_{i=0}^m \varphi_i z^i$, where $m \geq 1$. Then there exists a $C < \infty$, independent of m, and $\{d_{m,i}\}_{i=1}^{k-1}$ such that*

$$\phi_m^{-1}(z)\phi(z) = 1 - z^m \sum_{i=1}^{k-1} d_{m,i} z^i \eqqcolon 1 - d_m(z) z^m$$

for all $|z| \leq 1$ and $m \in \mathbb{N}$, and where $|d_{m,i}| \leq C\gamma_\phi^m$.

Proof. Since

$$1 = \phi^{-1}(z)\phi(z) = \phi(z)\left[\phi_m^{-1}(z) + \sum_{i=m+1}^{\infty} \varphi_i z^i\right]$$

for $|z| \leq 1$, it follows that

$$1 - \phi(z)\sum_{i=m+1}^{\infty} \varphi_i z^i = \phi(z)\phi_m^{-1}(z) = \left(1 - \sum_{i=1}^{k-1} \phi_i z^i\right)\sum_{i=0}^{m} \varphi_i z^i \eqqcolon 1 - \sum_{i=1}^{m+k-1} \vartheta_{m,i} z^i,$$

using that $\varphi_0 = 1$. Matching coefficients on the left and right hand sides, we obtain $\vartheta_{m,i} = 0$ for all $i \in \{1, \ldots, m\}$; while for $i \in \{m+1, \ldots, m+k-1\}$, $\vartheta_{m,i} = \sum_{j=0}^{i-(m+1)} \phi_j \varphi_{i-j}$. Taking $d_{m,i} \coloneqq \vartheta_{m,m+i} = \sum_{j=0}^{i-1} \phi_j \varphi_{m+i-j}$ for $i \in \{1, \ldots, k-1\}$, and noting that

$$|d_{m,i}| \leq \sum_{j=i}^{k-1} |\phi_j||\varphi_{m+i-j}| \leq C\gamma_\phi^m \sum_{j=1}^{k-1} |\phi_j|$$

yields the result. □

Lemma B.4. *Suppose Assumptions* A1–A4 *hold. Then for each $s \in \{0, \ldots, k-1\}$,*

(i) $T^{-1/2}\sum_{t=1}^T (u_t - y_t^-) \xrightarrow{d} \phi(1)[Y_{\theta_\phi}(1) - c_\phi \int Y_{\theta_\phi}(r) - b_0] - a$;
(ii) $\sum_{t=1}^T (y_t^-)^2 = o_p(T)$;
(iii) $T^{-1}\sum_{t=1}^T \Delta y_t \Delta y_{t-s} \xrightarrow{p} \sigma^2 \sum_{n=0}^\infty \varphi_n \varphi_{n+s}$; *and*
(iv) $\sum_{t=1}^T y_t \Delta y_{t-s} = O_p(T)$.

Proof. **(i).** Applying the factorisation (B.1) to (2.5), we get

$$u_t - y_t^- = -\alpha - (\beta - 1)y_{t-1} + \phi(L)\Delta y_t \tag{B.13}$$

and hence, recalling that $\phi(1) = 1 - \sum_{i=1}^{k-1} \phi_i$, $\alpha = T^{1/2}a$, and $T(\beta - 1) \to c$ under A2,

$$\begin{aligned}\frac{1}{T^{1/2}}\sum_{t=1}^T (u_t - y_t^-) &= -a - \frac{T(\beta-1)}{T^{3/2}}\sum_{t=1}^T y_{t-1} + \frac{1}{T^{1/2}}\left(\sum_{t=1}^T \Delta y_t - \sum_{i=1}^{k-1}\phi_i \sum_{t=1}^T \Delta y_{t-i}\right) \\ &= -a - \frac{T(\beta-1)}{T^{3/2}}\sum_{t=1}^T y_{t-1} + \frac{y_T - y_0}{T^{1/2}} - \sum_{i=1}^{k-1}\phi_i \frac{y_{T-i} - y_{0-i}}{T^{1/2}}\end{aligned}$$

$$\xrightarrow{d} -a - c\int_0^1 Y_{\theta_\phi}(r)\,\mathrm{d}r + \phi(1)[Y_{\theta_\phi}(1) - b_0]$$
$$= \phi(1)\left[Y_{\theta_\phi}(1) - c_\phi\int_0^1 Y_{\theta_\phi}(r)\,\mathrm{d}r - b_0\right] - a,$$

where convergence holds by Assumption A1, Theorem 3.2 and the CMT, recalling that $c_\phi = \phi(1)^{-1}c$.

(ii). The argument is analogous to that used to prove Lemma A.2(ii). Rewriting the factorisation (B.1) as

$$\beta z + (1-z)\sum_{i=1}^{k-1}\phi_i z^i = 1 - (1-\rho z)\left[1 - \sum_{i=1}^{k-1}\psi_i z^i\right] = \rho z + \sum_{i=1}^{k-1}\psi_i z^i(1-\rho z)$$

we have from (2.5) that

$$y_t^- = \left[\beta y_{t-1} + \sum_{i=1}^{k-1}\phi_i\Delta y_{t-i} + \alpha + u_t\right]_- = \left[\rho y_{t-1} + \sum_{i=1}^{k-1}\psi_i\Delta_\rho y_{t-i} + \alpha + u_t\right]_- = [\rho y_{t-1} + v_t]_-,$$

where $v_t \coloneqq \sum_{i=1}^{k-1}\psi_i\Delta_\rho y_{t-i} + \alpha + u_t$. Since $\rho y_{t-1} \geq 0$, it follows that

$$0 \geq y_t^- = (\rho y_{t-1} + v_t)\mathbf{1}\{\rho y_{t-1} + v_t \leq 0\} \geq v_t\mathbf{1}\{\rho y_{t-1} + v_t \leq 0\}.$$

By Lemma B.2, $\|\Delta_\rho y_t\|_{2+\delta_u}$ is bounded uniformly in t. Hence, under A3, so too is

$$\|v_t\|_{2+\delta_u} \leq \sum_{i=1}^{k-1}|\psi_i|\|\Delta_\rho y_{t-i}\|_{2+\delta_u} + |\alpha| + \|u_t\|_{2+\delta_u}.$$

whence it follows that $\max_{1\leq t\leq T}|y_t^-| \leq \max_{1\leq t\leq T}|v_t| = o_p(T^{1/2})$. Hence, using the result of part (i),

$$\sum_{t=1}^T (y_t^-)^2 \leq \max_{1\leq t\leq T}|y_t^-|\sum_{t=1}^T|y_t^-| = -\max_{1\leq t\leq T}|y_t^-|\sum_{t=1}^T y_t^-$$
$$= -\max_{1\leq t\leq T}|y_t^-|\left(\sum_{t=1}^T u_t + O_p(T^{1/2})\right) = o_p(T).$$

(iii). Let $s \in \{0,\ldots,k-1\}$. By Lemma B.3, applying $\phi_{t-1}^{-1}(L) = \sum_{i=0}^{t-1}\varphi_i L^i$ to both sides of (B.13) yields

$$\Delta y_t - d_{t-1}(L)\Delta y_1 = \phi_{t-1}^{-1}(L)\phi(L)\Delta y_t = \sum_{i=0}^{t-1}\varphi_i L^i[(\beta-1)y_{t-1} + \alpha + u_t - y_t^-]$$

and hence for $t \in \{1,\ldots,T\}$,

$$\Delta y_t = \sum_{i=0}^{t-1}\varphi_i(\alpha + u_{t-i}) + (\beta-1)\sum_{i=0}^{t-1}\varphi_i y_{t-1-i} - \sum_{i=0}^{t-1}\varphi_i y_{t-i}^- + d_{t-1}(L)\Delta y_1$$

$$=: w_t + r_{1,t} + r_{2,t} + r_{3,t}. \tag{B.14}$$

Decompose

$$w_t = \sum_{i=0}^{t-1} \varphi_i(\alpha + u_{t-i}) = \sum_{i=0}^{\infty} \varphi_i u_{t-i} - \sum_{i=t}^{\infty} \varphi_i u_{t-i} + \frac{a}{T^{1/2}} \sum_{i=0}^{t-1} \varphi_i =: \eta_t + r_{4,t} + r_{5,t}.$$

The sequence $\{\eta_t\}$ is a stationary linear process whose coefficients decay exponentially; hence by Phillips and Solo (1992, Thm. 3.7 and Rem. 3.9),

$$\frac{1}{T} \sum_{t=s+1}^{T} \eta_t \eta_{t-s} \xrightarrow{p} \sigma^2 \sum_{n=0}^{\infty} \varphi_n \varphi_{n+s}.$$

Since for all i, $|\varphi_i| < C\gamma_\phi^i$, $\gamma_\phi \in (0, 1)$, we have

$$\mathbb{E}\left[\frac{1}{T} \sum_{t=1}^{T} r_{4,t}^2\right] = \frac{\sigma^2}{T} \sum_{t=1}^{T} \sum_{i=t}^{\infty} \varphi_i^2 = O(T^{-1}), \qquad \frac{1}{T} \sum_{t=1}^{T} r_{5,t}^2 = \frac{a^2}{T^2} \sum_{t=1}^{T} \left(\sum_{i=0}^{t-1} \varphi_i\right)^2 = O(T^{-1}).$$

It follows by the CS inequality that $\frac{1}{T}\sum_{t=s+1}^{T}(\eta_t r_{\ell,t-s} + \eta_{t-s} r_{\ell,t}) = o_p(1)$ for $\ell \in \{4, 5\}$ and $\frac{1}{T}\sum_{t=s+1}^{T} r_{\ell,t} r_{\ell',t-s} = o_p(1)$ for $\ell, \ell' \in \{4, 5\}$, and hence

$$\frac{1}{T} \sum_{t=s+1}^{T} w_t w_{t-s} = \frac{1}{T} \sum_{t=s+1}^{T} \eta_t \eta_{t-s} + o_p(1) \xrightarrow{p} \sigma^2 \sum_{n=0}^{\infty} \varphi_n \varphi_{n+s} \tag{B.15}$$

for each $s \in \{0, \ldots, k-1\}$. Thus, if we can show that

$$\sum_{t=1}^{T} r_{1,t}^2 = O_p(1) \qquad \sum_{t=1}^{T} r_{2,t}^2 = o_p(T) \qquad \sum_{t=1}^{T} r_{3,t}^2 = O_p(1). \tag{B.16}$$

it will follow that

$$\begin{aligned}\frac{1}{T} \sum_{t=1}^{T} \Delta y_t \Delta y_{t-s} &=_{(1)} \frac{1}{T} \sum_{t=s+1}^{T} \Delta y_t \Delta y_{t-s} + o_p(1) \\ &=_{(2)} \frac{1}{T} \sum_{t=s+1}^{T} w_t w_{t-s} + o_p(1) \xrightarrow{p}_{(3)} \sigma^2 \sum_{n=0}^{\infty} \varphi_n \varphi_{n+s}.\end{aligned}$$

where $=_{(1)}$ holds by Lemma B.2, $=_{(2)}$ by (B.14)–(B.16) and the CS inequality, and $\xrightarrow{p}_{(3)}$ by (B.15).

It remains to prove (B.16). Since $\beta - 1 = O(T^{-1})$, there exists a $C < \infty$ such that

$$|r_{1,t}| \leq \frac{C}{T} \sum_{i=0}^{t-1} |\varphi_i||y_{t-1-i}| \leq \left(\sum_{i=0}^{\infty} |\varphi_i|\right) \frac{C}{\sqrt{T}} \max_{0 \leq t \leq T-1} \left|y_t / \sqrt{T}\right| = O_p(T^{-1/2})$$

uniformly in t by Theorem 3.2; the first part of (B.16) follows immediately. Next,

$$\sum_{t=1}^{T} r_{2,t}^2 = \sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\varphi_i y_{t-i}^-\right)^2 = \sum_{t=1}^{T}\sum_{i=0}^{t-1}\sum_{j=0}^{t-1}\varphi_i\varphi_j y_{t-i}^- y_{t-j}^-$$
$$= \sum_{i=0}^{T}\sum_{j=0}^{T}\varphi_i\varphi_j \sum_{t=\max\{i,j\}+1}^{T} y_{t-i}^- y_{t-j}^- = o_p(T)$$

by $\sum_{i=0}^{\infty}|\varphi_i| < \infty$, the result of part (ii), and the CS inequality. Finally, by Lemma B.3,

$$\sum_{t=1}^{T} r_{3,t}^2 = \sum_{t=1}^{T}\left(\sum_{i=1}^{k-1} d_{t-1,i}\Delta y_{1-i}\right)^2 \leq C^2\left(\sum_{i=1}^{k-1}|\Delta y_{1-i}|\right)^2 \sum_{t=1}^{T}\gamma_\phi^{2(t-1)} = O_p(1).$$

(iv). We first note that since $ab = \frac{1}{2}((a+b)^2 - a^2 - b^2)$,

$$\sum_{t=1}^{T} y_{t-1}\Delta y_t = \frac{1}{2}\sum_{t=1}^{T}(y_t^2 - y_{t-1}^2 - (\Delta y_t)^2) = \frac{1}{2}\left\{y_T^2 - y_0^2 - \sum_{t=1}^{T}(\Delta y_t)^2\right\}$$

which is $O_p(T)$ by Theorem 3.2 and part (iii) of this lemma. This gives the result when $s = 1$. To obtain the result for general $s \in \{0, \ldots, k-1\}$, we simply note that

$$\sum_{t=1}^{T} y_t\Delta y_t = \sum_{t=1}^{T}(\Delta y_t)^2 + \sum_{t=1}^{T} y_{t-1}\Delta y_t$$

and

$$\sum_{t=1}^{T} y_{t-s}\Delta y_t = \sum_{t=1}^{T}\left(y_t - \sum_{r=0}^{s-1}\Delta y_{t-r}\right)\Delta y_t = \sum_{t=1}^{T} y_t\Delta y_t - \sum_{r=0}^{s-1}\sum_{t=1}^{T}\Delta y_t\Delta y_{t-r}$$

when $s \geq 1$; the result then follows by a further appeal to part (iii) of this lemma. □

Proof of Theorem 3.4. Let $\boldsymbol{y}_t := (y_t, y_{t-1}, \ldots, y_{t-k+2})^\mathsf{T}$ and define

$$\mathcal{M}_T := \sum_{t=1}^{T}\begin{bmatrix} 1 & y_{t-1} & \Delta\boldsymbol{y}_{t-1}^\mathsf{T} \\ y_{t-1} & y_{t-1}^2 & y_{t-1}\Delta\boldsymbol{y}_{t-1}^\mathsf{T} \\ \Delta\boldsymbol{y}_{t-1} & y_{t-1}\Delta\boldsymbol{y}_{t-1} & \Delta\boldsymbol{y}_{t-1}\Delta\boldsymbol{y}_{t-1}^\mathsf{T}\end{bmatrix} \qquad m_T := \sum_{t=1}^{T}\begin{bmatrix}1 \\ y_{t-1} \\ \Delta\boldsymbol{y}_{t-1}\end{bmatrix}(u_t - y_t^-).$$

Then by rewriting (2.4) as

$$y_t = \alpha + \beta y_{t-1} + \boldsymbol{\phi}^\mathsf{T}\Delta\boldsymbol{y}_{t-1} + u_t - y_t^-,$$

the centred and rescaled OLS estimators $\hat{\mu}_T^\mathsf{T} := (\hat{\alpha}_T, \hat{\beta}_T, \hat{\boldsymbol{\phi}}_T^\mathsf{T})$ of $\mu^\mathsf{T} = (\alpha, \beta, \boldsymbol{\phi}^\mathsf{T})$ are equal to

$$\begin{bmatrix} T^{1/2}(\hat{\alpha}_T - \alpha) \\ T(\hat{\beta}_T - \beta) \\ \hat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}\end{bmatrix} = D_{2,T}(\hat{\mu}_T - \mu) = (D_{1,T}^{-1}\mathcal{M}_T D_{2,T}^{-1})^{-1} D_{1,T}^{-1} m_T \tag{B.17}$$

where $D_{1,T} := \mathrm{diag}\{T^{1/2}, T, I_{k-1}T\}$, $D_{2,T} := \mathrm{diag}\{T^{1/2}, T, I_{k-1}\}$,

$$D_{1,T}^{-1}\mathcal{M}_T D_{2,T}^{-1} = \begin{bmatrix} 1 & T^{-3/2}\sum_{t=1}^T y_{t-1} & T^{-1/2}\sum_{t=1}^T \Delta\boldsymbol{y}_{t-1}^{\mathsf{T}} \\ T^{-3/2}\sum_{t=1}^T y_{t-1} & T^{-2}\sum_{t=1}^T y_{t-1}^2 & T^{-1}\sum_{t=1}^T y_{t-1}\Delta\boldsymbol{y}_{t-1}^{\mathsf{T}} \\ T^{-3/2}\sum_{t=1}^T \Delta\boldsymbol{y}_{t-1} & T^{-2}\sum_{t=1}^T y_{t-1}\Delta\boldsymbol{y}_{t-1} & T^{-1}\sum_{t=1}^T \Delta\boldsymbol{y}_{t-1}\Delta\boldsymbol{y}_{t-1}^{\mathsf{T}} \end{bmatrix} \tag{B.18}$$

and

$$D_{1,T}^{-1} m_T = \begin{bmatrix} T^{-1/2}\sum_{t=1}^T (u_t - y_t^-) \\ T^{-1}\sum_{t=1}^T y_{t-1}(u_t - y_t^-) \\ T^{-1}\sum_{t=1}^T \Delta\boldsymbol{y}_{t-1}(u_t - y_t^-) \end{bmatrix}. \tag{B.19}$$

It remains to determine the weak limits of the elements of (B.18) and (B.19). We consider (B.18) first. By Theorem 3.2 and the CMT,

$$\mathcal{Y}_T := \begin{bmatrix} 1 & T^{-3/2}\sum_{t=1}^T y_{t-1} \\ T^{-3/2}\sum_{t=1}^T y_{t-1} & T^{-2}\sum_{t=1}^T y_{t-1}^2 \end{bmatrix} \xrightarrow{d} \begin{bmatrix} 1 & \int Y_{\theta_\phi}(r)\,\mathrm{d}r \\ \int Y_{\theta_\phi}(r)\,\mathrm{d}r & \int Y_{\theta_\phi}^2(r)\,\mathrm{d}r \end{bmatrix} =: \mathcal{Y}_{\theta_\phi}. \tag{B.20}$$

By Theorem 3.2 and Lemma B.4(iv),

$$\Xi_T := \begin{bmatrix} T^{-1/2}\sum_{t=1}^T \Delta\boldsymbol{y}_{t-1}^{\mathsf{T}} \\ T^{-1}\sum_{t=1}^T y_{t-1}\Delta\boldsymbol{y}_{t-1}^{\mathsf{T}} \end{bmatrix} = \begin{bmatrix} T^{-1/2}(\boldsymbol{y}_{T-1} - \boldsymbol{y}_{-1})^{\mathsf{T}} \\ T^{-1}\sum_{t=1}^T y_{t-1}\Delta\boldsymbol{y}_{t-1}^{\mathsf{T}} \end{bmatrix} = O_p(1). \tag{B.21}$$

By Lemma B.4(iii),

$$\Omega_T := T^{-1}\sum_{t=1}^T \Delta\boldsymbol{y}_{t-1}\Delta\boldsymbol{y}_{t-1}^{\mathsf{T}} \xrightarrow{p} \Omega \tag{B.22}$$

where $\Omega_{ij} := \sigma^2\sum_{n=0}^\infty \varphi_n\varphi_{n+|i-j|}$. It follows by the partitioned matrix inversion formula and the continuity of matrix inversion that

$$(D_{1,T}^{-1}\mathcal{M}_T D_{2,T}^{-1})^{-1} = \begin{bmatrix} \mathcal{Y}_T & \Xi_T \\ o_p(1) & \Omega_T \end{bmatrix}^{-1} = \begin{bmatrix} \mathcal{Y}_T^{-1} & -\mathcal{Y}_T^{-1}\Xi_T\Omega_T^{-1} \\ 0 & \Omega_T^{-1} \end{bmatrix} + o_p(1). \tag{B.23}$$

We turn next to (B.19). By Lemma B.4(i),

$$\mathcal{Z}_T^{(\alpha)} := \frac{1}{T^{1/2}}\sum_{t=1}^T (u_t - y_t^-) \to \phi(1)\left[Y_{\theta_\phi}(1) - c_\phi\int_0^1 Y_{\theta_\phi}(r)\,\mathrm{d}r - b_0\right] - a := \mathcal{Z}_{\theta_\phi}^{(\alpha)}. \tag{B.24}$$

Next, since only one of y_t and y_t^- can be nonzero, $y_{t-1}y_t^- = -\Delta y_t y_t^-$, and hence

$$\begin{aligned} \mathcal{Z}_T^{(\beta)} &:= \frac{1}{T}\sum_{t=1}^T y_{t-1}(u_t - y_t^-) = \frac{1}{T}\sum_{t=1}^T y_{t-1}u_t + \frac{1}{T}\sum_{t=1}^T \Delta y_t y_t^- \\ &=_{(1)} \frac{1}{T}\sum_{t=1}^T y_{t-1}u_t + o_p(1) \xrightarrow{d}_{(2)} \sigma\int_0^1 Y_{\theta_\phi}(r)\,\mathrm{d}W(r) := \mathcal{Z}_{\theta_\phi}^{(\beta)} \end{aligned} \tag{B.25}$$

where $\xrightarrow{d}_{(2)}$ holds by Theorem 3.2 and Liang *et al.* (2016, Theorem 2.1), and $=_{(1)}$ since

$$\left|\frac{1}{T}\sum_{t=1}^{T}\Delta y_t y_t^-\right|^2 \le \frac{1}{T}\sum_{t=1}^{T}(\Delta y_t)^2\frac{1}{T}\sum_{t=1}^{T}(y_t^-)^2 = o_p(1) \tag{B.26}$$

by the CS inequality and Lemma B.4(ii)–(iii). Finally, for $s \in \{1,\dots,k-1\}$,

$$\frac{1}{T}\sum_{t=1}^{T}\Delta y_{t-s}(u_t - y_t^-) = \frac{1}{T}\sum_{t=1}^{T}\Delta y_{t-s}u_t - \frac{1}{T}\sum_{t=1}^{T}\Delta y_{t-s}y_t^- = \frac{1}{T}\sum_{t=1}^{T}\Delta y_{t-s}u_t + o_p(1),$$

by the same argument as which yielded (B.26). Further, by Lemma B.2,

$$\mathbb{E}\left(\frac{1}{T}\sum_{t=1}^{T}\Delta y_{t-s}u_t\right)^2 = \frac{\sigma^2}{T^2}\sum_{t=1}^{T}\mathbb{E}(\Delta y_{t-s})^2 = O(T^{-1}).$$

Hence

$$\frac{1}{T}\sum_{t=1}^{T}\Delta y_{t-s}(u_t - y_t^-) = O_p(T^{-1/2}) + o_p(1) = o_p(1). \tag{B.27}$$

Letting $\mathcal{Z}_T \coloneqq (\mathcal{Z}_T^{(\alpha)}, \mathcal{Z}_T^{(\beta)})^\mathsf{T}$ and $\mathcal{Z}_{\theta_\phi} \coloneqq (\mathcal{Z}_{\theta_\phi}^{(\alpha)}, \mathcal{Z}_{\theta_\phi}^{(\beta)})^\mathsf{T}$, it follows from (B.19), (B.24), (B.25) and (B.27) that

$$D_{1,T}^{-1}m_T = \begin{bmatrix}\mathcal{Z}_T\\ o_p(1)\end{bmatrix} \xrightarrow{d} \begin{bmatrix}\mathcal{Z}_{\theta_\phi}\\ 0\end{bmatrix}.$$

Therefore, recalling (B.17), and using (B.20)–(B.23), we obtain

$$\begin{aligned}D_{2,T}(\hat{\mu}_T - \mu) &= \left(\begin{bmatrix}\mathcal{Y}_T^{-1} & -\mathcal{Y}_T^{-1}\Xi_T\Omega_T^{-1}\\ 0 & \Omega_T^{-1}\end{bmatrix} + o_p(1)\right)\begin{bmatrix}\mathcal{Z}_T\\ o_p(1)\end{bmatrix}\\ &= \begin{bmatrix}\mathcal{Y}_T^{-1}\mathcal{Z}_T\\ 0\end{bmatrix} + o_p(1) \xrightarrow{d} \begin{bmatrix}\mathcal{Y}_{\theta_\phi}^{-1}\mathcal{Z}_{\theta_\phi}\\ 0\end{bmatrix}.\end{aligned}$$

□

Proof of Corollary 3.2 ($k \ge 2$). We first show that $\hat{\sigma}_T^2 \xrightarrow{p} \sigma^2$. Adapting the argument from the $k = 1$ case, we have

$$\begin{aligned}\sum_{t=1}^{T}[\hat{u}_t^2 - (u_t - y_t^-)^2] &= \sum_{t=1}^{T}[(\alpha - \hat{\alpha}_T) + (\beta - \hat{\beta}_T)y_{t-1} + (\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}_T)^\mathsf{T}\Delta\boldsymbol{y}_{t-1}][\hat{u}_t + (u_t - y_t^-)]\\ &= (\alpha - \hat{\alpha}_T)\sum_{t=1}^{T}(u_t - y_t^-) + (\beta - \hat{\beta}_T)\sum_{t=1}^{T}y_{t-1}(u_t - y_t^-)\\ &\quad + (\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}_T)^\mathsf{T}\sum_{t=1}^{T}\Delta\boldsymbol{y}_{t-1}(u_t - y_t^-)\\ &= O_p(T^{-1/2})O_p(T^{1/2}) + O_P(T^{-1})O_p(T) + o_p(1)o_p(T) = o_p(T)\end{aligned}$$

where the the orders of the sums follow from Lemma B.4(i), (B.25) and (B.27), and the rates of convergence of the OLS estimators from Theorem 3.4. It follows that

$$\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2 = \frac{1}{T}\sum_{t=1}^{T}(u_t - y_t^-)^2 + o_p(1) = \frac{1}{T}\sum_{t=1}^{T}u_t^2 + o_p(1) \xrightarrow{p} \sigma^2$$

by Lemma B.4(ii), the LLN and the CS inequality.

To complete the proof, we note that since $\phi(1)^{-1}J_{\theta_\phi}(r) = Y_{\theta_\phi}(r)$,

$$\begin{aligned}\mathcal{Y}_{\theta_\phi} &= \begin{bmatrix} 1 & \int_0^1 Y_{\theta_\phi}(r)\,\mathrm{d}r \\ \int_0^1 Y_{\theta_\phi}(r)\,\mathrm{d}r & \int_0^1 Y_{\theta_\phi}^2(r)\,\mathrm{d}r \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & \phi(1)^{-1} \end{bmatrix} \begin{bmatrix} 1 & \int_0^1 J_{\theta_\phi}(r)\,\mathrm{d}r \\ \int_0^1 J_{\theta_\phi}(r)\,\mathrm{d}r & \int_0^1 J_{\theta_\phi}^2(r)\,\mathrm{d}r \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \phi(1)^{-1} \end{bmatrix}.\end{aligned}$$

It follows that the result of Theorem 3.4 can be rewritten as

$$\begin{bmatrix} T^{1/2}(\hat{\alpha}_T - \alpha) \\ T(\hat{\beta}_T - \beta) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \mathfrak{a}_{\theta_\phi} \\ \phi(1)\mathfrak{b}_{\theta_\phi} \end{bmatrix}.$$

By (B.20) and (B.23), the upper left 2×2 block of $D_{1,T}^{-1}\mathcal{M}_T D_{2,T}^{-1}$ converges to $\mathcal{Y}_{\theta_\phi}$. Thus, $D_{2,T}\mathcal{M}_T^{-1}D_{1,T}$ converges to $\mathcal{Y}_{\theta_\phi}^{-1}$, and so $T\mathcal{M}_T^{-1}(1,1) \xrightarrow{d} \mathcal{J}_{\theta_\phi}^{-1}(1,1)$ and $T^2\mathcal{M}_T^{-1}(2,2) \xrightarrow{d} \phi(1)^2\mathcal{J}_{\theta_\phi}^{-1}(2,2)$. Hence, the result follows by the CMT. □

References

Andrews, D. W. K. and Chen, H. Y. (1994). Approximately median-unbiased estimation of autoregressive models. *Journal of Business & Economic Statistics*, **12** (2), 187–204.

Aruoba, S. B., Mlikota, M., Schorfheide, F. and Villalvazo, S. (2021). SVARs with occasionally-binding constraints. *Journal of Econometrics*, in press, DOI: 10.1016/j.jeconom.2021.07.013.

Baillie, R. T. and Bollerslev, T. (1989). Common stochastic trends in a system of exchange rates. *Journal of Finance*, **44** (1), 167–181.

Brezigar-Masten, A., Masten, I. and Volk, M. (2021). Modeling credit risk with a Tobit model of days past due. *Journal of Banking & Finance*, **122**, 105984.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: theory and methods*. Springer.

Bykhovskaya, A. (2021). Time series approach to the evolution of networks: prediction and estimation. *Journal of Business & Economic Statistics*, in press, DOI: 10.1080/07350015.2021.2006669.

— and Phillips, P. C. B. (2018). Boundary limit theory for functional local to unity regression. *Journal of Time Series Analysis*, **39** (4), 523–562.

Cavaliere, G. (2004). The asymptotic distribution of the Dickey–Fuller statistic under nonnegativity constraint. *Econometric Theory*, **20** (4), 808–810.

— (2005). Limited time series with a unit root. *Econometric Theory*, **21** (5), 907–945.

— and Xu, F. (2014). Testing for unit roots in bounded time series. *Journal of Econometrics*, **178**, 259–272.

Chan, K.-S. (ed.) (2009). *Exploration of a Nonlinear World: an appreciation of Howell Tong's contributions to statistics*. World Scientific.

Chan, N. H. and Wei, C. Z. (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics*, pp. 1050–1063.

de Jong, R. and Herrera, A. M. (2011). Dynamic censored regression and the Open Market Desk reaction function. *Journal of Business & Economic Statistics*, **29** (2), 228–237.

Demiralp, S. and Jordà, O. (2002). The announcement effect: evidence from Open Market Desk data. *FRBNY Economic Policy Review*, **8**, 29–48.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, **74** (366), 427–431.

Dong, D., Schmit, T. M. and Kaiser, H. (2012). Modelling household purchasing behaviour to analyse beneficial marketing strategies. *Applied Economics*, **44** (6), 717–725.

Duffy, J. A., Mavroeidis, S. and Wycherley, S. (2022). Stationarity, unit roots and cointegration in an SVAR with an occasionally binding constraint, Working paper, University of Oxford.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: nonparametric and parametric methods*. Springer.

Gao, J. (2007). *Nonlinear Time Series: semiparametric and nonparametric methods*. Chapman and Hall/CRC.

—, Tjøstheim, D. and Yin, J. (2013). Estimation in threshold autoregressive models with a stationary and a unit root regime. *Journal of Econometrics*, **172** (1), 1–13.

Giraitis, L. and Phillips, P. C. B. (2006). Uniform limit theory for stationary autoregression. *Journal of Time Series Analysis*, **27** (1), 51–60.

Hahn, J. and Kuersteiner, G. (2010). Stationarity and mixing properties of the dynamic Tobit model. *Economics Letters*, **107** (2), 105–111.

Hansen, B. E. (1999). The grid bootstrap and the autoregressive model. *Review of Economics and Statistics*, **81** (4), 594–607.

Hertrich, M. (2022). Foreign exchange interventions under a minimum exchange rate regime and the Swiss franc. *Review of International Economics*, **30** (2), 450–489.

Hong, S. H. and Phillips, P. C. B. (2010). Testing linearity in cointegrating relations with an application to purchasing power parity. *Journal of Business & Economic Statistics*, **28** (1), 96–114.

Jordan, T. J. (2016). The euro and Swiss monetary policy. Speech at the Europa Forum, Lucerne, 2 May.

Jungers, R. (2009). *The Joint Spectral Radius: theory and applications*. Springer.

Karatzas, I. and Shreve, S. (2012). *Brownian Motion and Stochastic Calculus*. Springer Science & Business Media.

Liang, H., Phillips, P. C. B., Wang, H. and Wang, Q. (2016). Weak convergence to stochastic integrals for econometric applications. *Econometric Theory*, **32** (6), 1349–1375.

Liebscher, E. (2005). Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes. *Journal of Time Series Analysis*, **26** (5), 669–689.

Liu, L., Moon, H. R. and Schorfheide, F. (2019). Forecasting with a panel tobit model. NBER working paper 28571.

Liu, W., Ling, S. and Shao, Q.-M. (2011). On non-stationary threshold autoregressive models. *Bernoulli*, **17** (3), 969–986.

Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. USA: C.U.P.

Mavroeidis, S. (2021). Identification at the zero lower bound. *Econometrica*, **69** (6), 2855–2885.

Michel, J. and de Jong, R. (2018). Mixing properties of the dynamic tobit model with mixing errors. *Economics Letters*, **162**, 112–115.

Mikusheva, A. (2007). Uniform inference in autoregressive models. *Econometrica*, **75** (5), 1411–1452.

— (2012). One-dimensional inference in autoregressive models with the potential presence of a unit root. *Econometrica*, **80** (1), 173–212.

Moran, P. A. P. (1953). The statistical analysis of the Canadian lynx cycle, I: structure and prediction. *Australian Journal of Zoology*, **1**, 163–173.

Phillips, P. C. B. (1987a). Time series regression with a unit root. *Econometrica*, pp. 277–301.

— (1987b). Towards a unified asymptotic theory for autoregression. *Biometrika*, **74** (3), 535–547.

— and Solo, V. (1992). Asymptotics for linear processes. *Annals of Statistics*, **20** (2), 971–1001.

Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion*. Springer.

Rossi, B. (2013). Exchange rate predictability. *Journal of Economic Literature*, **51** (4), 1063–1119.

Saikkonen, P. (2008). Stability of regime switching error correction models under linear cointegration. *Econometric Theory*, **24** (1), 294–318.

Sarno, L. and Valente, G. (2006). Deviations from purchasing power parity under different exchange rate regimes: Do they revert and, if so, how? *Journal of Banking & Finance*, **30** (11), 3147–3169.

Terasvirta, T., Tjøstheim, D. and Granger, C. W. J. (2010). *Modelling Nonlinear Economic Time Series*. O.U.P.

Tobin, J. (1958). Estimation of relationship for limited dependent variables. *Econometrica*, **26** (1), 24–36.

von Schweinitz, G., Tonzer, L. and Buchholz, M. (2021). Monetary policy through exchange rate pegs: the removal of the Swiss franc–euro floor and stock price reactions. *International Review of Finance*, **21** (4), 1382–1406.

Wei, S. X. (1999). A bayesian approach to dynamic Tobit models. *Econometric Reviews*, **18** (4), 417–439.

(Anna Bykhovskaya) Duke University
Email address: anna.bykhovskaya@duke.edu

(James A. Duffy) University of Oxford
Email address: james.duffy@economics.ox.ac.uk