Nonstationary Censored Regression

Ling Hu* Robert de Jong†

March 30, 2006

Abstract

This paper considers the censored regression model under the assumption that the regressors are integrated. We show that Maximum Likelihood estimation is superconsistent and asymptotically mixed normal, implying that standard inference techniques remain valid, and that in general least squares estimation based on the positive observations only is superconsistent, but not mixed normal. An exception to this is the case of a single integrated regressor; in that case, least squares on the positive observations will be asymptotically mixed normal and asymptotically equivalent to the Tobit Maximum Likelihood estimator. We also derive a test for the null of Tobit cointegration, and apply this test to price floors of agricultural commodities.

1 Introduction

There has been a substantial amount of recent econometric work on the estimation of nonlinear models with integrated regressors and on nonlinear functions of integrated regressors. Park and Phillips (1999) derived the asymptotic theory for averages of nonlinear functions of integrated processes. Based on this work, Park and Phillips (2000) analyzed the binary choice model in the presence of integrated regressors, and Park and Phillips (2001) analyzed nonlinear least squares with integrated regressors. This paper analyzes the Tobit model when regressors are integrated.

The likelihood function of the Tobit model is nonlinear, and results from the literature of estimation in the presence of integrated processes are not able to deal with the Tobit model directly. Also, trying to follow the line of proof of Park and Phillips (2000) and Park and Phillips (2001) for dealing with the Tobit model is not feasible. Because an orthogonality condition in the quadratic variation process fails to hold (see Park and Phillips (2000, p. 1272, Equation (29)), we were unable to use an analogue of the approach of Park and Phillips (2000) for deriving the asymptotics for the nonstationary Tobit model, and instead

*Department of Economics, 429 Arps Hall, Ohio State University, 1945 N. High Street, Columbus, OH 43210, email linghu@econ.ohio-state.edu.

†Department of Economics, 403 Arps Hall, Ohio State University, 1945 N. High Street, Columbus, OH 43210, email dejong@econ.ohio-state.edu.

we have used a different analytical approach that completely avoids the use of stochastic calculus. For the derivation of our mixed normality result, we apply a central limit theorem type result from Hall and Heyde (1980). While this approach works well in our setting, it does force us to make a stronger exogeneity assumption. Also, an important tool used is an approach suggested by Pötscher (2004) to calculate bounds for the expectation of averages of functions of integrated processes.

We will show in this paper that Maximum Likelihood (ML) estimation of the Tobit model in the presence of integrated regressors is superconsistent and asymptotically mixed normal. This implies validity of standard inference procedures, such as t statistics and F tests. Therefore, the asymptotic behavior of the Tobit ML estimator is in this respect reminiscent of that of the least squares estimator in a cointegrating regression under exogeneity and uncorrelatedness assumptions. In addition, it will be shown that the Ordinary Least Squares (OLS) estimator of the nonstationary Tobit model using only the positive observations is also superconsistent, but not mixed normal in general, which implies that standard inference techniques based on OLS are not correct in general, even under assumptions of exogeneity and uncorrelated errors. As in the stationary case, OLS on all observations remains inconsistent.

An interesting result that does not appears to have parallels elsewhere is that if OLS is applied to the positive observations in the case of a one-dimensional regressor, we do obtain mixed normality. Therefore, OLS on the positive observations yields incorrect asymptotic inference, except for the case when the regressor is one-dimensional. This opens up the possibility to derive a residual-based test for the null of Tobit cointegration when there is a single regressor. We do so by amending the cointegration test of Shin (1994) with a Tobit adjustment, and we tabulate the critical values of the resulting test statistic. We apply the test in price dynamics of dairy products where there are binding price floors.

The first full treatment of the asymptotic theory for the Tobit model in the case of i.i.d. data appears to be Amemiya (1973). Using standard minimization theory for dependent variables, these results can be extended to cover the case of data that have fading memory properties.

The literature on estimation in the presence of nonlinear functions of integrated processes is rapidly growing. In our context, it may be useful to note that Park and Phillips (2000) and Park and Phillips (2001) make assumptions of exogeneity and uncorrelatedness, which leads to mixed normality and therefore the validity of standard inference. We will follow this type of approach in this paper. Recently, Park and Shintani (2005), Chang and Park (2004), de Jong (2002) and de Jong, Wang and Bae (2005) have considered dynamic OLS and fully modified type approaches to deal with possible issues of correlation and endogeneity, but we will not pursue this in this paper, because the structure of the Tobit model does not seem to allow for such extensions without additional analytical complications and because the present analysis is already far from straightforward.

Section 2 of this paper will consider Maximum Likelihood estimation of the Tobit model. Section 3 considers OLS estimation of the Tobit model. Section 4 contains our simulations, and an empirical application is given in Section 5. Section 6 concludes. Appendix A gives some useful lemmas and the proofs of the main results are given in Appendix B.

2 The model and ML estimation

2.1 The model

We consider the regression model given by

$$y_t^* = x_t'\beta_0 + \varepsilon_t, \quad \text{for} \quad t = 1, \ldots, T \tag{1}$$

and

$$y_t = \begin{cases} y_t^* & \text{when} \quad y_t^* > 0, \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where ε_t is assumed to be i.i.d. $N(0, \sigma_0^2)$ conditional on $x_1, \ldots, x_T$. We will consider Maximum Likelihood (ML) estimation of the parameter $\theta_0 = (\beta_0', \sigma_0^2)'$ under the assumption of error normality.

As in Park and Phillips (2000), we make the following assumption on the data generating process of x_t. Below, we define $|X| = (\text{tr}(X'X))^{1/2}$.

Assumption 1 *Let $x_t = x_{t-1} + v_t$ with $x_0 = 0$ and where*

$$v_t = \Pi(L)e_t = \sum_{i=0}^{\infty} \Pi_i e_{t-i}, \tag{3}$$

$\Pi(1)$ *nonsingular and* $|\Pi_i| \leq Ci^{-2-\eta}$ *for some* $C > 0$ *and* $\eta > 0$. *The innovations* e_t *are iid with mean zero and* $E|e_t|^r < \infty$ *for some* $r \geq 4$, *have a distribution that is absolutely continuous with respect to Lebesgue measure, and have characteristic function* $\varphi(t)$ *which satisfies* $\lim_{|t|\to\infty} |t|^\kappa \varphi(t) = 0$ *for some* $\kappa > 1$.

Assumption 2 ε_t *is i.i.d. and* $\varepsilon_t | x_1, \ldots, x_T$ *is distributed* $N(0, \sigma_0^2)$.

Our assumption on the parameter space is the following.

Assumption 3 $\Theta = B \times \Sigma$ *is compact,* $\beta_0 \in \text{int}(B)$ *and* $\sigma_0^2 \in \text{int}(\Sigma)$, *and* $\sigma_{min} = \inf \Sigma > 0$.

Assumption 1 implies that x_t is a multivariate I(1) process without cointegrating relationships that satisfies $T^{-1/2}x_{[rT]} \Rightarrow V(r)$, where $V(r)$ is a multivariate rescaled Brownian motion process. Here and elsewhere in this paper, "$\Rightarrow$" denotes weak convergence. We strengthened Park and Phillips' (2000) moment condition on e_t to finiteness of a fourth moment. In addition, we strengthened Park and Phillips' (2000) assumption on the characteristic function. As our assumption stands here, the existence of a uniformly density function $f_t(\cdot)$ of $t^{-1/2}x_t$ is implied for $t \geq 1$, because $\kappa > 1$ by assumption. This assumption can be relaxed to $\kappa > 0$, at the expense of additional complications; Pötscher (2004) documents that this latter assumption implies the existence of a uniformly bounded density from some value of t onwards, and his treatment of such a situation should also apply to our setting.

As in Park and Phillips (2000), we rotate the regressor space, and assume that $\beta_0 \neq 0$ and use an orthogonal matrix $A = (a_1, A_2)$ with $a_1 = \beta_0/(\beta_0'\beta_0)^{1/2}$. Let $(\alpha^1, \alpha^{2'})' = \alpha = A'\beta$ and define $\alpha_0 = (\alpha_0^1, \alpha_0^{2'})' = A'\beta_0$. Then we can write (1) as

$$
\begin{aligned}
y_t^* &= x_t'\beta_0 + \varepsilon_t \\
&= x_t'AA'\beta_0 + \varepsilon_t \\
&= (A'x_t)'A'\beta_0 + \varepsilon_t \\
&= x_{1t}\alpha_0^1 + x_{2t}'\alpha_0^2 + \varepsilon_t, \qquad (4)
\end{aligned}
$$

where

$$A'x_t = (x_{1t}, x_{2t}')', \quad x_{1t} = a_1'x_t \quad \text{and} \quad x_{2t} = A_2'x_t, \qquad (5)$$

$$\alpha_0^1 = a_1'\beta_0 = (\beta_0'\beta_0)^{1/2} \quad \text{and} \quad \alpha_0^2 = A_2'\beta_0 = 0. \qquad (6)$$

Accordingly, we now define

$$V_1 = a_1'V \quad \text{and} \quad V_2 = A_2'V, \qquad (7)$$

which are Brownian motions of dimensions 1 and $(m-1)$, respectively.

2.2 Preliminary results regarding ML estimation

The likelihood, the scores and the Hessians are given in Amemiya (1973), and can be written as

$$
\begin{aligned}
\log L_T(\theta) = &\sum_{t=1}^{T} \log(1 - \Phi(x_t'\beta/\sigma))\mathbf{1}(x_t'\beta_0 + \varepsilon_t \leq 0) \\
&-(1/2)(\log(2\pi\sigma^2)) \sum_{t=1}^{T} \mathbf{1}(x_t'\beta_0 + \varepsilon_t > 0) \\
&-(1/2)\sigma^{-2} \sum_{t=1}^{T} (y_t - x_t'\beta)^2 \mathbf{1}(x_t'\beta_0 + \varepsilon_t > 0), \qquad (8)
\end{aligned}
$$

and $S_T(\theta) = (S_\beta(\theta)', S_{\sigma^2}(\theta))'$ where

$$
\begin{aligned}
S_\beta(\theta) = (\partial \log L_T(\theta)/\partial\beta) = &-\sigma^{-1} \sum_{t=1}^{T} \phi(x_t'\beta/\sigma)(1 - \Phi(x_t'\beta/\sigma))^{-1} x_t \mathbf{1}(x_t'\beta_0 + \varepsilon_t \leq 0) \\
&+\sigma^{-2} \sum_{t=1}^{T} (y_t - x_t'\beta) x_t \mathbf{1}(x_t'\beta_0 + \varepsilon_t > 0), \qquad (9)
\end{aligned}
$$

$$S_{\sigma^2}(\theta) = (\partial \log L_T(\theta)/\partial\sigma^2)$$

$$= (1/2)\sigma^{-3}\sum_{t=1}^{T} x_t'\beta\phi(x_t'\beta/\sigma)(1-\Phi(x_t'\beta/\sigma))^{-1}\mathbf{1}(x_t'\beta_0+\varepsilon_t \le 0)$$

$$-(1/2)\sigma^{-2}\sum_{t=1}^{T}\mathbf{1}(x_t'\beta_0+\varepsilon_t > 0) + (1/2)\sigma^{-4}\sum_{t=1}^{T}(y_t - x_t'\beta)^2\mathbf{1}(x_t'\beta_0+\varepsilon_t > 0), \quad (10)$$

and

$$H_T(\theta) = \begin{pmatrix} H_{\beta\beta'}(\theta) & H_{\sigma^2\beta'}(\theta) \\ H_{\sigma^2\beta'}(\theta)' & H_{\sigma^2\sigma^2}(\theta) \end{pmatrix} \quad (11)$$

where

$$H_{\beta\beta'}(\theta) = (\partial^2 \log L_T(\theta)/\partial\beta\partial\beta') = -\sigma^{-2}\sum_{t=1}^{T} g_1(x_t'\beta/\sigma)\mathbf{1}(x_t'\beta_0+\varepsilon_t \le 0)x_t x_t'$$

$$-\sigma^{-2}\sum_{t=1}^{T} x_t x_t'\mathbf{1}(x_t'\beta_0+\varepsilon_t > 0), \quad (12)$$

$$H_{\sigma^2\beta'}(\theta) = (\partial^2 \log L_T(\theta)/\partial\sigma\partial\beta') = -\sigma^{-3}\sum_{t=1}^{T} g_2(x_t'\beta/\sigma)x_t\mathbf{1}(x_t'\beta_0+\varepsilon_t \le 0)$$

$$-\sigma^{-4}\sum_{t=1}^{T}(y_t - x_t'\beta)x_t\mathbf{1}(x_t'\beta_0+\varepsilon_t > 0), \quad (13)$$

and

$$H_{\sigma^2\sigma^2}(\theta) = (\partial^2 \log L_T(\theta)/\partial(\sigma^2)^2)$$

$$= \sigma^{-4}\sum_{t=1}^{T} g_3(x_t'\beta/\sigma)\mathbf{1}(x_t'\beta_0+\varepsilon_t \le 0) + (1/2)\sigma^{-4}\sum_{t=1}^{T}\mathbf{1}(x_t'\beta_0+\varepsilon_t > 0)$$

$$-\sigma^{-6}\sum_{t=1}^{T}(y_t - x_t'\beta)^2\mathbf{1}(x_t'\beta_0+\varepsilon_t > 0), \quad (14)$$

where

$$g_1(y) = \phi(y)(1-\Phi(y))^{-2}[\phi(y) - (1-\Phi(y))y], \quad (15)$$

$$g_2(y) = (1/2)\phi(y)(1-\Phi(y))^{-2}[(1-\Phi(y))y^2 - (1-\Phi(y)) - y\phi(y)], \quad (16)$$

and

$$g_3(y) = (1/4)\phi(y)(1-\Phi(y))^{-2}[(1-\Phi(y))y^3 - 3(1-\Phi(y))y - y^2\phi(y)]. \quad (17)$$

The following result will help us simplifying the analysis of the Hessian matrix.

Lemma 1 *Let Assumptions 1 and 2 hold. Let* $g(\cdot)$ *be a function such that*

$$\int_{-\infty}^{\infty} |g(y)|(1-\Phi(y))dy < \infty. \tag{18}$$

Let x_{2t}^{κ} *denote the* κ*-times tensor product of* x_{2t} *with itself. Then*

$$T^{-1-\kappa/2}\left|\sum_{t=1}^{T} g(x_{1t})\mathbf{1}(x_t'\beta_0+\varepsilon_t \le 0)x_{2t}^{\kappa}\right| \xrightarrow{p} 0. \tag{19}$$

With some computations, we can verify that the nonlinear functions in all three Hessians (i.e. $g_1(\cdot)$, $g_2(\cdot)$ and $g_3(\cdot)$) satisfy the property of $g(\cdot)$ in Lemma 1. Therefore, the integrability condition of Lemma 1 holds, and we can show the asymptotic negligibility of the corresponding terms in the Hessian, and the limit result for the Hessian then easily follows.

Lemma 2 below provides limit results for sample moments and covariance functions and assists in analyzing the asymptotic behavior of the score function and Hessian.

Lemma 2 *Let Assumption 1 hold. Then as* $T \to \infty$*,*

(a) $T^{-1}\sum_{t=1}^{T}\Phi(x_{1t}) \xrightarrow{d} \int_0^1 \mathbf{1}(V_1(r)>0)dr$,

(b) $T^{-3/2}\sum_{t=1}^{T}\Phi(x_{1t})x_{1t} \xrightarrow{d} \int_0^1 \mathbf{1}(V_1(r)>0)V_1(r)dr$,

(c) $T^{-2}\sum_{t=1}^{T}\Phi(x_{1t})x_{1t}^2 \xrightarrow{d} \int_0^1 \mathbf{1}(V_1(r)>0)V_1(r)^2dr$,

(d) $T^{-3/2}\sum_{t=1}^{T}\Phi(x_{1t})x_{2t} \xrightarrow{d} \int_0^1 \mathbf{1}(V_1(r)>0)V_2(r)V_2(r)'dr$,

(e) $T^{-2}\sum_{t=1}^{T}\Phi(x_{1t})x_{2t}x_{2t}' \xrightarrow{d} \int_0^1 \mathbf{1}(V_1(r)>0)V_2(r)V_2(r)'dr$,

and

(f) $T^{-2}\sum_{t=1}^{T}\Phi(x_{1t})x_{1t}x_{2t} \xrightarrow{d} \int_0^1 \mathbf{1}(V_1(r)>0)V_1(r)V_2(r)dr$.

2.3 Main Results

Let $\hat\theta_T = (\hat\beta_T', \hat\sigma_T^2)'$ be the ML estimator of $\theta_0 = (\beta_0', \sigma_0^2)'$. As usual in ML limit theory, the asymptotic distribution of $\hat\theta_T$ will be obtained from the expansion

$$0 = S_T(\hat\theta_T) = S_T(\theta_0) + H_T(\tilde\theta_T)(\hat\theta_T - \theta_0), \tag{20}$$

or in partitioned form,

$$0 = \begin{pmatrix} S_\beta(\hat\theta_T) \\ S_{\sigma^2}(\hat\theta_T) \end{pmatrix} = \begin{pmatrix} S_\beta(\theta_0) \\ S_{\sigma^2}(\theta_0) \end{pmatrix} + \begin{pmatrix} H_{\beta\beta}(\tilde\theta_T) & H_{\sigma^2\beta'}(\tilde\theta_T) \\ H_{\sigma^2\beta'}(\tilde\theta_T)' & H_{\sigma^2\sigma^2}(\tilde\theta_T) \end{pmatrix} \begin{pmatrix} \hat\beta_T - \beta_0 \\ \hat\sigma_T^2 - \sigma_0^2 \end{pmatrix}, \tag{21}$$

where $\tilde{\theta}$ is a mean value on the line segment between $\hat{\theta}_T$ and θ_0. Corresponding to the rotation in the regressors and parameters, define

$$G = \begin{pmatrix} A & 0 \\ 0 & 1 \end{pmatrix}. \tag{22}$$

Pre-multiplying (20) by G' and remembering that $G'G = I$ gives

$$0 = G'S_T(\hat{\theta}_T) = G'S_T(\theta_0) + G'H_T(\tilde{\theta}_T)GG'(\hat{\theta}_T - \theta_0), \tag{23}$$

and the asymptotic distribution for $\hat{\theta}_T - \theta_0$ will be derived from the above equation.

The limiting distribution of the Hessian and the score functions are given below by Lemma 3 and Lemma 4 respectively.

Lemma 3 *Let Assumptions 1 and 2 hold. Then*

$$\begin{pmatrix} T^{-2}A'H_{\beta\beta'}(\theta_0)A & T^{-3/2}A'H_{\sigma^2\beta'}(\theta_0) \\ T^{-3/2}H_{\sigma^2\beta'}(\theta_0)'A & T^{-1}H_{\sigma^2\sigma^2}(\theta_0) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} -Q & 0 \\ 0 & -q \end{pmatrix}, \tag{24}$$

where

$$Q = \sigma_0^{-2} \begin{bmatrix} (\alpha_0^1)^2 \int_0^1 V_1(r)^2 \mathbf{1}(V_1(r) > 0)dr & \alpha_0^1 \int_0^1 \mathbf{1}(V_1(r) > 0)V_1(r)V_2(r)'dr \\ \alpha_0^1 \int_0^1 \mathbf{1}(V_1(r) > 0)V_1(r)V_2(r)dr & \int_0^1 \mathbf{1}(V_1(r) > 0)V_2(r)V_2(r)'dr \end{bmatrix} \tag{25}$$

and

$$q = (1/2)\sigma_0^{-4} \int_0^1 \mathbf{1}(V_1(r) > 0)dr. \tag{26}$$

Lemma 4 *Under Assumptions 1 and 2,*

$$(T^{-1}S_\beta(\theta_0)'A, T^{-1/2}S_{\sigma^2}(\theta_0))' \xrightarrow{d} MN\left(0, \begin{pmatrix} -Q & 0 \\ 0 & -q \end{pmatrix}\right), \tag{27}$$

jointly with the convergence in Lemma 3.

Notice that the linear and nonlinear components are of the same order in the score functions, but that they are of different orders in the Hessian. Lemma 3 establishes that the nonlinear components are of smaller order, and are dominated in the Hessian by the linear components. The asymptotics of the linear components are then derived using Lemma 2.

We can now establish an order result for our ML estimator:

Lemma 5 *Under Assumptions 1, 2 and 3,* $A'T(\hat{\beta}_T - \beta_0) = O_p(1)$ *and* $T^{1/2}(\hat{\sigma}_T^2 - \sigma_0^2) = O_p(1)$.

Hence the ML estimator of β is superconsistent, which is similar to the superconsistency of the least squares estimator in a linear cointegration regression. With the above lemmas we can further show asymptotic normality of the ML estimators, which is our main result:

Theorem 1 *Under Assumptions 1, 2 and 3, there exists a sequence of ML estimators for which* $\hat{\alpha}_T \xrightarrow{p} \alpha_0$ *and* $\hat{\sigma}_T^2 \xrightarrow{p} \sigma_0^2$ *and*

$$TA'(\hat{\beta}_T - \beta_0) = T(\hat{\alpha}_T - \alpha_0) \xrightarrow{d} MN(0, Q^{-1}), \tag{28}$$

and

$$T^{1/2}(\hat{\sigma}_T^2 - \sigma_0^2) \xrightarrow{d} MN(0, 1/q). \tag{29}$$

Since the estimator is asymptotically mixed normal, the above result implies that standard t-tests and F-tests are asymptotically valid. This is the same conclusion as reached in Park and Phillips (2000).

3 The OLS estimator

It is well known that the OLS estimation of a stationary Tobit model is inconsistent. When the regressors are unit root nonstationary, we find that the OLS estimator is superconsistent.

Let $\hat{\beta}_T^{OLS}$ denote the OLS estimator of β that uses the positive observations only,

$$\hat{\beta}_T^{OLS} = (\sum_{t=1}^{T} x_t x_t' \mathbf{1}(x_t'\beta_0 + \varepsilon_t > 0))^{-1} \sum_{t=1}^{T} x_t y_t \mathbf{1}(x_t'\beta_0 + \varepsilon_t > 0). \tag{30}$$

$\hat{\beta}_T^{OLS}$ can be rewritten as

$$TA'(\hat{\beta}_T^{OLS} - \beta_0) = (T^{-2} \sum_{t=1}^{T} A'x_t(A'x_t)' \mathbf{1}(x_t'\beta_0 + \varepsilon_t > 0))^{-1} T^{-1} \sum_{t=1}^{T} A'x_t \varepsilon_t \mathbf{1}(x_t'\beta_0 + \varepsilon_t > 0). \tag{31}$$

Define

$$B_T = (T^{-2} \sum_{t=1}^{T} A'x_t(A'x_t)' \mathbf{1}(x_t'\beta_0 + \varepsilon_t > 0))^{-1} T^{-1} \sum_{t=1}^{T} A'x_t \sigma_0 \phi(\sigma_0^{-1}\beta_0' x_t). \tag{32}$$

Theorem 2 *Under Assumption 1, 2 and 3,*

$$TA'(\hat{\beta}_T^{OLS} - \beta_0) - B_T \xrightarrow{d} MN(0, Q^{-1}). \tag{33}$$

The proof is straightforward (hence omitted) since the numerator in B_T is asymptotically equivalent to the nonlinear term in $S_\beta(\theta)$, which is the difference between the OLS and ML estimators. Notice that the numerator of B_T can be written as

$$T^{-1} \sum_{t=1}^{T} A'x_t \sigma_0 \phi(\sigma_0^{-1}\beta_0' x_t) = \begin{bmatrix} T^{-1} \sum_{t=1}^{T} x_{1t} \sigma_0 \phi(\sigma_0^{-1}\alpha_0^1 x_{1t}) \\ T^{-1} \sum_{t=1}^{T} x_{2t} \sigma_0 \phi(\sigma_0^{-1}\alpha_0^1 x_{1t}) \end{bmatrix} = \begin{bmatrix} O_p(T^{-1/2}) \\ O_p(1) \end{bmatrix}, \tag{34}$$

by Lemma 2 in Park and Phillips (2000). Therefore, asymptotically the bias of the OLS estimator comes from the direction that is orthogonal to the true parameters.

The above result implies that standard inference based on the OLS estimator is not valid in general. However, one special case that is of interest is when x_t is one-dimensional. For one-dimensional x_t, OLS estimation is asymptotically mixed normal, implying that standard t- and F-test procedures are asymptotically valid. This is because in the case of one-dimensional x_t the second vector element on the right-hand of Equation (34) disappears and B_T becomes asymptotically negligible. Therefore in this case, the OLS estimator is in fact asymptotically equivalent to the ML estimator and the difference is of order $O_p(T^{-1/2})$.

4 Simulation results

In this section, we report some simulation results of the finite sample performance of the ML estimator and the OLS estimator based on positive observations. The exogenous covariates are generated by a bivariate vector autoregression of the form

$$x_{it} = x_{i,t-1} + v_{it}, \quad i = 1, 2 \tag{35}$$

with $v_t = (v_{1t}, v_{2t})'$ i.i.d. and $N(0, I_2)$. The two regression coefficients are set as $\beta_0^1 = 1$ and $\beta_0^2 = 0$. The regression error is generated from a standard normal distribution. We study both the ML estimator and the truncated OLS estimator. For the truncated OLS estimator, we estimate the parameters only when there are at least 5 positive observations of the dependent variables. The number of replications is 5000.

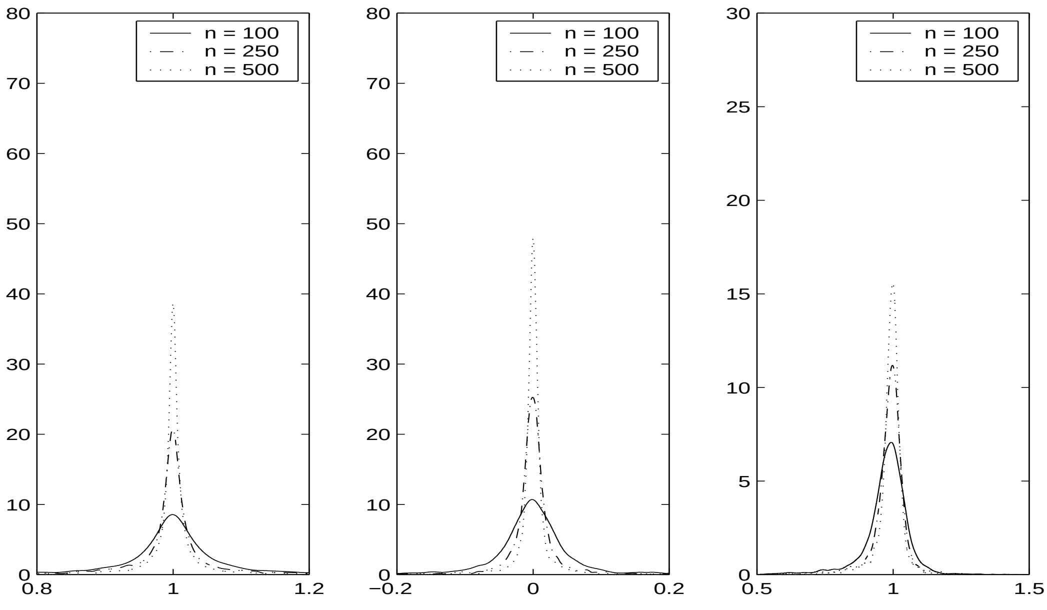

Figure 1: Densities of ML estimators of $\beta_0^1 = 1$, $\beta_0^2 = 0$ and $\sigma_0^2 = 1$

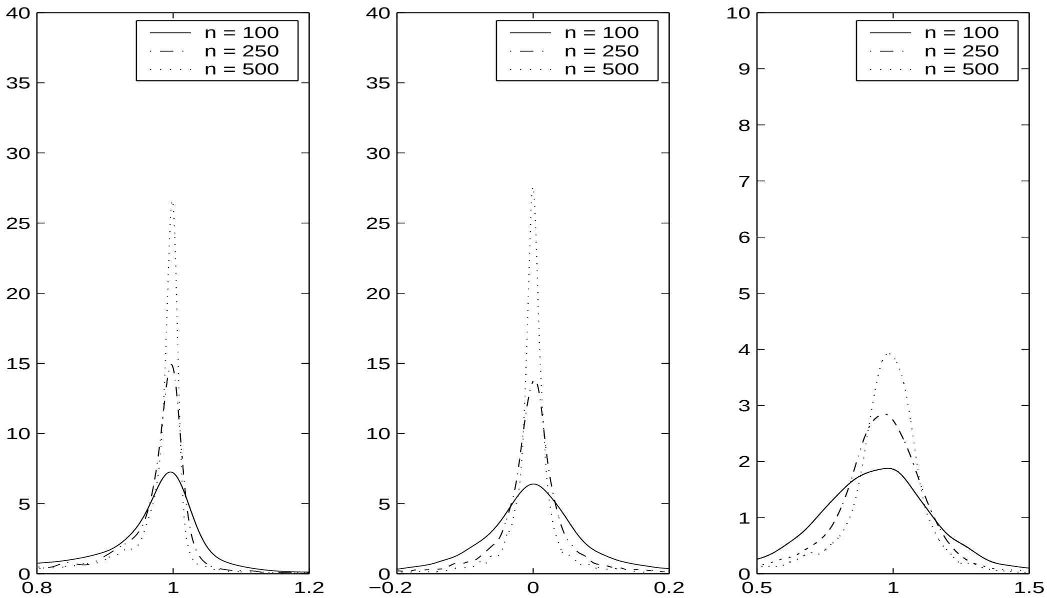

Figure 2: Densities of truncated OLS estimators of $\beta_0^1 = 1$, $\beta_0^2 = 0$ and $\sigma_0^2 = 1$

Figure 1 and Figure 2 plot the densities of the ML and truncated OLS estimators respectively for various sample sizes. It appears that the ML estimator is more efficient than the truncated OLS estimator when there are more than one regressors. When there is one regressor, these two estimators are asymptotically equivalent. Figure 3 plots the densities of these two estimators in univariate Tobit models and they are very similar.

5 A test for Tobit cointegration with an empirical application

When the nonstationary Tobit model is estimated in empirical applications, it is desirable to have a valid test statistic of the null hypothesis that the error term is indeed stationary. If the error term is nonstationary, the regression will be spurious and the inference results derived in this paper will be invalid. In this section, we derive a consistent test for the null of Tobit cointegration and we apply the test to an empirical study of the dairy market.

In linear regressions a widely used cointegration test is Shin's test; see Shin (1994). The null hypothesis of the test is cointegration, and the alternative is a spurious regression. The Shin test procedure consists of computing the KPSS statistic on the regression residuals, and using the appropriate critical values, which are different from those used for the KPSS test applied to a regular series. In this paper, we propose a similar test in univariate censored regressions, and we employ a Tobit adjustment. Recall that when x_t is a scalar, the OLS estimator based on positive observations is asymptotically equivalent to the ML estimator. Therefore the test can be conducted by running an OLS regression on the positive

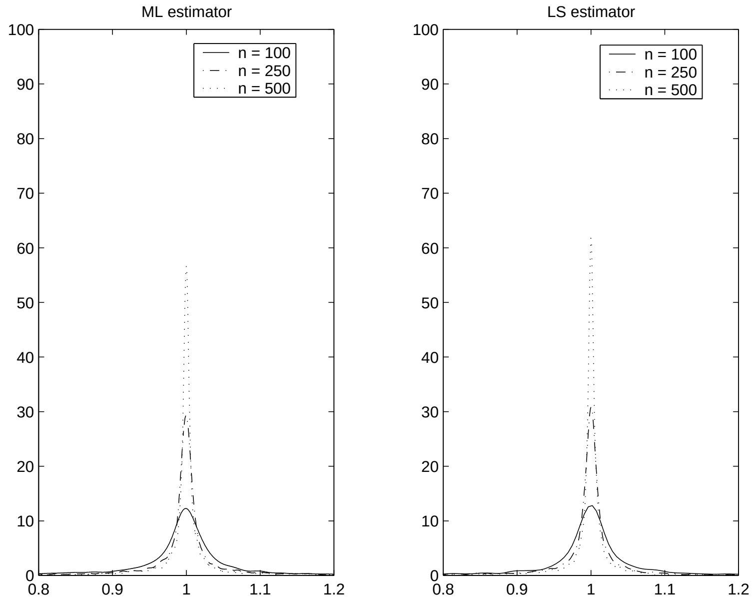

Figure 3: Densities of ML and truncated OLS estimators of β when $\beta_0 = 1$, $\sigma_0^2 = 1$

observations and computing a "Tobit-adjusted" Shin statistic.

The Tobit-adjusted Shin statistic we consider is

$$\hat{\sigma}_T^{-2} \int_0^1 G_T(r)^2 dr \tag{36}$$

where

$$G_T(r) = T^{-1/2} \sum_{t=1}^{[rT]} ((y_t - \hat{\beta}_T x_t)\mathbf{1}(y_t > 0) - \hat{\sigma}_T \phi(\hat{\beta}_T x_t / \hat{\sigma}_T)) \tag{37}$$

and $x_t \in \mathbb{R}$, $\beta \in \mathbb{R}$.

Our main result for this test statistic is as follows.

Theorem 3 *Under Assumption 1, 2 and 3,*

$$G_T(r) \Rightarrow G(r) \tag{38}$$

for some process $G(\cdot)$ that is mixed normal for any $r \in [0, 1]$.

The critical values of $\int_0^1 G(r)^2 dr/\sigma_0^2$ can be tabulated, and they are given in Table 1. These critical values were calculated by a simulation computer program based on 100,000,000 replications and $n = 1000$. Notice that the critical values are smaller than those for Shin's test for linear cointegration.

There are various possible applications of our nonstationary censored regression cointegration test to prices and quantities for agricultural commodities, where the presence of price

Table 1: Critical values of the Tobit-adjusted Shin's cointegration test

Fractile	Critical values
0.900	0.330
0.950	0.505
0.975	0.710
0.990	1.014

and/or quantity ceilings and price floors are common. Many agricultural products are protected with government support programs. For instance, the Milk Price Support Program (MPSP) was established in 1949 to provide farmers a parity level of income. To do so, "the MPSP purchases dairy products from processors and vendors to allow farmers to be paid the mandated support price for their milk"[1]. Since 1981, the mandated support price has been established by Congress either at specific price levels or by some formulas. The program has been active till now and is authorized till 2007. Therefore, in the market for dairy products, the observed price dynamics is subject to censoring, and the market prices are unobservable when they are below the support prices.

Due to the censored nature of price data in markets with support prices, Tobit regression seems to be an appropriate tool. However, in empirical work the possibility of nonstationarity of the data has been left undiscussed. For instance, Chavas and Kim (2004) uses a multivariate dynamic Tobit model to study the price dynamics of three dairy products, using inventory stocks as the exogenous explanatory variables. Both the prices and the inventory stocks data show I(1) features, but Chavas and Kim's paper treats the model as a stationary nonlinear regression, without any unit root test or cointegration test.

Below we use similar dataset as in Chavas and Kim (2004) and test the null of Tobit cointegration in dairy market. We consider the time period from 1991 to 2004, in monthly frequency (making for a sample size of 168). We chose to analyze American cheese in this paper because the support prices of cheese during these fourteen years have been almost constant and only mildly changed between \$1.10 and \$1.14 per pound (for blocks of 40 pounds). The monthly average prices and support prices are plotted in Figure 4. It can be seen that most of the time the market prices are above the support prices. The prices dropped to the support prices in 16 months, about 10% of the time.

The inventory stocks data (the sum of government stocks and commercial stocks) is plotted in Figure 5. Table 2 reports some unit root test statistics of the sequence. Both ADF test and Phillips-Perron test cannot reject the null of unit root. The KPSS test rejects the null of stationarity at 1% level. We use Schwarz information criterion in selecting number of lags when we conduct the tests. We also try different numbers of lags and the conclusion

[1]Source: Fact Sheet of Milk Price Support Program of USDA, July 2004. The data of this paper are also provided by USDA. The monthly prices are obtained from the Dairy Market News (USDA) and the inventory stocks are provided by Economic Research Service (USDA).

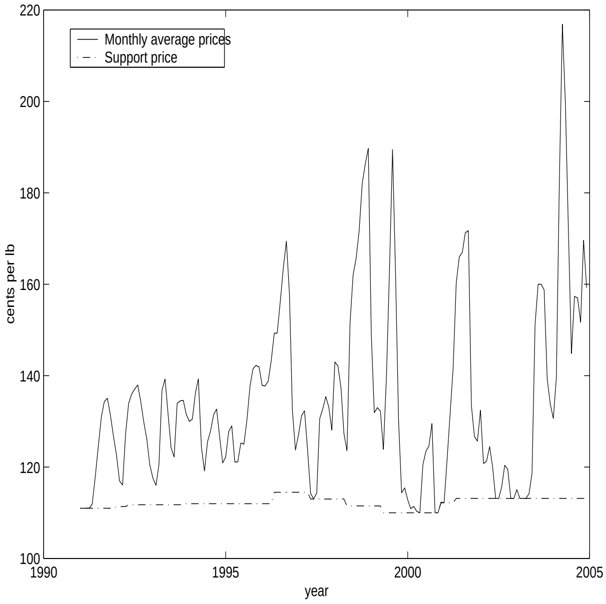

Figure 4: Monthly average prices and support prices of American cheese, 1991.1 - 2004.12

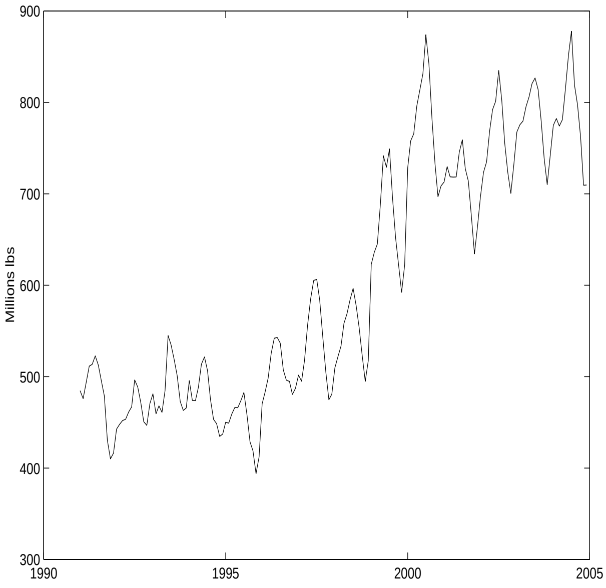

Figure 5: Inventory stocks of American cheese, 1991.1 - 2004.12

remain the same. This suggests that the data of inventory stocks is very likely to be unit root nonstationary, and therefore that inference based on conventional stationary Tobit model asymptotics is inappropriate.

Table 2: Unit root test statistics of the inventory stocks of American cheese

Test	Statistic	1% c.v.	5% c.v.	10% c.v.
ADF t-test	-1.034	-3.47	-2.88	-2.57
Phillips Perron t-test	-1.40	-3.47	-2.88	-2.57
KPSS test	1.48	0.74	0.46	0.38

In this application, we will test whether the market prices and inventory stocks are cointegrated. The model can be written as

$$p_t^* = \beta s_t + \varepsilon_t \tag{39}$$

and

$$p_t = \begin{cases} p_t^* & \text{when} \quad p_t^* > c, \\ c & \text{otherwise} \end{cases} \tag{40}$$

where p_t^* is the latent market prices, p_t is observed prices, c is the support prices, and s_t is the inventory stocks. This is a general Tobit model where the threshold is not necessarily zero. Our earlier results can be easily extended to incorporate this situation.

We first ran an OLS regression using the observations that were above the support price. We then computed the Tobit-adjusted Shin statistic, which equaled 2.6185. Since this value exceeded the critical value at the 1% level, we rejected the null of Tobit cointegration. Therefore, we concluded that regression (39) between prices and stocks of American cheese is spurious. Note though that in this application, the linear Shin test also rejected the null of cointegration.

6 Conclusion

In this paper we consider Tobit models with integrated regressors. We study both the ML estimator and the OLS estimator. Both estimators are superconsistent, and the ML estimator is asymptotically mixed normal. When there is only one regressor, the OLS estimator based on the positive observations is asymptotically equivalent to the ML estimator, and therefore asymptotically mixed normal. However when there is more than one integrated regressor, the OLS estimator is not asymptotically mixed normal, and standard testing procedures become asymptotically invalid. We also derive a nonlinear Tobit cointegration test based on the residuals from univariate LS censored regression.

The nonstationary Tobit model is potentially useful in time series settings where price ceilings, price floors or other forms of government interventions occur. We apply our cointegration test to prices of American cheese, and the test rejects the null of Tobit cointegration.

References

Amemiya, T. (1973) Regression Analysis When the Dependent Variable is Truncated Normal. *Econometrica* 41, 977-1016.

Billingsley, P. (1968) *Convergence of Probability Measures.* New York: Wiley.

Chang, Y. and J. Park (2004) Endogeneity in Nonlinear Regressions with Integrated Time Series. Mimeo, Rice University.

Chavas, J. and K. Kim (2004) A Heteroskedastic Multivariate Tobit Analysis of Price Dynamics in the Presence of Price Floors. *American Journal of Agricultural Economics* 86, 576-593.

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

de Jong, R.M. (2002) Nonlinear Minimization Estimators in the Presence of Cointegrating Relations. *Journal of Econometrics* 110, 241-259.

de Jong, R.M., C. Wang and Y. Bae (2005) Correlation Robust Threshold Unit Root Tests. Working paper.

Hall, P. and C.C. Heyde (1980) *Martignale Limit Theory and its Applications.* New York: Academic Press.

Park, J. and P.C.B. Phillips (1999) Asymptotics for Nonlinear Transformations of Integrated Time Series. *Econometric Theory* 15, 269-298.

Park, J. and P.C.B. Phillips (2000) Nonstationary Binary Choice. *Econometrica* 68, 1249-1280.

Park, J. and P.C.B. Phillips (2001) Nonlinear Regressions with Integrated Time Series. *Econometrica* 69, 1452-1498.

Park, J. and M. Shintani (2005) Testing for a Unit Root against Transitional Autoregressive Models. Working paper.

Pötscher B.M. (2004) Nonlinear Functions and Convergence to Brownian Motion: Beyond the Continuous Mapping Theorem. *Econometric Theory* 20, 1-22.

Shin, Y. (1994) A Residual-Based Test of the Null of Cointegration agaist the Alternative of No Cointegration. *Econometric Theory* 10, 91-115.

Wooldridge, J.M. (1994) Estimation and Inference for Dependent Processes. In *Handbook of Econometrics*, Vol. IV, ed. by R.F. Engle and D.L. McFadden. Amsterdam: North Holland.

A Useful Lemmas and Proofs

The following result forms the basis of our mixed normality result:

Lemma A.1 *Suppose that the probability space* $(\Omega_T, \mathcal{F}_T, P_T)$ *supports the square-integrable random variables* $S_{T1}, \ldots, S_{T,k_T}$*, and that the* S_{Tt} *are adapted to the sigma-fields* $\mathcal{F}_{Tt}$*, where* $\mathcal{F}_{T1} \subseteq \mathcal{F}_{T2} \subseteq \cdots \subseteq \mathcal{F}_{T,k_T} \subseteq \mathcal{F}_T$*. Let* $X_{Tt} = S_{Tt} - S_{T,t-1}$ $(S_{T0} = 0)$ *and* $U_{Tj}^2 = \sum_{t=1}^{j} X_{Tt}^2$*. If* $\mathcal{G}_T$ *is a sub-*σ*-field of* $\mathcal{F}_T$*, let* $\mathcal{G}_{Tt} = \mathcal{F}_{Tt} \vee \mathcal{G}_T$ *(the* σ*-field generated by* $\mathcal{F}_{Tt} \cup \mathcal{G}_T$*) and let* $\mathcal{G}_{T0} = \{\Omega_T, \phi\}$ *denote the trivial* σ*-field. Suppose further that*

$$\max_{1 \le t \le T} |X_{Tt}| \xrightarrow{p} 0, \qquad E(\max_{1 \le t \le T} X_{Tt}^2) \quad \text{is bounded in } T,$$

and that there exist σ*-fields* $\mathcal{G}_T \subseteq \mathcal{F}_T$ *and* $\mathcal{G}_T$*-measurable random variables* u_T^2 *such that*

$$U_{T,k_T}^2 - u_T^2 \xrightarrow{p} 0,$$

$$\sum_t E(X_{Tt}|\mathcal{G}_{T,t-1}) \xrightarrow{p} 0, \quad \text{and} \quad \sum_t |E(X_{Tt}|\mathcal{G}_{T,t-1})|^2 \xrightarrow{p} 0.$$

Then $U_{T,k_T}^{-1} S_{T,k_T}^2 \xrightarrow{d} N(0,1)$*, and if* $U_{T,k_T}^2 \xrightarrow{d} \eta^2$*,* $S_{T,k_T}^2 \xrightarrow{d} Z$*, where the characteristic function of the random variable* Z *satisfies* $E\exp(irZ) = \exp(-(1/2)\eta^2 r^2)$*.*

Proof of Lemma A.1: See Hall and Heyde (1980, Theorem 3.4). □

The following result regarding the boundedness of the density of $t^{-1/2}x_{1t}$ will be used at various locations in the proofs of this paper:

Lemma A.2 *Under Assumption 1, the density* $f_t(\cdot)$ *of* $t^{-1/2}x_t$ *satisfies* $\sup_{t \in \mathbb{N}, y \in \mathbb{R}} f_t(y) < \infty$*.*

Proof of Lemma A.2: This follows from Section 3.1 in Pötscher (2004). □

Lemma A.3 *Assume that Assumption 1 holds and that* $H(\cdot)$ *is Borel measurable and satisfies* $\int_{-\infty}^{\infty} |H(y)|dy < \infty$ *and that Assumption 1 holds. Then*

$$\sup_{T \ge 1} ET^{-1/2} \sum_{t=1}^{T} |H(x_t)| < \infty.$$

Proof of Lemma A.3: Let $f_t(\cdot)$ denote the density of $t^{-1/2}x_t$. Then

$$ET^{-1/2}\sum_{t=1}^{T}|H(x_t)| = T^{-1/2}\sum_{t=1}^{T}\int_{-\infty}^{\infty}|H(t^{1/2}y)|f_t(y)dy$$

$$\leq \sup_{t\in\mathbb{N},y\in\mathbb{R}} f_t(y)T^{-1/2}\sum_{t=1}^{T}t^{-1/2}\int_{-\infty}^{\infty}|H(y)|dy$$

$$\leq C\int_{-\infty}^{\infty}|H(y)|dy,$$

which proves the result. □

Lemma A.4 *Under Assumption 1,*

$$\sup_{T\geq 1} T^{-2}E\max_{1\leq t\leq T}|x_t|^4 < \infty.$$

Proof of Lemma A.4: By setting $a_k = |k+1|^{-1}|\log(|k|+1)|^{-2}$ in Equation (16.26) of Lemma 16.8 of Davidson (1994), for some constant C,

$$T^{-2}E\max_{1\leq t\leq T}|x_t|^4 \leq CT^{-2}(\sum_{k=0}^{\infty}|k+1|^3|\log(|k|+1)|^6 E|Y_{Tk}|^4)$$

where

$$Y_{Tk} = \sum_{t=1}^{T}(E(v_t|e_{t-k},e_{t-k-1},\ldots) - E(v_t|e_{t-k-1},e_{t-k-2},\ldots)) = \Pi_k\sum_{t=1}^{T}e_{t-k}.$$

Since $Ee_t^4 < \infty$ and e_t is i.i.d., it follows that $\sup_{T\geq 1}T^{-2}E|\sum_{t=1}^{T}e_{t-k}|^4 < \infty$. Therefore, for some constant C',

$$E\max_{1\leq t\leq T}|x_t|^4 \leq C'\sum_{k=0}^{\infty}|k+1|^3|\log(|k|+1)|^6|\Pi_k|^4,$$

and because $|\Pi_k| \leq C''k^{-2-\eta}$ for some $C'' > 0$ and $\eta > 0$, the last summation is finite. □

Lemma A.5 *Assume that Assumption 1 holds and that* $H_1 : \mathbb{R} \to \mathbb{R}$ *and* $H_2 : \mathbb{R} \to \mathbb{R}$ *are continuous, uniformly bounded, and measurable functions satisfying* $|H_1(y)| \leq C|y|^p$ *and* $\int_{-\infty}^{\infty}|y|^p|H_2(y)|dy < \infty$. *Let* c_T *denote a deterministic sequence such that* $c_T \to 0$ *as* $T \to \infty$. *Then*

$$T^{-1/2}\sum_{t=1}^{T}(H_1(x_t+c_T) - H_1(x_t))H_2(x_t) \xrightarrow{p} 0.$$

Proof of Lemma A.5: By Lemma A.2,

$$ET^{-1/2}\sum_{t=1}^{T}|H_1(x_t+c_T)-H_1(x_t)||H_2(x_t)|$$

$$\leq T^{-1/2}\sum_{t=1}^{T}\int_{-\infty}^{\infty}|H_1(t^{1/2}y+c_T)-H_1(t^{1/2}y)||H_2(t^{1/2}y)|f_t(y)dy,$$

and by substituting $x=t^{1/2}y$, it follows that the last expression can be bounded by

$$\sup_{t,y}|f_t(y)|T^{-1/2}\sum_{t=1}^{T}t^{-1/2}\int_{-\infty}^{\infty}|H_1(x+c_T)-H_1(x)||H_2(x)|dx.$$

Since $\sup_{T\geq 1}T^{-1/2}\sum_{t=1}^{T}t^{-1/2}<\infty$ and $|H_1(x+c_T)|\leq C(|x|+\sup_{T\geq 1}|c_T|)^p$ and $\int_{-\infty}^{\infty}|y|^p|H_2(y)|dy<\infty$ by assumption, by the bounded convergence theorem, the last expression converges to zero, which proves the result. □

Lemma A.6 *Assume that for all $K>0$ and some function $\nu(\cdot)$ and $G(\cdot)$*

$$\int_{-K}^{K}|\nu(T^{1/2})^{-1}G(T^{1/2}x)-H(x)|dx\to 0$$

and that the density $f_t(\cdot)$ of $t^{-1/2}y_{t1}$ satisfies $\sup_{t\in\mathbb{N},y\in\mathbb{R}}f_t(y)<\infty$. Also assume that $g(\cdot)$ is continuous on $\mathbb{R}^m$ and that $(T^{-1/2}y_{1,[rT]},T^{-1/2}y_{2,[rT]})\Rightarrow(W_1(r),W_2(r))$. Then

$$|\nu(T^{1/2})^{-1}T^{-1}\sum_{t=1}^{T}G(y_{t1})g(T^{-1/2}y_{t2})-T^{-1}\sum_{t=1}^{T}H(T^{-1/2}y_{t1})g(T^{-1/2}y_{t2})|\xrightarrow{p}0.$$

Proof of Lemma A.6: First observe that

$$|\nu(T^{1/2})^{-1}T^{-1}\sum_{t=1}^{T}G(y_{t1})g(T^{-1/2}y_{t2})-T^{-1}\sum_{t=1}^{T}H(T^{-1/2}y_{t1})g(T^{-1/2}y_{t2})|$$

$$\leq\sup_{1\leq t\leq T}|g(T^{-1/2}y_{t2})|T^{-1}\sum_{t=1}^{T}|\nu(T^{1/2})^{-1}G(y_{t1})-H(T^{-1/2}y_{t1})|.$$

Since $\sup_{1\leq t\leq T}|T^{-1/2}y_{t2}|=O_p(1)$ and because $g(\cdot)$ is continuous under the assumptions of this lemma, it suffices to show that for all $K>0$,

$$T^{-1}\sum_{t=1}^{T}E|\nu(T^{1/2})^{-1}G(y_{t1})-H(T^{-1/2}y_{t1})|\mathbf{1}(T^{-1/2}|y_{t1}|\leq K)=o(1).$$

Because the density $f_t(\cdot)$ of $t^{-1/2}y_{t1}$ is well-defined and satisfies $\sup_{t\in\mathbb{N},y\in\mathbb{R}} f_t(y) < \infty$ by assumption, setting $x = T^{-1/2}t^{1/2}y$,

$$T^{-1}\sum_{t=1}^{T} E|\nu(T^{1/2})^{-1}G(y_{t1}) - H(T^{-1/2}y_{t1})|\mathbf{1}(T^{-1/2}|y_{t1}| \leq K)$$

$$= T^{-1}\sum_{t=1}^{T}\int_{-\infty}^{\infty} |\nu(T^{1/2})^{-1}G(t^{1/2}y) - H(T^{-1/2}t^{1/2}y)|\mathbf{1}(T^{-1/2}t^{1/2}|y| \leq K)f_t(y)dy$$

$$\leq \sup_{t\in\mathbb{N},y\in\mathbb{R}} f_t(y) \sup_{T\in\mathbb{N}}(T^{-1/2}\sum_{t=1}^{T} t^{-1/2}) \int_{-K}^{K} |\nu(T^{1/2})G(T^{1/2}x) - H(x)|dx \to 0$$

by assumption. This completes the proof. □

To study the asymptotic behavior of the Hessian, define

$$\tilde{H}_{\beta\beta'}(\theta_0) = -\sigma_0^{-2}\sum_{t=1}^{T}\mathbf{1}(x_t'\beta_0 + \varepsilon_t > 0)x_t x_t',$$

$$\tilde{H}_{\sigma^2\beta'}(\theta_0) = -\sigma_0^{-4}\sum_{t=1}^{T}\varepsilon_t\mathbf{1}(x_t'\beta_0 + \varepsilon_t > 0)x_t,$$

$$\tilde{H}_{\sigma^2\sigma^2}(\theta_0) = (1/2)\sigma_0^{-4}\sum_{t=1}^{T}\mathbf{1}(x_t'\beta_0 + \varepsilon_t > 0) - \sigma_0^{-6}\sum_{t=1}^{T}\varepsilon_t^2\mathbf{1}(x_t'\beta_0 + \varepsilon_t > 0),$$

and

$$\breve{H}_{\beta\beta'}(\theta_0) = -\sigma_0^{-2}\sum_{t=1}^{T}\Phi(\sigma_0^{-1}\beta_0'x_t)x_t x_t',$$

$$\breve{H}_{\sigma^2\beta'}(\theta_0) = -\sigma_0^{-3}\sum_{t=1}^{T}\phi(\sigma_0^{-1}\beta_0'x_t)x_t,$$

$$\breve{H}_{\sigma^2\sigma^2}(\theta_0) = (1/2)\sigma_0^{-4}\sum_{t=1}^{T}\Phi(\sigma_0^{-1}\beta_0'x_t) - \sigma_0^{-4}\sum_{t=1}^{T}(\Phi(\sigma_0^{-1}\beta_0'x_t) - \phi(\sigma_0^{-1}\beta_0'x_t)\sigma_0^{-1}x_{1t}\alpha_0^1).$$

The asymptotic equivalence of rescaled versions of the $H_{\cdot,\cdot}(\cdot)$ to the $\breve{H}_{\cdot,\cdot}(\cdot)$ can now be established:

Lemma A.7 *Let Assumption 1 and 2 hold. Then*

$$T^{-2}A'(H_{\beta\beta'}(\theta_0) - \tilde{H}_{\beta\beta'}(\theta_0))A \xrightarrow{p} 0,$$

$$T^{-3/2}A'(H_{\sigma^2\beta'}(\theta_0) - \tilde{H}_{\sigma^2\beta'}(\theta_0)) \xrightarrow{p} 0,$$

$$T^{-1}(H_{\sigma^2\sigma^2}(\theta_0) - \tilde{H}_{\sigma^2\sigma^2}(\theta_0)) \xrightarrow{p} 0,$$

and

$$T^{-2}A'(\tilde{H}_{\beta\beta'}(\theta_0) - \breve{H}_{\beta\beta'}(\theta_0))A \xrightarrow{p} 0,$$

$$T^{-3/2}A'(\tilde{H}_{\sigma^2\beta'}(\theta_0) - \breve{H}_{\sigma^2\beta'}(\theta_0)) \xrightarrow{p} 0,$$

$$T^{-1}(\tilde{H}_{\sigma^2\sigma^2}(\theta_0) - \breve{H}_{\sigma^2\sigma^2}(\theta_0)) \xrightarrow{p} 0.$$

Proof of Lemma A.7: The first step in this lemma consists of verifying the conditions of Lemma 1, and the second step is a martingale difference result. To see that $g_1(\cdot)(1-\Phi(y))$, $g_2(\cdot)(1-\Phi(y))$ and $g_3(\cdot)(1-\Phi(y))$ are absolutely integrable, note that this is straightforward as $y \to -\infty$ and $\phi(y)(1-\Phi(y))^{-1} = O(y)$ as $y \to \infty$ (see Park and Phillips (2000, p.1254)). Therefore, for y large enough and some constant C, using the definitions of $g_1(\cdot)$, $g_2(\cdot)$ and $g_3(\cdot)$ of Section 2.2,

$$|g_1(y)|(1-\Phi(y)) \leq Cy|(1-\Phi(y))y - \phi(y)|,$$

and the last function is clearly integrable. For $g_2(\cdot)$ and $g_3(\cdot)$, a similar argument holds. To show the second part of the lemma, note that since $\max_{1\leq t\leq T} T^{-1/2}|x_t| = O_p(1)$, it suffices to consider for each $K > 0$ the truncated difference between $\tilde{H}_{\beta\beta'}(\theta_0)$ and $\breve{H}_{\beta\beta'}(\theta_0)$, viz.

$$T^{-2}\sigma_0^{-2}\sum_{t=1}^{T}(\mathbf{1}(x_t'\beta_0 + \varepsilon_t > 0) - \Phi(\beta_0'x_t/\sigma_0))x_t x_t'\mathbf{1}(T^{-1/2}|x_t| \leq K).$$

The squared expectation of each element of the above matrix can be bounded by

$$T^{-2}\sigma_0^{-4}K^4\sum_{t=1}^{T}E(1-\Phi(\beta_0'x_t/\sigma_0))\Phi(\beta_0'x_t/\sigma_0) \leq T^{-1}\sigma_0^{-4}K^4 \to 0,$$

and an analogous proof is easily provided for both other terms. □

Lemma A.8 *Let* $h(x) = (\phi(x)/(1-\Phi(x)))^m x^n$ *where* m *and* n *are positive integers. For any* $\varepsilon > 0$, *let*

$$\bar{h}(x) = \sup_{|a-1|<\varepsilon}\sup_{|b|<\varepsilon}|h(ax+b)|.$$

Then

$$\int_{-\infty}^{\infty}\bar{h}(x)(1-\Phi(x))dx < \infty.$$

Proof of Lemma A.8: First notice that the function is integrable for $x < 0$. Therefore below we just consider integration over $(0, \infty)$. Let $\tilde{x}$ be an intermediate value between x and $(1+\varepsilon)x + \varepsilon$. Then by the Taylor series expansion

$$\begin{aligned}
& \int_0^\infty \bar{h}(x)(1-\Phi(x))dx \\
= & \int_0^\infty \sup_{|a-1|<\varepsilon} \sup_{|b|<\varepsilon} |h(x) + [m\phi(\tilde{x})^{m-1}(1-\Phi(\tilde{x}))^{-(m-1)}(\phi(\tilde{x})(1-\Phi(\tilde{x}))^{-1} \\
& -\phi(\tilde{x})^2(1-\Phi(\tilde{x}))^{-2}) - n\phi(\tilde{x})^m(1-\Phi(\tilde{x}))^{-m}\tilde{x}^{n-1}]((a-1)x+b)/2|(1-\Phi(x))dx \\
\leq & \int_0^\infty |h(x)|(1-\Phi(x))dx + \int_0^\infty C[(1+\varepsilon)x+\varepsilon]^{m+n-1}(x\varepsilon+\varepsilon)(1-\Phi(x))dx
\end{aligned}$$

where C is a constant larger than $m+n$. Clearly the first term is integrable; the second term is integrable too since

$$\int_0^\infty x^m(1-\Phi(x))dx < \infty$$

for any positive integer m. This finishes the proof. □

The following lemmas concern our Tobit cointegration test.

Lemma A.9 $\sup_{r\in[0,1]} |G_T(r) - H_T(r)| \xrightarrow{p} 0,$

where $G_T(r)$ is defined as in (37) and

$$H_T(r) = (T^{-1/2}\sum_{t=1}^{[rT]} \varepsilon_t \mathbf{1}(x_t > 0))$$

$$-(T^{-1}\sum_{t=1}^{T} \varepsilon_t x_t \mathbf{1}(x_t > 0))(T^{-3/2}\sum_{t=1}^{[rT]} x_t \mathbf{1}(x_t > 0))/(T^{-2}\sum_{t=1}^{T} x_t^2 \mathbf{1}(x_t > 0).$$

Proof of Lemma A.9: From the reasoning at the end of the proof of Lemma 5, it follows that

$$T^{-1/2}\sum_{t=1}^{T} |\hat{\sigma}_T\phi(\hat{\beta}_T x_t/\hat{\sigma}_T)) - \sigma_0\phi(\beta_0 x_t/\sigma)| = o_p(1),$$

implying that we only need to show

$$\sup_{r\in[0,1]} |T^{-1/2}\sum_{t=1}^{[rT]} \varepsilon_t(\mathbf{1}(y_t > 0) - \sigma_0\phi(\beta_0 x_t/\sigma_0) - \mathbf{1}(x_t > 0))| = o_p(1),$$

$$T^{-1}\sum_{t=1}^{T}\varepsilon_t x_t(\mathbf{1}(x_t>0)-\mathbf{1}(y_t>0))=o_p(1),$$

$$T^{-3/2}\sum_{t=1}^{T}|x_t||\mathbf{1}(x_t>0)-\mathbf{1}(y_t>0)|=o_p(1),$$

and

$$T^{-2}\sum_{t=1}^{T}x_t^2|\mathbf{1}(x_t>0)-\mathbf{1}(y_t>0)|=o_p(1).$$

The last two expressions converge to zero in probability if

$$T^{-1}\sum_{t=1}^{T}E|\mathbf{1}(x_t>0)-\mathbf{1}(y_t>0)|^2\to 0.$$

This follows because

$$T^{-1}\sum_{t=1}^{T}E|\mathbf{1}(x_t>0)-\mathbf{1}(\varepsilon_t+\beta_0 x_t>0)|^2$$

$$=T^{-1}\sum_{t=1}^{T}(E\mathbf{1}(x_t>0)-2E\mathbf{1}(x_t>0)\Phi(\beta_0 x_t/\sigma_0)+E\Phi(\beta_0 x_t/\sigma_0))$$

$$\leq 3ET^{-1}\sum_{t=1}^{T}|\mathbf{1}(x_t>0)-\Phi(\beta_0 x_t/\sigma_0)|,$$

and because $|\mathbf{1}(x>0)-\Phi(\beta_0 x/\sigma)|$ is an integrable function, it follows that the last expression is $O(T^{-1/2})$. Also, to deal with the second expression, note that

$$T^{-1}\sum_{t=1}^{T}\varepsilon_t x_t(\mathbf{1}(x_t>0)-\mathbf{1}(y_t>0))$$

$$=T^{-1}\sum_{t=1}^{T}(\varepsilon_t x_t(\mathbf{1}(x_t>0)-\mathbf{1}(y_t>0))-\sigma_0 x_t\phi(\beta_0 x_t/\sigma_0))+T^{-1}\sum_{t=1}^{T}x_t\sigma_0\phi(\beta_0 x_t/\sigma_0),$$

and because $|x|\phi(\beta_0 x/\sigma_0)$ is integrable, the second term is $O_p(T^{-1/2})$; also, the first term has a mean of zero and a variance of

$$T^{-2}\sum_{t=1}^{T}E\varepsilon_t^2 x_t^2(\mathbf{1}(x_t>0)-\mathbf{1}(y_t>0)-\sigma_0\phi(\beta_0 x_t/\sigma_0))^2$$

which is $O(T^{-1/2})$ by the earlier arguments. Finally, to show that the first term is $o_p(1)$, note that pointwise convergence holds by a simple variance calculation and Lemma A.3, because summands are mean zero martingale differences. Therefore, it remains to show stochastic equicontinuity to strengthen the pointwise convergence in probability to uniform convergence in probability. To show stochastic equicontinuity of a process $J_T(\cdot)$, it suffices to show

$$E|J_T(r) - J_T(r')|^4 \leq C|r - r'|^2;$$

see for example Billingsley (1968, p.95). Such a result can be established in our case by noting that, by Burkholder's inequality and the norm inequality,

$$E|T^{-1/2} \sum_{t=[r'T]+1}^{[rT]} \varepsilon_t(\mathbf{1}(y_t > 0) - \sigma_0\phi(\beta_0 x_t/\sigma_0) - \mathbf{1}(x_t > 0))|^4$$

$$\leq CE|T^{-1} \sum_{t=[r'T]+1}^{[rT]} \varepsilon_t^2(\mathbf{1}(x_t > 0) - \mathbf{1}(y_t > 0) - \sigma_0\phi(\beta_0 x_t/\sigma_0)|^2$$

$$\leq C'|r' - r|^2 E\varepsilon_t^4.$$

This completes the proof. □

Lemma A.10 *Under Assumption 1, 2 and 3, for each* $(r_1, r_2) \in [0, 1]^2$*,*

$$(T^{-1/2} \sum_{t=1}^{[r_1 T]} \varepsilon_t \mathbf{1}(x_t > 0), T^{-1} \sum_{t=1}^{T} \varepsilon_t x_t \mathbf{1}(x_t > 0), T^{-1/2} x_{[r_2 T]})$$

$$\xrightarrow{d} (\sigma_0(Z_1, Z_2)B(r_1)^{1/2}, V_1(r_2))$$

where (Z_1, Z_2) *is bivariate normal,* Z_1 *and* Z_2 *are independent of each other, and* (Z_1, Z_2) *is independent of* $V_1(\cdot)$*, and*

$$B(r_1) = \begin{pmatrix} \int_0^{r_1} \mathbf{1}(V_1(r) > 0)dr & \int_0^{r_1} V_1(r)\mathbf{1}(V_1(r) > 0)dr) \\ \int_0^{r_1} V_1(r)\mathbf{1}(V_1(r) > 0)dr) & \int_0^{r_1} V_1(r)^2\mathbf{1}(V_1(r) > 0)dr) \end{pmatrix}.$$

Proof of Lemma A.10:

We show convergence in distribution of $\lambda_1 T^{-1/2}\sum_{t=1}^{[r_1T]}\varepsilon_t\mathbf{1}(x_t>0)+\lambda_2T^{-1/2}x_{[r_2T]}$ for any (λ_1,λ_2); including $T^{-1}\sum_{t=1}^{T}\varepsilon_t x_t\mathbf{1}(x_t>0)$ does not add any further complications, but will complicate notation. The characteristic function of this expression is, by conditioning on $(x_1,\ldots,x_T)$ and the normality of ε_t given $x_1,\ldots,x_T$,

$$E\exp(i\lambda_1T^{-1/2}\sum_{t=1}^{[r_1T]}\varepsilon_t\mathbf{1}(x_t>0)+\lambda_2T^{-1/2}x_{[r_2T]})$$

$$=EE(\exp(i\lambda_1T^{-1/2}\sum_{t=1}^{[r_1T]}\varepsilon_t\mathbf{1}(x_t>0)+\lambda_2T^{-1/2}x_{[r_2T]}|x_1,\ldots,x_T)$$

$$=E\exp(i\lambda_2T^{-1/2}x_{[r_2T]})\prod_{t=1}^{[r_1T]}\exp(-(1/2)\sigma_0^2\lambda_1^2T^{-1}\mathbf{1}(x_t>0))$$

$$=E\exp(i\lambda_2T^{-1/2}x_{[r_2T]})\exp(-(1/2)\lambda_1^2\sigma_0^2T^{-1}\sum_{t=1}^{[r_1T]}\mathbf{1}(x_t>0)).$$

By the joint convergence in distribution of both terms, the last expression converges to

$$E\exp(i\lambda_2V_1(r_2)+\lambda_1^2\sigma_0^2\int_0^{r_1}\mathbf{1}(V_1(r)>0)dr),$$

which corresponds to the joint distribution of $(\sigma_0Z(\int_0^{r_1}\mathbf{1}(V_1(r)>0)dr)^{1/2},V_1(r_2))$. □

B Proofs of the main results

Proof of Lemma 1: Write

$$T^{-(1+\kappa/2)}\sum_{t=1}^{T}g(x_{1t})\mathbf{1}(x_t'\beta_0+\varepsilon_t\le 0)x_{2t}^{\kappa}$$

$$=T^{-(1+\kappa/2)}\sum_{t=1}^{T}g(x_{1t})\mathbf{1}(x_t'\beta_0+\varepsilon_t\le 0)x_{2t}^{\kappa}\mathbf{1}(T^{-1/2}|x_{2t}|\le K)$$

$$+T^{-(1+\kappa/2)}\sum_{t=1}^{T}g(x_{1t})\mathbf{1}(x_t'\beta_0+\varepsilon_t\le 0)x_{2t}^{\kappa}\mathbf{1}(T^{-1/2}|x_{2t}|>K).$$

Since $\sup_{1\leq t\leq T} T^{-1/2}|x_{2t}| = O_p(1)$, the second term can be made negligible by choosing K large enough. Using the law of iterated expectations, the absolute expectation of the first expression can be bounded by

$$T^{-1}\sum_{t=1}^{T} E|g(x_{1t})|(1-\Phi(x_{1t}\alpha_0^1))K^{\kappa},$$

and because $\int_{-\infty}^{\infty}|g(y)|(1-\Phi(y))dy < \infty$ by assumption, the result now follows from Lemma A.3. This completes the proof. □

Proof of Lemma 2: We apply Lemma A.6 to show the six results and we will use $\nu(\lambda) = \lambda^{\kappa}$ in this lemma. From Lemma A.2, the assumed uniform boundedness of the densities follows. Weak convergence follows from Assumption 1. The result of (a) follows by setting $G(s_1) = F(s_1)$ and $g(s_2) = 1$. For (b), we use $G(s_1) = s_1\Phi(s_1)$ and $g(s_2) = 1$; for (c), set $G(s_1) = s_1^2\Phi(s_1)$ and $g(s_2) = 1$, and $\kappa = 0$ for these cases. To show (d), we use $G(s_1) = \Phi(s_1)$ and $g(s_2) = s_2$, and $\kappa = 1$; for (e), $G(s_1) = \Phi(s_1)$ and $g(s_2) = s_2 s_2'$, for $\kappa = 2$; and for (f), $G(s_1) = \Phi(s_1)s_1$, $g(s_2) = s_2$, and $\kappa = 1$. For all these cases, we have for all $K > 0$

$$\int_{-K}^{K} |\nu(T^{1/2})^{-1}G(T^{1/2}x) - H(x)|dx \to 0.$$

This completes the proof. □

Proof of Lemma 3: By Lemma A.7, it suffices to find the limit distributions of $\breve{H}_{\beta\beta'}(\theta_0)$, $\breve{H}_{\sigma^2\beta'}(\theta_0)$, and $\breve{H}_{\sigma^2\sigma^2}(\theta_0)$. The result for $\breve{H}_{\beta\beta'}(\theta_0)$ follows from an application of Lemma 2. To show the result for $\breve{H}_{\sigma^2\beta'}(\theta_0)$, we observe that this statistic consists of two parts, viz. $\sigma_0^{-2}\sum_{t=1}^{T}\phi(\sigma_0^{-1}\alpha_0^1 x_{t1})x_{t1}$ and $\sigma_0^{-2}\sum_{t=1}^{T}\phi(\sigma_0^{-1}\alpha_0^1 x_{t1})x_{t2}$. By Lemma 2 of Park and Phillips (2000), p. 1256, it follows that the first term is $O_p(T^{1/2})$, while the second is $O_p(T)$, which establishes the result. To show the result for $\breve{H}_{\sigma^2\sigma^2}(\theta_0)$, first note that $T^{-1/2}\sum_{t=1}^{T}\phi(\sigma_0^{-1}x_{1t}\alpha_0^1)x_{1t}\alpha_0^1$ converges in distribution by Lemma 2 of Park and Phillips (2000) because the function $\phi(y)y$ is integrable. By Lemma A.7, the distributional limit now follows for this term, which completes the proof. □

Proof of Lemma 4: We will apply Lemma A.1. Define $\mathcal{G}_T = \sigma(x_1,\ldots,x_T)$ and $\mathcal{F}_{Tj} = \sigma(\varepsilon_1,\ldots,\varepsilon_j,x_1,\ldots,x_T)$. Note that the summands of $(S_\beta(\theta_0), S_{\sigma^2}(\theta_0))$ are martingale differences with respect to the $\mathcal{G}_{Tj}$ as defined in Lemma A.1. We will show that $\lambda_1' S_\beta(\theta_0) +$

$\lambda_2 S_{\sigma^2}(\theta_0)$ converges in distribution to the appropriate limit, from which the result follows by the Cramèr-Wold device. Then for any $\lambda = (\lambda_1', \lambda_2)'$,

$$U_{TT}^2 = \sigma_0^{-2} T^{-2} \sum_{t=1}^{T} \phi(\beta_0' x_t/\sigma_0)^2 (1 - \Phi(\beta_0' x_t/\sigma_0))^{-2} (\lambda_1' x_t)^2 \mathbf{1}(x_t'\beta_0 + \varepsilon_t \leq 0)$$

$$+\sigma_0^{-4} T^{-2} \sum_{t=1}^{T} \varepsilon_t^2 (\lambda_1' x_t)^2 \mathbf{1}(x_t'\beta_0 + \varepsilon_t > 0)$$

$$+\lambda_2^2 \sigma_0^{-6} (1/4) T^{-1} \sum_{t=1}^{T} (\beta_0' x_t)^2 \phi(\beta_0' x_t/\sigma_0)^2 (1 - \Phi(\beta_0' x_t/\sigma_0))^{-2} \mathbf{1}(x_t'\beta_0 + \varepsilon_t \leq 0)$$

$$+\lambda_2^2 \sigma_0^{-8} T^{-1} \sum_{t=1}^{T} (\varepsilon_t^2 - \sigma_0^2)^2 \mathbf{1}(x_t'\beta_0 + \varepsilon_t > 0)$$

$$+\lambda_2 \sigma_0^{-4} (1/2) T^{-3/2} \sum_{t=1}^{T} (\beta_0' x_t)(\lambda_1' x_t) \phi(\beta_0' x_t/\sigma_0)^2 (1 - \Phi(\beta_0' x_t/\sigma_0))^{-2} \mathbf{1}(x_t'\beta_0 + \varepsilon_t \leq 0)$$

$$+\lambda_2 \sigma_0^{-6} T^{-3/2} \sum_{t=1}^{T} \varepsilon_t (\lambda_1' x_t) \mathbf{1}(x_t'\beta_0 + \varepsilon_t > 0)(\varepsilon_t^2 - \sigma_0^2), \tag{41}$$

and define

$$u_T^2 = \sigma_0^{-2} T^{-2} \sum_{t=1}^{T} (\lambda_1' x_t)^2 (\Phi(\beta_0' x_t/\sigma_0) - (\sigma_0^{-1} \beta_0' x_t) \phi(\sigma_0^{-1} \beta_0' x_t))$$

$$+\lambda_2^2 \sigma_0^{-4} T^{-1} \sum_{t=1}^{T} (-(x_t'\beta_0/\sigma_0)^3 \phi(x_t'\beta_0/\sigma_0) + 2\Phi(x_t'\beta_0/\sigma_0) - (x_t'\beta_0/\sigma_0)\phi(x_t'\beta_0/\sigma_0)). \tag{42}$$

The difference $U_{TT}^2 - u_T^2$ is now a sum of martingale differences; to see this, note that it is used here that for any nonrandom $a \in \mathbb{R}$,

$$E\varepsilon \mathbf{1}(\varepsilon > -a) = \sigma_0 \phi(a/\sigma_0),$$

$$E\varepsilon^2 \mathbf{1}(\varepsilon > -a) = \sigma_0^2 (\Phi(a/\sigma_0) - (a/\sigma_0)\phi(a/\sigma_0)),$$

$$E\varepsilon^3 \mathbf{1}(\varepsilon > -a) = \sigma_0^3 ((a/\sigma_0)^2 \phi(a/\sigma_0) + 2\phi(a/\sigma_0)),$$

and

$$E\varepsilon^4 \mathbf{1}(\varepsilon > -a) = \sigma_0^4 (-(a/\sigma_0)^3 \phi(a/\sigma_0) + 3\Phi(a/\sigma_0) - 3(a/\sigma_0)\phi(a/\sigma_0)).$$

To show that

$$U_{TT}^2 - u_T^2 \xrightarrow{p} 0,$$

first observe that since $\max_{1\le t\le T} T^{-1/2}|x_t| = O_p(1)$, for showing that the first term of U_{TT}^2 is negligible, it suffices to show that

$$ET^{-1}\sum_{t=1}^{T}\phi(\beta_0' x_t/\sigma_0)^2(1-\Phi(\beta_0' x_t/\sigma_0))^{-2}\mathbf{1}(x_t'\beta_0+\varepsilon_t\le 0)\to 0$$

and by Lemma 1 and the integrability of $|x|\phi(x)$, this result easily follows. A similar argument holds for the third and fifth term of Equation (41). To show that the second term of Equation (41) is asymptotically equivalent to the second term of Equation (42), again using that $\max_{1\le t\le T} T^{-1/2}|x_t| = O_p(1)$, note that it suffices to show that

$$T^{-2}\sum_{t=1}^{T}(\varepsilon_t^2\mathbf{1}(x_t'\beta_0+\varepsilon_t\le 0) - E(\varepsilon_t^2\mathbf{1}(x_t'\beta_0+\varepsilon_t\le 0)|x_1,\ldots,x_T))(\lambda_1' x_t)^2\mathbf{1}(T^{-1/2}|x_t|\le K)$$
$$\xrightarrow{p} 0.$$

This result follows from a simple variance calculation and because $E\varepsilon_t^4 < \infty$. A similar argument holds for the fourth term. For the sixth term, a similar argument applies. Also, to show that $\max_{1\le t\le T}|X_{Tt}| \xrightarrow{p} 0$, note that again we need to deal with six terms. We can assume again that $\max_{1\le t\le T} T^{-1/2}|x_t|\le K$ and for the first term, it suffices to show that for all $K>0$,

$$P(|\lambda_1|K\max_{1\le t\le T} T^{-1/2}\phi(\beta_0' x_t/\sigma_0)(1-\Phi(\beta_0' x_t/\sigma_0))^{-1}\mathbf{1}(x_t'\beta_0+\varepsilon_t\le 0)>\delta)$$

$$\le \sum_{t=1}^{T} P(T^{-1/2}K|\lambda_1|\phi(\beta_0' x_t/\sigma_0)(1-\Phi(\beta_0' x_t/\sigma_0))^{-1}\mathbf{1}(x_t'\beta_0+\varepsilon_t\le 0)>\delta)$$

$$\le \delta^{-2}K^2|\lambda_1|^2\sum_{t=1}^{T} T^{-1}E\phi(\beta_0' x_t/\sigma_0)^2(1-\Phi(\beta_0' x_t/\sigma_0))^{-2}\mathbf{1}(x_t'\beta_0+\varepsilon_t\le 0)\to 0$$

by Lemma 1. For the five other terms, similar arguments apply. In addition

$$E(\max_{1\le t\le T} X_{Tt}^2) = O(T^{-2}E(\max_{1\le t\le T}|x_{t1}|^2\max_{1\le t\le T}|x_t|^2)) = O(T^{-2}E\max_{1\le t\le T}|x_t|^4),$$

and by Lemma A.4, the last expression is $O(1)$, as required.
The result now follows by observing that by construction, $E(X_{Tt}|\mathcal{G}_{T,t-1}) = 0$ because by Assumption 2, $\varepsilon_t|x_1,\ldots,x_T$ is distributed $N(0,\sigma_0^2)$.
Also,

$$u_T^2 \xrightarrow{d} \sigma_0^{-4}\int_0^1(\lambda_1' V(r))^2\mathbf{1}(V_1(r)>0)dr + 2\lambda_2^2\sigma_0^{-4}\int_0^1\mathbf{1}(V_1(r)>0)dr.$$

Therefore, for any λ, $\lambda_1' S_\beta(\theta_0)+\lambda_2 S_{\sigma^2}(\theta_0)$ converges in distribution to a mixed normal limit, and the Cramèr-Wold device therefore renders the result of the theorem. □

Proof of Lemma 5:

By the Taylor expansion,

$$0 = (\partial/\partial\beta)\log L_T((\hat{\beta}_T', \hat{\sigma}_T^2)') = (\partial/\partial\beta)\log L_T((\beta_0', \hat{\sigma}_T^2)')$$
$$+(\partial^2/\partial\beta\partial\beta')\log L_T((\tilde{\beta}_T', \hat{\sigma}_T^2)')(\hat{\beta}_T - \beta_0),$$

so (presuming the matrix is invertible)

$$T(\hat{\beta}_T - \beta_0) = [(\partial^2/\partial\beta\partial\beta')\log L_T((\tilde{\beta}_T', \hat{\sigma}_T^2)')]^{-1}(\partial/\partial\beta)\log L_T((\beta_0', \hat{\sigma}_T^2)')$$
$$= [-T^{-2}\sum_{t=1}^T g_1(x_t'\tilde{\beta}_T/\hat{\sigma}_T)\mathbf{1}(x_t'\beta_0 + \varepsilon_t \leq 0)x_t x_t' - T^{-2}\sum_{t=1}^T \mathbf{1}(x_t'\beta_0 + \varepsilon_t > 0)x_t x_t']^{-1}$$
$$\times[-\hat{\sigma}_T T^{-1}\sum_{t=1}^T \phi(\beta_0' x_t/\hat{\sigma}_T)(1 - \Phi(\beta_0' x_t/\hat{\sigma}_T))^{-1} x_t \mathbf{1}(x_t'\beta_0 + \varepsilon_t \leq 0)$$
$$+T^{-1}\sum_{t=1}^T (y_t - x_t'\beta_0)x_t\mathbf{1}(x_t'\beta_0 + \varepsilon_t > 0)].$$

The proof that $T(\hat{\beta}_T - \beta_0) = O_p(1)$ is now complete if we can show the following four results:

1. $T^{-2}\sum_{t=1}^T g_1(x_t'\beta/\sigma)\mathbf{1}(x_t'\beta_0 + \varepsilon_t \leq 0)x_t x_t'$ is positive semidefinite for any $(\beta, \sigma) \in B \times \Sigma$;
2. $T^{-2}\sum_{t=1}^T \mathbf{1}(x_t'\beta_0 + \varepsilon_t > 0)x_t x_t'$ converges in distribution to an a.s. positive definite matrix;
3. $\sup_{\sigma\in\Sigma} T^{-1}\sum_{t=1}^T \phi(\beta_0' x_t/\sigma)(1 - \Phi(\beta_0' x_t/\sigma))^{-1}|x_t|\mathbf{1}(x_t'\beta_0 + \varepsilon_t \leq 0) = O_p(1)$;
4. $T^{-1}\sum_{t=1}^T \varepsilon_t x_t \mathbf{1}(x_t'\beta_0 + \varepsilon_t > 0)$ converges in distribution.

Result 1 follows because $g_1(\cdot) \geq 0$. Result 2 was derived in the proofs of Lemma A.7 and Lemma 4. Result 4 was derived in Lemma 4. To show the remaining result 3, write

$$T^{-1}\sum_{t=1}^T \phi(\beta_0' x_t/\sigma)(1 - \Phi(\beta_0' x_t/\sigma))^{-1}\mathbf{1}(x_t'\beta_0 + \varepsilon_t \leq 0)|x_t|$$
$$= T^{-1}\sum_{t=1}^T \phi(\beta_0' x_t/\sigma)(1 - \Phi(\beta_0' x_t/\sigma))^{-1}\mathbf{1}(x_t'\beta_0 + \varepsilon_t \leq 0)\mathbf{1}(\beta_0' x_t > 0)|x_t|\mathbf{1}(T^{-1/2}|x_t| \leq K)$$
$$+T^{-1}\sum_{t=1}^T \phi(\beta_0' x_t/\sigma)(1 - \Phi(\beta_0' x_t/\sigma))^{-1}\mathbf{1}(x_t'\beta_0 + \varepsilon_t \leq 0)\mathbf{1}(\beta_0' x_t > 0)|x_t|\mathbf{1}(T^{-1/2}|x_t| > K)$$

$$+T^{-1}\sum_{t=1}^{T}\phi(\beta_0'x_t/\sigma)(1-\Phi(\beta_0'x_t/\sigma))^{-1}\mathbf{1}(x_t'\beta_0+\varepsilon_t\leq 0)\mathbf{1}(\beta_0'x_t\leq 0)|x_t|\mathbf{1}(T^{-1/2}|x_t|\leq K)$$

$$+T^{-1}\sum_{t=1}^{T}\phi(\beta_0'x_t/\sigma)(1-\Phi(\beta_0'x_t/\sigma))^{-1}\mathbf{1}(x_t'\beta_0+\varepsilon_t\leq 0)\mathbf{1}(\beta_0'x_t\leq 0)|x_t|\mathbf{1}(T^{-1/2}|x_t|> K).$$

The second and fourth term can be made arbitrarily small by choosing K large because $\sup_{1\leq t\leq T}T^{-1/2}|x_t| = O_p(1)$. All terms are monotone in σ and positive. Therefore, the largest possible value when maximized over σ for each term is attained at either $\sigma_1=\max\Sigma$ or $\sigma_2=\inf\Sigma$. Therefore, it suffices to show the $O_p(1)$ property for the first and third term. To show this, note that the absolute expectation of the first and third term is bounded by

$$\sup_{T\geq 1}T^{-1/2}\sum_{t=1}^{T}E\phi(\beta_0'x_t/\sigma_j)(1-\Phi(\beta_0'x_t/\sigma_j))^{-1}(1-\Phi(\beta_0'x_t/\sigma_0))\mathbf{1}(T^{-1/2}|x_t|\leq K)K<\infty,$$

from which result (3) follows; the last result follows from Lemma A.3 from noting that

$$\phi(y/\sigma_j)(1-\Phi(y/\sigma_j))^{-1}(1-\Phi(y/\sigma_0))$$

is integrable over y because

$$\phi(y)(1-\Phi(y))^{-1}\mathbf{1}(y>1)\leq C|y|$$

(see Park and Phillips (2000), p. 1254) and because $|y|(1-\Phi(cy))\mathbf{1}(y>0)$ is integrable for all $c>0$.

To show the rate for $\hat{\sigma}_T^2$, note that from the expression for the score function of Equation (10), it is easily seen that

$$\hat{\sigma}_T^2 = O_p(T^{-1/2})+\sigma^2+O_p(\sup_{\sigma\in\Sigma}T^{-1}\sum_{t=1}^{T}|x_t'\hat{\beta}_T|\phi(x_t'\hat{\beta}_T/\sigma)(1-\Phi(x_t'\hat{\beta}_T/\sigma))^{-1}\mathbf{1}(x_t'\beta_0+\varepsilon_t\leq 0)).$$

Because $\phi(y)(1-\Phi(y))^{-1}$ is increasing in y, and because $\hat{\beta}_T'x_t\leq\beta_0'x_t+|\hat{\beta}_T-\beta_0||x_t|$, and because $T^{-1/2}\sup_{1\leq t\leq T}|x_t| = O_p(1)$, it suffices to find for all $K>0$ an $O_p(T^{-1/2})$ rate for

$$T^{-1}\sum_{t=1}^{T}(x_t'\beta_0+T^{-1/2}K)\phi((x_t'\beta_0+T^{-1/2}K)/\sigma_{min})(1-\Phi((x_t'\beta_0+T^{-1/2}K)/\sigma_{min}))^{-1}\mathbf{1}(x_t'\beta_0+\varepsilon_t\leq 0).$$

Taking absolute expectations, we see that it suffices to find a $O(T^{-1/2})$ rate for

$$ET^{-1}\sum_{t=1}^{T}|x_t'\beta_0+T^{-1/2}K|\phi((x_t'\beta_0+T^{-1/2}K)/\sigma_{min})(1-\Phi((x_t'\beta_0+T^{-1/2}K)/\sigma_{min}))^{-1}(1-\Phi(\beta_0'x_t)).$$

Because $|y|\phi(y)(1-\Phi(y))^{-1} = O(y^2)$, by Lemma A.5, the last expression is asymptotically equivalent to

$$ET^{-1}\sum_{t=1}^{T}|x_t'\beta_0|\phi(x_t'\beta_0/\sigma_{min}).$$

Because $\beta_0'x_t = \alpha_0^1 x_{t1}$, the above expression involves the average of the summation of an integrable function of an integrated process, and is $O(T^{-1/2})$ by Lemma A.3. That concludes the proof. □

Proof of Theorem 1: We start with the expansion (23), or

$$0 = S_T(\hat{\theta}_T) = S_T(\theta_0) + H_T(\theta_0)(\hat{\theta}_T - \theta_0) + [H_T(\tilde{\theta}_T) - H_T(\theta_0)](\hat{\theta}_T - \theta_0).$$

Define $D_T = \text{diag}(TI_m, T^{1/2})$. Then write

$$0 = D_T^{-1}S_T(\theta_0) + [D_T^{-1}H_T(\theta_0)D_T^{-1}]D_T(\hat{\theta}_T - \theta_0) + (D_T^{-1}[H_T(\tilde{\theta}_T) - H_T(\theta_0)]D_T^{-1})D_T(\hat{\theta}_T - \theta_0).$$

We will employ Theorem 10.1 of Wooldridge (1994), which applies to nonlinear models with nonstationary data. Condition (i) and (ii) are satisfied by assumption and condition (iv) is given in Lemma 3 and Lemma 4 in the paper. Therefore we only need to verify Wooldridge's condition (iii), which is

$$\sup_{\{\theta:|C_T(\theta-\theta_0)|\leq 1|\}} |C_T^{-1}[H_T(\theta) - H_T(\theta_0)]C_T^{-1}| = o_p(1) \tag{43}$$

where $C_T = D_T T^{-\delta}$ for some $\delta > 0$. For (43) to hold, it is sufficient that

$$T^{-2+2\delta}|H_{\beta\beta'}(\theta) - H_{\beta\beta}(\theta_0)| \xrightarrow{p} 0,$$

$$T^{-3/2+2\delta}|H_{\sigma^2\beta'}(\theta) - H_{\sigma^2\beta'}(\theta_0)| \xrightarrow{p} 0,$$

and

$$T^{-1+2\delta}|H_{\sigma^2\sigma^2}(\theta) - H_{\sigma^2\sigma^2}(\theta_0)| \xrightarrow{p} 0$$

uniformly for all $\theta = (\beta', \sigma^2)'$ satisfying

$$|\beta - \beta_0| \leq T^{-1+\delta} \quad \text{and} \quad |\sigma^2 - \sigma_0^2| \leq T^{-1/2+\delta} \tag{44}$$

for some $\delta > 0$. By the Taylor series expansion of the Hessian at $\theta = \theta_0$, for some mean value $\tilde{\theta} = (\tilde{\beta}, \tilde{\sigma}^2)$,

$$H_{\beta\beta'}(\theta) - H_{\beta\beta'}(\theta_0)$$

$$= \sum_{t=1}^{T} (\partial(-g_1(x_t'\beta/\sigma)\sigma^{-2})/\partial\beta|_{\theta=\tilde{\theta}}(\beta-\beta_0) + \partial(-g_1(x_t'\beta/\sigma)\sigma^{-2})/\partial\sigma^2|_{\theta=\tilde{\theta}}(\sigma^2-\sigma_0^2))$$

$$\cdot\mathbf{1}(x_t'\beta_0+\varepsilon_t \le 0)x_t x_t' + \tilde{\sigma}^{-4}(\sigma^2-\sigma_0^2)\sum_{t=1}^{T}\mathbf{1}(x_t'\beta_0+\varepsilon_t>0)x_t x_t',$$

$$H_{\sigma^2\beta'}(\theta) - H_{\sigma^2\beta'}(\theta_0)$$

$$= \sum_{t=1}^{T} (\partial(-g_2(x_t'\beta/\sigma)\sigma^{-3})/\partial\beta|_{\theta=\tilde{\theta}}(\beta-\beta_0) + \partial(-g_2(x_t'\beta/\sigma)\sigma^{-3})/\partial\sigma^2|_{\theta=\tilde{\theta}}(\sigma^2-\sigma_0^2))$$

$$\cdot\mathbf{1}(x_t'\beta_0+\varepsilon_t \le 0)x_t + 2\tilde{\sigma}^{-6}(\sigma^2-\sigma_0^2)\sum_{t=1}^{T}\mathbf{1}(x_t'\beta_0+\varepsilon_t>0)(\varepsilon_t+(\tilde{\beta}-\beta_0)'x_t)x_t$$

$$+2\tilde{\sigma}^{-4}\sum_{t=1}^{T}((\beta-\beta_0)'x_t)\mathbf{1}(x_t'\beta_0+\varepsilon_t>0)x_t,$$

$$H_{\sigma^2\sigma^2}(\theta) - H_{\sigma^2\sigma^2}(\theta_0)$$

$$= \sum_{t=1}^{T} (\partial(g_3(x_t'\beta/\sigma)\sigma^{-4})/\partial\beta|_{\theta=\tilde{\theta}}(\beta-\beta_0) + \partial(g_3(x_t'\beta/\sigma)\sigma^{-4})/\partial\sigma^2|_{\theta=\tilde{\theta}}(\sigma^2-\sigma_0^2))\mathbf{1}(x_t'\beta_0+\varepsilon_t \le 0)$$

$$-\tilde{\sigma}^{-6}(\sigma^2-\sigma_0^2)\sum_{t=1}^{T}\mathbf{1}(x_t'\beta_0+\varepsilon_t>0) + 3\tilde{\sigma}^{-8}(\sigma^2-\sigma_0^2)\sum_{t=1}^{T}(\varepsilon_t-(\tilde{\beta}-\beta_0)'x_t)^2\mathbf{1}(x_t'\beta_0+\varepsilon_t>0)$$

$$-4\tilde{\sigma}^{-6}\sum_{t=1}^{T}(\varepsilon_t-(\tilde{\beta}-\beta_0)'x_t)((\beta-\beta_0)'x_t)\mathbf{1}(x_t'\beta_0+\varepsilon_t>0).$$

We first consider the linear components of each Hessian function. The linear term in $T^{-2+2\delta}|H_{\beta\beta'}(\theta) - H_{\beta\beta}(\theta_0)|$ can be bounded by

$$\sup T^{-2+2\delta}\tilde{\sigma}^{-4}|\sigma^2-\sigma_0^2|\sum_{t=1}^{T}\mathbf{1}(x_t'\beta_0+\varepsilon_t>0)|x_t x_t'| \le \sigma_{min}^{-4}T^{-5/2+3\delta}\sum_{t=1}^{T}|x_t x_t'| = O_p(T^{-1/2+3\delta}).$$

The linear terms in $T^{-3/2+2\delta}|H_{\sigma^2\beta'}(\theta) - H_{\sigma^2\beta'}(\theta_0)|$ can be bounded by

$$\sup 2\tilde{\sigma}^{-6}|\sigma^2-\sigma_0^2||T^{-3/2+2\delta}\sum_{t=1}^{T}\mathbf{1}(x_t'\beta_0+\varepsilon_t>0)x_t\varepsilon_t + x_t(\tilde{\beta}-\beta_0)'x_t|$$

$$\le 2\sigma_{min}^{-6}T^{-2+3\delta}\sum_{t=1}^{T}|x_t\varepsilon_t| + 2\sigma_{min}^{-6}T^{-3+4\delta}\sum_{t=1}^{T}|x_t x_t'|$$

$$= O_p(T^{-1/2+3\delta}) + O_p(T^{-1+4\delta}),$$

and

$$\sup 2\tilde{\sigma}^{-4}|T^{-2+2\delta}\sum_{t=1}^{T}[(\beta-\beta_0)'x_t]\mathbf{1}(x_t'\beta_0+\varepsilon_t>0)x_t|$$

$$\leq 2\sigma_{min}^{-4}T^{-3+3\delta}\sum_{t=1}^{T}|x_tx_t'| = O_p(T^{-1+3\delta}).$$

The linear terms in $T^{-1+2\delta}|H_{\sigma^2\sigma^2}(\theta) - H_{\sigma^2\sigma^2}(\theta_0)|$ can be bounded by

$$\sigma_{min}^{-2}T^{-1/2+3\delta} + 3\sigma_{min}^{-8}T^{-3/2+2\delta}\sum_{t=1}^{T}\varepsilon_t^2 + 3\sigma_{min}^{-8}T^{-5/2+4\delta}\sum_{t=1}^{T}|\varepsilon_t x_t|$$

$$+3\sigma_{min}^{-8}T^{-7/2+5\delta}\sum_{t=1}^{T}|x_tx_t'| + 4\sigma_{min}^{-6}T^{-2+3\delta}\sum_{t=1}^{T}|\varepsilon_t x_t| + 4\sigma_{min}^{-6}T^{-3+4\delta}\sum_{t=1}^{T}|x_tx_t'|$$

$$= O_p(T^{-1/2+3\delta}) + O_p(T^{-1+4\delta}) + O_p(T^{-3/2+5\delta}).$$

Therefore, if we let $0 < \delta < 1/6$, all linear components goes to zero uniformly. For the nonlinear component, note that a typical term in the Hessian functions takes the form

$$\sum_{t=1}^{T}\sigma^{-l}\phi^k(x_t'\beta/\sigma)(1-\Phi(x_t'\beta/\sigma))^{-k}(x_t'\beta/\sigma)^p x_t^{\kappa},$$

where l, k, p, κ are some nonnegative integers. Note that

$$\begin{aligned}
& (\partial/\partial\beta)(\phi(x_t'\beta/\sigma)/(1-\Phi(x_t'\beta/\sigma))) \\
= \; & -\phi(x_t'\beta/\sigma)(1-\Phi(x_t'\beta/\sigma))^{-1}(x_t'\beta/\sigma)(x_t/\sigma) + \phi^2(x_t'\beta/\sigma)(1-\Phi(x_t'\beta/\sigma))^{-2}(x_t/\sigma) \\
& (\partial/\partial\sigma^2)(\phi(x_t'\beta/\sigma)/(1-\Phi(x_t'\beta/\sigma))) \\
= \; & (1/2\sigma^2)[\phi(x_t'\beta/\sigma)(1-\Phi(x_t'\beta/\sigma))^{-1}(x_t'\beta/\sigma)^2 + \phi^2(x_t'\beta/\sigma)(1-\Phi(x_t'\beta/\sigma))^{-2}(x_t'\beta/\sigma)]
\end{aligned}$$

Therefore, if we let $\tilde{\phi}_t = \phi(x_t'\tilde{\beta}/\tilde{\sigma})$, $\tilde{\Phi}_t = \Phi(x_t'\tilde{\beta}/\tilde{\sigma})$, and for some nonnegative integers l, k, p, j, n, q, κ, a typical term of the nonlinear component evaluated at $\tilde{\theta}$ can be written as[2]

$$\sup|T^{-1+\kappa/2+2\delta}\sum_{t=1}^{T}(\tilde{\sigma}^{-l}\tilde{\phi}_t^k(1-\tilde{\Phi}_t)^{-k}(x_t'\tilde{\beta}/\tilde{\sigma})^p x_t'(\beta-\beta_0) + \tilde{\sigma}^{-j}\tilde{\phi}_t^n(1-\tilde{\Phi}_t)^{-n}(x_t'\tilde{\beta}/\tilde{\sigma})^q(\sigma^2-\sigma_0^2))$$

[2]For instance,

$$\begin{aligned}
\partial g_1(\tilde{\theta})/\partial\beta &= -\tilde{\phi}_t(1-\tilde{\Phi}_t)^{-1}(x_t/\tilde{\sigma}) + (2\tilde{\phi}_t(1-\tilde{\Phi}_t)^{-1} - x_t'\tilde{\beta}/\tilde{\sigma})(\partial/\partial\tilde{\beta})(\tilde{\phi}_t/(1-\tilde{\Phi}_t)) \\
\partial g_1(\tilde{\theta})/\partial\sigma^2 &= -\tilde{\phi}_t(1-\tilde{\Phi}_t)^{-1}(x_t/\tilde{\sigma})(1/2\tilde{\sigma}^2) + (2\tilde{\phi}_t(1-\tilde{\Phi})^{-1} - x_t'\tilde{\beta}/\tilde{\sigma})(\partial/\partial\tilde{\sigma}^2)(\tilde{\phi}_t/(1-\tilde{\Phi}_t))
\end{aligned}$$

$$
\begin{aligned}
&\cdot \mathbf{1}(x_t'\beta_0 + \varepsilon_t \le 0) x_t^\kappa | \\
&\le \sup T^{-3/2+\kappa/2+2\delta} \sum_{t=1}^T (\tilde\sigma^{-l} \tilde\phi_t^k (1-\tilde\Phi_t)^{-k} |x_t'\tilde\beta/\tilde\sigma|^p |x_t|/\sqrt{T} + \tilde\sigma^{-j}\tilde\phi_t^n (1-\tilde\Phi_t)^{-n} |x_t'\tilde\beta/\tilde\sigma|^q) \\
&\quad \cdot \mathbf{1}(x_t'\beta_0 + \varepsilon_t \le 0)|x_t|^\kappa \\
&= \sup T^{-3/2+\kappa/2+2\delta} \sum_{t=1}^T (\tilde\sigma^{-l} \tilde\phi_t^k (1-\tilde\Phi_t)^{-k} (x_t'\tilde\beta/\tilde\sigma)^p |x_t|/\sqrt{T} + \tilde\sigma^{-j}\tilde\phi_t^n (1-\tilde\Phi_t)^{-n} (x_t'\tilde\beta/\tilde\sigma)^q) \\
&\quad \cdot \mathbf{1}(x_t'\tilde\beta > 0)\mathbf{1}(x_t'\beta_0 + \varepsilon_t \le 0)|x_t|^\kappa \\
&+ \sup T^{-3/2+\kappa/2+2\delta} \sum_{t=1}^T (\tilde\sigma^{-l} \tilde\phi_t^k (1-\tilde\Phi_t)^{-k} |x_t'\tilde\beta/\tilde\sigma|^p |x_t|/\sqrt{T} + \tilde\sigma^{-j}\tilde\phi_t^n (1-\tilde\Phi_t)^{-n} |x_t'\tilde\beta/\tilde\sigma|^q) \\
&\quad \cdot \mathbf{1}(x_t'\tilde\beta \le 0)\mathbf{1}(x_t'\beta_0 + \varepsilon_t \le 0)|x_t|^\kappa \\
&\le \sup T^{-3/2+\kappa/2+2\delta} \sum_{t=1}^T (\sigma_{min}^{-l} h_1(x_t'\tilde\beta/\tilde\sigma)|x_t|/\sqrt{T} + \sigma_{min}^{-j} h_2(x_t'\tilde\beta/\tilde\sigma))\mathbf{1}(x_t'\tilde\beta > 0) \\
&\quad \cdot \mathbf{1}(x_t'\beta_0 + \varepsilon_t \le 0)|x_t|^\kappa \qquad (45) \\
&+ \sup T^{-3/2+\kappa/2+2\delta} \sum_{t=1}^T (2^k \sigma_{min}^{-l} \tilde\phi_t^k |x_t'\tilde\beta/\tilde\sigma|^p |x_t|/\sqrt{T} + 2^n \sigma_{min}^{-j} \tilde\phi_t^n |x_t'\tilde\beta/\tilde\sigma|^q) \\
&\quad \cdot \mathbf{1}(x_t'\tilde\beta \le 0)\mathbf{1}(x_t'\beta_0 + \varepsilon_t \le 0)|x_t|^\kappa \qquad (46)
\end{aligned}
$$

where

$$h_1(x) = \phi(x)^k (1-\Phi(x))^{-k} x^p \quad \text{and} \quad h_2(x) = \phi(x)^n (1-\Phi(x))^{-n} x^q.$$

It is clear that since ϕ is an integrable function, the term in (46) is of order $O_p(T^{-1+2\delta})$; see Lemma A2 in Park and Phillips (2000). For the term in (45), choose K large enough so that the probability that $T^{-1/2}|x_t| > K$ is negligible for all $t \le T$. Then term (45) is bounded by

$$T^{-3/2+2\delta} \sum_{t=1}^T (\sigma_{min}^{-l} \bar h_1(x_t'\beta_0/\sigma_0) K + \sigma_{min}^{-j} \bar h_2(x_t'\beta_0/\sigma_0)) \mathbf{1}(x_t'\beta_0 + \varepsilon_t \le 0) K^\kappa, \qquad (47)$$

where $\bar h_1(x)$ and $\bar h_2(x)$ are defined as in Lemma A.8. The absolute expectation of (47) can be bounded by $T^{-1/2+2\delta} G_T$, where

$$
\begin{aligned}
G_T &= T^{-1} \sum_{t=1}^T (\sigma_{min}^{-l} E|\bar h_1(x_t'\beta_0/\sigma_0)|K + \sigma_{min}^{-j} E|\bar h_2(x_t'\beta_0/\sigma_0)|)(1-\Phi(x_t'\beta_0/\sigma_0))K^\kappa \\
&= o_p(1)
\end{aligned}
$$

following Lemma A.3 and Lemma A.8. Therefore, the term (45) is of order $T^{-1/2+2\delta}$. In summary, when $0 < \delta < 1/6$, condition (43) is satisfied. Then by Theorem 10.1 of Wooldridge (1994), we have the limit distribution of the ML estimator. □

Proof of Theorem 3: Without loss of generality, we assume that $\beta_0 > 0$ in the proof. This result follows from Lemma A.9 and Lemma A.10 if we can show that $H_T(\cdot)$ is stochastically equicontinuous on $[0, 1]$. To show this, we only need to show the stochastic equicontinuity of

$$T^{-3/2} \sum_{t=1}^{[rT]} x_t \mathbf{1}(x_t > 0)$$

and

$$T^{-1/2} \sum_{t=1}^{[rT]} \varepsilon_t \mathbf{1}(x_t > 0)).$$

For the first term, we have

$$|T^{-3/2} \sum_{t=1}^{[rT]} x_t \mathbf{1}(x_t > 0) - T^{-3/2} \sum_{t=1}^{[r'T]} x_t \mathbf{1}(x_t > 0)|$$

$$\leq (T^{-1/2} \max_{1 \leq t \leq T} |x_t|)|r - r'|,$$

from which stochastic equicontinuity of this term follows. In addition, by the Burkholder inequality and the norm inequality,

$$E|T^{-1/2} \sum_{t=[r'T]+1}^{[rT]} \varepsilon_t \mathbf{1}(x_t > 0))|^4$$

$$\leq CE|T^{-1} \sum_{t=[r'T]+1}^{[rT]} \varepsilon_t^2 \mathbf{1}(x_t > 0)|^2 \leq C'|r - r'|^2 E\varepsilon_t^4,$$

and from the reasoning of the proof of Lemma A.9, it follows that $T^{-1/2} \sum_{t=1}^{[rT]} \varepsilon_t \mathbf{1}(x_t > 0)$ is stochastically equicontinuous. □