Department of Economics, Hitotsubashi University

Discussion Paper Series #2014-20

Empirical Likelihood Confidence Intervals for Nonparametric Nonlinear Nonstationary Regression Models

Ryota Yabe

December 2014

Empirical Likelihood Confidence Intervals for Nonparametric Nonlinear Nonstationary Regression Models

Ryota Yabe*

Abstract

By using the empirical likelihood (EL), we consider the construction of pointwise confidence intervals (CIs) for nonparametric nonlinear nonstationary regression models with nonlinear nonstationary heterogeneous errors. It is well known that the EL-based CI has attractive properties such as data dependency and automatic studentization in cross-sectional and weak-dependence models. We extend EL theory to the nonparametric nonlinear nonstationary regression model and show that the log-EL ratio converges to a chi-squared random variable with one degree of freedom. This means that Wilks' theorem holds even if the covariate follows a nonstationary process. We also conduct empirical analysis of Japan's inverse money demand to demonstrate the data-dependency property of the EL-based CI.

1 Introduction

The aim of this paper is to use the empirical likelihood (EL) to construct pointwise confidence intervals (CIs) of the unknown regression function for a nonparametric nonstationary regression model. The nonlinear nonstationary regression model can be interpreted as an extension of the linear cointegrating model, which has been exhaustively studied for the last three decades and applied in many empirical studies. However, linearity of the cointegrating model has recently been considered restrictive in empirical studies such as discrete choice models and—in macroeconomics—money demand functions. (See, for example, Park and Phillips (2000) and

*E-mail address:k131014a@r.hit-u.ac.jp

Hu and Phillips (2004a) for discrete choice models and Hu and Phillips (2004b) and Bae and de Jong (2007) for money demand functions.) To ease this linearity restriction, attention has turned to the (parametric and nonparametric) nonlinear nonstationary regression model.

There are many studies of the parametric nonlinear nonstationary model. For example, Park and Phillips (1999) and Park and Phillips (2001) provided a quite general asymptotic theory and proposed a nonlinear least-squares estimation method. Parametric modeling is useful and works well given a priori information on the functional form. However, empirical studies are hampered by the little information provided by economic theory about functional form. To avoid misspecification, recent attention has turned to the nonstationary nonparametric regression model, which requires no a priori information on functional form. Karlsen et al. (2007) have studied the asymptotic properties of the local-constant estimator by using the recurrent Markov chain and Wang and Phillips (2009a) have studied local-time asymptotic techniques. See Tjøstheim (2012) and Park (2014) for a survey of recent studies of parametric and nonparametric nonlinear nonstationary regression models.

Wang and Phillips (2009b) have also constructed pointwise CIs of the covariate function based on the residual from a kernel regression for the structural nonlinear model. Because the CI based on the residual is artificially symmetric around the estimate, the CI is inflexible and problematic. Owen (1988) points out the well-known problem that the residual-based CI does not work well if the distribution of the error process is asymmetric. Another problem is that any a priori information on the data, such they are positive or bounded, may be meaningless. For example, although much macroeconomic data take nonnegative values, the residual-based CI can include negative values if the dependent variable is close to zero, as is the case with interest and inflation rates.

To overcome these symmetry problems, we propose the EL-based CI based on the EL ratio for the nonparametric nonstationary regression model. Xu (2009) has adopted this approach for nonparametric diffusion processes. The nonparametric regression model with independent and identically distributed regressors has been studied by Chen and Qin (2000). Xu (2009) applied EL-based CI techniques to the nonparametric diffusion model and showed that it works better than the residual-based CI. The EL-based CI has two attractive features. One is that the shape of the CI is data dependent, unlike the residual-based CI. We demonstrate the data-dependency property of the EL-based CI by conducting an empirical study of Japan's inverse money demand function. The second one is the EL-based CI's automatic studentization property. In cross-sectional data, the log-EL ratio is asymptotically pivotal, which means that one need not estimate the variance of the estimator. For time-series parametric models, Kitamura (1997) showed that the time direction correlation of the asymptotic distribution of the log EL must be estimated. Therefore, the asymptotically pivotal property of the EL is redundant in time-series parametric models. However, having recently considered pointwise CIs for the kernel density of weakly dependent processes, citetXiong2012 have shown that the asymptotic distribution of the log EL converges to the chi-squared distribution. This asymptotic result is known as Wilks' theorem in the EL literature. In this paper, we show that the log EL of our model exhibits the asymptotically pivotal property by proving that Wilks' theorem holds.

We extend the literature by incorporating a nonlinear nonstationary heterogeneous (NNH) disturbance term, as originally proposed by Park (2002). Such an error term is often used in nonparametric regression and allows the volatility of the error to depend on nonstationary regressors. Wang and Wang (2013) have proved uniform consistency of the local estimator with NNH errors. Park (2002) has pointed out that the NNH error is a powerful alternative to the ARCH error. We show that the log-EL statistic is pivotal even if the error follows an NNH process, which means that a pointwise CI of the regression function can be constructed without estimating nuisance parameters. This property enables us to extend the nonparametric nonstationary regression model considered by Wang and Phillips (2009a) to the NNH error model proposed by Wang and Wang (2013) without making specific modifications.

Following Chen et al. (2003), Wang and Wang (2013) have shown that, under exogeneity with stationary regressors, Wilks' theorem holds for the log EL at the true parameter values of the nonparametric regression model. Combining this result with our main result reveals that the EL-based CI is a unified way of constructing CIs between stationary and nonstationary regressors. That is, one can construct a pointwise CI for the covariate function without testing whether the regressor is stationary or nonstationary.

The rest of this paper is organized as follows. In Section 2, we describe the model and its assumptions.

In Section 3, we present our main result in the form of the asymptotic distribution of the EL and show that Wilks' theorem holds. In Section 4, we use simulation results to compare the performance of our proposed EL-based CIs with that of residual-based CIs. In Section 5, we conduct an empirical study of Japan's inverse money demand function to demonstrate the flexibility of EL-based CIs. Section 6 concludes the paper.

2 Model and Assumptions

We consider the nonlinear nonstationary regression model with an NNH error defined as follows:

$$y_t = f(x_t) + \sigma(x_t)u_t, \quad t = 1, \ldots, n, \tag{2.1}$$

where $f(\cdot) : \mathbb{R} \to \mathbb{R}$ is an unknown function to be estimated, $\{x_t\}$ is a scalar nonstationary process and $\{u_t\}$ is a weakly dependent process. $\sigma(\cdot) : \mathbb{R} \to \mathbb{R}^{++}$ is the same NNH function used by Park (2002) and Wang and Wang (2013). The heterogeneity generating function captures heterogeneity in the error, which is explained by the nonstationary regressor. Rigorous definitions and assumptions are given below.

Assumption 1. *$x_t = \rho x_{t-1} + \eta_t$, where $x_0 = 0$, $\rho = 1 - \kappa/n$ with κ being a positive constant, and $\eta_t = \sum_{k=0}^{\infty} \phi_k \epsilon_{t-k}$ with $\phi \equiv \sum_{k=0}^{\infty} \phi_k \neq 0$ and $\sum_{k=0}^{\infty} |\phi| < \infty$. $\{\epsilon\}$ is a sequence of zero mean independent and identically distributed random variables with $\epsilon_1^2 = 1$, and with a characteristic function $\phi(t)$ satisfying $\int_{-\infty}^{\infty} |\phi(t)| < \infty$.*

Assumption 2. *The disturbance process $u_t = u(\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-m_0+1}, \lambda_t, \lambda_{t-1}, \ldots, \lambda_{t-m_0+1})$ satisfies $E[u_t] = 0$, $E[u_t^2] = 1$ and $E[u_t^4] < \infty$ for $t \geq m_0$, where $u(x_1, \ldots, x_{m_0}, y_1, \ldots, y_{m_0})$ is a real measurable function on $\mathbf{R}^{2m_0}$. The initial values of $\{u_t\}$ also satisfy $u_t = 0$ for $1 \leq t \leq m_0 - 1$.*

Assumption 3. *The NNH function $\sigma(\cdot) : \mathbb{R} \to \mathbb{R}^{++}$ satisfies the condition that there is a constant C such that*

$$|\sigma(x) - \sigma(y)| < C|x - y|^{\delta}, \ \textit{for some} \ 0 < \delta \leq 1.$$

Assumptions 1–3 apply to the data generating process given by equation (2.1). The first part of Assumption 1 ensures that the regressor process $\{x_t\}$ belongs to the local-to-unity class. The remainder of Assumption

1 ensures that the normalized regressor $\{x_t/d_t\}$ has a uniformly bounded density for t, where $d_t^2 = \mathrm{E}[x_t^2]$. (For a detailed derivation, see Wang and Phillips (2009a)'s proof of Corollary 2.2.) To incorporate structural economic models, Assumption 2 introduces endogeneity by allowing correlation between the regressor and error processes. Standardization of the error variance in the last condition is used for identification. However, because inference about the error process is beyond our scope, this restriction is not necessarily required. Assumption 3 is the same one made to ensure smoothness of the NNH function by Park (2002) and Wang and Wang (2013).

We consider two different ELs. One uses the first-order condition (FOC) of the local-constant estimator as an estimating equation. The other is the EL based on the local linear estimating equation. These estimating equations are based on the work of Hall and Owen (1993), Chen and Qin (2000) and Xiong and Lin (2012).

Let a be a candidate for the covariate function ($f(x)$) at given x. The profile EL function $L_c(a)$ based on the local-constant estimating equation is defined as

$$L_c(a) = \left\{ \max_{p_t} \prod_{t=1}^{n} p_t, \sum_{t=1}^{n} p_t = 1, \sum_{t=1}^{n} p_t g_{t,c}(a) = 0 \right\}, \tag{2.2}$$

where $g_{t,c}(a) = (y_t - a)K[(x_t - x)/h]$ is the FOC of the local-constant estimator. The local linear EL $L_l(a)$ is similarly defined by replacing $g_{t,c}$ in (2.2) with the concentrated FOC of the local linear estimator $g_{t,l}(a) = (y_t - a)[1 - s_{n,1}/s_{n,2}(x_t - x)/h]K[(x_t - x)/h]$, where $s_{n,i} = \sum_{t=1}^{n}[(x_t - x)/h]^i K[(x_t - x)/h]$ for $i = 1, 2$.

Solving the Lagrangian maximization problem defined in (2.2) yields the local-constant log EL defined as

$$l_c(a) = 2 \sum_{t=1}^{n} \log(1 + \lambda g_{t,c}(a)), \tag{2.3}$$

where the Lagrangian multiplier $\lambda_c(a)$ is the solution of the following equation:

$$\sum_{t=1}^{n} \frac{g_{t,c}(a)}{1 + \lambda_c(a) g_{t,c}(a)} = 0. \tag{2.4}$$

The local linear log EL ($l_l(a)$) can also be defined by replacing $g_{t,c}(a)$ in (2.3) with $g_{t,l}(a)$.

The next set of assumptions concerns the kernel and covariate functions for the local-constant EL and the linear EL.

Assumption 4. *The compactly supported kernel function $K(\cdot)$ satisfies*

$$\int_{-\infty}^{\infty} K(s)dy = 1 \text{ and } \int_{-\infty}^{\infty} K^2(y)dy < \infty.$$

Assumption 5. *The covariate function $f(\cdot)$ satisfies the condition that there is a function $f_1(x, y)$ such that for given x, sufficiently small h and for some $0 < \gamma < 1$,*

$$|f(x + hy) - f(x)| < h^{\gamma} f_1(x, y) \text{ for all } y \in \mathbb{R} \quad \text{and} \quad \int_{-\infty}^{\infty} f_1(x, y)dy < \infty.$$

Assumption 6. *The kernel function $K(\cdot)$ with compact support satisfies*

$$\int_{-\infty}^{\infty} K(y)dy = 1, \int_{-\infty}^{\infty} yK(y)dy = 0, \quad \text{and} \quad \int_{-\infty}^{\infty} y^2K(y)dy \neq 0.$$

Assumption 7. *For given x, the covariate function $f(\cdot)$ is continuously twice differentiable in a neighborhood of x.*

Assumptions 4 and 5 for the local-constant EL ($L_c(a)$) are the same as those used by Wang and Phillips (2009a) and Wang and Phillips (2009b). Assumptions 6 and 7 for the local linear EL ($L_L(a)$) are slightly weaker than those used by Wang and Phillips (2011). The contrast with Wang and Phillips (2011) is that we use a second-order kernel in Assumption 6 and assume a twice differentiable covariate function in Assumption 7; to analyze higher-order asymptotic properties, Wang and Phillips (2011) assumes a higher-order kernel and higher-order differentiability of the covariate function. Because such higher-order analysis for bias correction is beyond our scope, we make weaker assumptions.

3 Results

In this section, we derive the asymptotic distribution of the log-EL ratios for the true value of the covariate function at a given x to construct the EL-based CIs. If there is no endogeneity, under certain regularity conditions, combining the results of Chen et al. (2003) with routine calculation shows that Wilks' theorem holds for an α-mixing process. Therefore, showing that Wilks' theorem holds in the nonstationary nonparametric regression model means that, for the time-series nonparametric regression model, EL-based construction of the CI represents a unified approach.

Theorem 1. *Under Assumptions 1–5, for any $h \to 0$, $nh^2 \to \infty$ and $nh^{2+4\gamma} \to 0$, at given x, the asymptotic distributions of the Lagrange multiplier $\lambda(f(x))$ and the local-constant log EL for $a = f(x)$ are given, respectively, by*

$$(nh^2)^{1/4}\lambda_c(f(x)) \Rightarrow NL^{-1/2}(1,0)\left(\int_{-\infty}^{\infty} K^2(s)ds\right)^{-1/2}\left(\frac{E[u_{m0}^2]}{|\phi|}\right)^{-1/2}, \tag{3.1}$$

$$l_c(f(x)) \Rightarrow \chi_1^2, \tag{3.2}$$

where N is a standard normal random variable, $L(1,0)$ is a local time function and χ_1^2 is a chi-squared random variable with one degree of freedom.

Theorem 1 shows that the asymptotic log EL is a chi-squared random variable with one degree of freedom. This means that Wilks' theorem holds even if the regressor is a nonstationary process. By using Theorem 1, we can construct the pointwise $100(1-\alpha)\%$ level EL-based CI (I_c) of the covariate function at a given x such that

$$I_c = \left\{a | l_c(a) \leq \chi_{1,1-\alpha}^2\right\}, \tag{3.3}$$

where $\chi_{1,1-\alpha}^2$ is a $100(1-\alpha)\%$ quantile of a chi-squared random variable with one degree of freedom.

As noted above, Wilks' theorem also holds for an α-mixing process under certain regularity conditions; see Chen et al. (2003) for details of the assumptions required. However, each bandwidth convergence order is different. For an α-mixing process, the bandwidth must satisfy $h = O(n^{-1/3})$ under the assumptions of

Theorem 1 and those of Chen et al. (2003). Using $h = c/n^{1/3}$ for any positive constant as the bandwidth satisfies the conditions for both weak and strong data dependence. This confirms that the local-constant EL-based CI is unified.

Theorem 2. *Under Assumptions 1–3, 6 and 7, for any $h \to 0$, $nh^2 \to \infty$ and $nh^{10} \to 0$, at given x, the asymptotic distributions of the Lagrange multiplier $\lambda(f(x))$ and the local-linear log EL, $l_l(f(x))$, are the same as those given in Theorem 1.*

The Lagrange multipliers and log ELs of the local-constant estimator and the linear estimator have the same asymptotic distributions. This parallels the equivalence of the asymptotic distributions of the local-constant and linear kernel regression estimators considered by Wang and Phillips (2009a) and Wang and Phillips (2011), respectively.

Based on the local linear log EL, we can use (3.3) to construct the CI (I_l). Because, as noted above, Wilks' theorem holds for a weak-dependence process, the EL-based CIs are valid for weak-dependence and nonstationary processes. However, EL-based CIs for weak and nonstationary processes have bandwidths of different orders. The bandwidth for a weak-dependence process must satisfy $O(n^{-1/5})$ under the assumptions of Theorem 2 and those made by Chen et al. (2003). Therefore, we can construct a unified local-constant EL-based CI for stationary and nonstationary time-series models by using $h = c/n^{1/5}$ for any positive constant c.

In this section, we have shown that the log ELs of the local-constant and linear estimators satisfy Wilks' theorem and represent unified approaches for stationary and nonstationary regressors. Therefore, one can draw inferences without testing whether the regressors are stationary or nonstationary.

4 Simulation

We conduct a simulation analysis to compare the proposed El-based method to the residual-based one proposed by Wang and Phillips (2009b). The simulation design is as follows:

$$y_t = 0.3x_t + 0.5x_t^2 \sin(x_t) + 0.5(1 + x_t)u_t, \tag{4.1}$$

where

$$x_t = x_{t-1} + \epsilon_t \quad \text{and} \quad u_t = \frac{v_t + 2\epsilon_t}{\sqrt{1 + 2^2\sigma_\epsilon^2}}, \tag{4.2}$$

$\{v_t\}$ and $\{\epsilon_t\}$ are sequences of independent normal random variables with variances of 1 and σ_ϵ^2 (specified later), respectively. The number of replications is 10,000 with sample sizes of $n = 500$ and $1,000$. The bandwidths of the local-constant and linear-based CIs are set to $h = 1/n^{1/3}$ and $1/n^{1/5}$, respectively. The disturbance term $\{u_t\}$ in equation (4.2) captures endogeneity and has the same structure as that used by Wang and Phillips (2009b). We calculate the empirical coverage probabilities of the two-sided 95%-level residual-based and EL-based CIs by using the local-constant estimator and the linear estimator and their corresponding moment conditions, at $x = 0, 0.01, \ldots, 1$. When the number of local observations is less than two, we suppose that the CI in the experiment does not include the true value. This is because, in such an experiment, the CI cannot be constructed.

We consider two cases for the variance of the regression error in equation (4.2): $\sigma_\epsilon^2 = 0.01$ and 0.1. Because the regressor is nonstationary, there are much fewer observations included in the support of the kernel function than in the stationary case. For details of how nonstationary regressors affect nonparametric estimation problems, see the simulation study of Honda (2013).

Figures 1 and 2 show the empirical coverage probabilities for the variance cases $\sigma_\epsilon^2 = 0.01$ and 0.1, respectively. In both cases, all CIs overreject the true value of $f(x)$, particularly when x is large. This is because of the small number observations at each x caused by the nonstationarity of the regressor. In all cases, the local linear EL-based CI (LLEL) is the closest to the 95% nominal level. The local-constant residual-based (LC) and EL-based CIs (LCEL) are too strict because of their small number of observations. The bandwidth of the local linear CIs is wider than that of the local-constant CIs, because of the larger number of observations for the former. Although the LLEL only seems to perform better because of its larger bandwidth, the result for the local linear residual-based CI (LL) in figure 1 shows that the result for the LLEL is attributable not only to the large bandwidth, but also to the properties of the EL.

From our simulation results, we can conclude that the LLEL is the best in terms of coverage probability. However, because it is so strict, this CI's finite-sample performance must be improved by, for example, correcting the higher-order bias. For details of Bartlett correction of the EL in the cross-sectional nonparametric model, see DiCiccio et al. (1991) and Chen and Qin (2000).

5 Empirical Study

In this section, we conduct an empirical study of the inverse money demand function to illustration our procedure. The money demand function is a typical empirical application of the nonlinear nonstationary regression model. It describes the relationship between the interest rate and money demand, and provides useful insights for monetary policy analysis. Empirical studies using the nonlinear nonstationary model include those of Choi and Saikkonen (2010), who applied their proposed specification test to the money demand function and Bae et al. (2006) and Bae and de Jong (2007), who estimated money demand by using the parametric nonlinear model. Rather than study the money demand function itself, we study its inverse. This is for two reasons. One is to illustrate the flexibility of the EL-based CI when the data range of the dependent variable is restricted, which allows us to demonstrate the data-dependence property of the EL-based CI. In this case, the interest rate takes nonnegative values by definition. If the interest rate is close to zero, the residual-based CI can contain negative values because of the symmetry of the conventional CI. However, because the EL-based CI is data dependent, it is not expected to contain negative values.

The other reason is that the interest rate—the regressor of the money demand function—does not seem to have a unit root. Interest rate series are often modeled using an ARMA(1,1) model with a large negative root close to -1 in the MA(1) term. The asymptotic properties of this process are quite different to those of typical unit-root processes. For details, see, for example, Pantula (1991) and Nabeya and Perron (1994).

Ng and Perron (2001) has proposed a unit-root test for this process. Because the interest rate cannot be considered to satisfy Assumption 1, we analyze the inverse money demand function, which uses real money demand as the regressor.

Following Bae and de Jong (2007), we analyze the inverse money demand function given by

$$i_t = f(m_t) + u_t, \tag{5.1}$$

where $m_t = M_t/(P_t Y_t)$, M_t is the money supply, P_t is the gross domestic product (GDP) deflator and Y_t is GDP, which are all in natural logarithms and i_t is the call rate.

We use 91 quarterly observations from 1985Q3 to 2008Q1. The data on i_t are quarterly averages of monthly averaged uncollateralized overnight call rates taken from the Bank of Japan's website (data section). M_t is M1, generated by taking quarterly averages of the seasonally adjusted monthly averaged M1. Seasonally adjusted monthly averaged data on M1, the GDP deflator and real GDP are from the Nikkei database. Bandwidths of $h = 1/(3.5n^{1/3}) \fallingdotseq 0.074$ and $1/(4n^{1/6}) \fallingdotseq 0.116$ are chosen for the local-constant CI and the local linear CI, respectively.

The results are shown in Figures 3, 4, 5 and 6. According to Figures 3 and 5, the EL-based CIs (LCEL and LLEL) are narrower than the corresponding residual-based CIs (LC and LL). The residual-based CI from the local-constant estimator in Figure 3 seems too wide because almost all observations are included.

Figures 4 and 6 illustrate the CIs based on interest rates close to zero. The residual-based CIs include negative values because of their symmetry, whereas the EL-based CIs do not. Moreover, LC in Figure 4 ranges from -3.0 to -2.65 for similar values of the dependent variable. By contrast, the LCEL is narrow in this region. Figure 6 reveals a similar phenomenon for LL and LLEL in the range from -2.85 to -2.7. Because the dependent-variable values in this range are similar and close to zero, the variance based on the local estimators is large. This overestimated variance seems to widen the CIs. Hence, the EL-based CIs succeed in reducing the spread of the CIs.

6 Concluding Remarks

In this paper, we derived EL-based pointwise CIs for the nonstationary nonparametric regression model. Simulation results showed that the empirical coverage probabilities of the local-constant EL-based CI outperform the residual-based ones. Our empirical study of Japan's inverse money demand function shows that the EL-based CI works well and exhibits the data-dependency property. Because our results show that the EL-based CIs do not include negative values of the interest rate, the EL-based CI approach contributes to the problems associated with working with dependent-variable data that are restricted.

Acknowledgments

In 2014, this paper was presented at IMS-APRM, the Joint Statistical Meeting and the international conference RSS. The author gratefully acknowledges the very helpful comments of Eiji Kurozumi, Toshio Honda and Yohei Yamamoto. This research was supported in part by Grant-in-Aid for JSPS Fellows Grant Number 819680.

References

Bae, Y. and de Jong, R. M. (2007). Money demand function estimation by nonlinear cointegration. *Journal of Applied Econometrics*, 22(4):767–793.

Bae, Y., Kakkar, V., and Ogaki, M. (2006). Money Demand in Japan and Nonlinear Cointegration. *Journal of Money, Credit, and Banking*, 38(6):1659–1667.

Chen, S. X., Härdle, W., and Li, M. (2003). An empirical likelihood goodness-of-fit test for time series. *J Royal Statistical Soc B*, 65:663–678.

Chen, S. X. and Qin, Y. S. (2000). Empirical Likelihood Confidence Intervals for Local Linear Smoothers. *Biometrika*, 87(4):946–953.

Choi, I. and Saikkonen, P. (2010). Tests for Nonlinear Cointegration. *Econometric Theory*, 26(03):682–709.

DiCiccio, T., Hall, P., and Romano, J. (1991). Empirical Likelihood is Bartlett-Correctable. *The Annaals of Statistics*, 19(2):1053–1061.

Hall, P. and Owen, A. B. (1993). Empirical Likelihood Confidence Bands in Density Estimation. *Journal of Computational and Graphical Statistics*, 2(3):273–289.

Honda, T. (2013). Nonparametric LAD cointegrating regression. *Journal of Multivariate Analysis*, 117:150–162.

Hu, L. and Phillips, P. C. B. (2004a). Dynamics of the federal funds target rate: a nonstationary discrete choice approach. *Journal of Applied Econometrics*, 19(7):851–867.

Hu, L. and Phillips, P. C. B. (2004b). Nonstationary discrete choice. *Journal of Econometrics*, 120(1):103–138.

Karlsen, H. A., Myklebust, T., and Tjøstheim, D. (2007). Nonparametric estimation in a nonlinear cointegration type model. *Ann. Statist.*, 35(1):252–299.

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Ann. Statist.*, 25(5):2084–2102.

Nabeya, S. and Perron, P. (1994). Local asymptotic distribution related to the AR(1) model with dependent errors. *Journal of Econometrics*, 62(2):229–264.

Ng, S. and Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, 69(6):1519–1554.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249.

Pantula, S. G. (1991). Asymptotic Distributions of Unit-Root Tests When the Process Is Nearly Stationary. *Journal of Business & Economic Statistics*, 9(1):63–71.

Park, J. Y. (2002). Nonstationary nonlinear heteroskedasticity. *Journal of Econometrics*, 110:383–415.

Park, J. Y. (2014). Nonstationary Nonlinearity: a Survey on Peter Phillips's Contributions with a New Perspective. *Econometric Theory*, pages 1–29.

Park, J. Y. and Phillips, P. C. B. (1999). Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory*.

Park, J. Y. and Phillips, P. C. B. (2000). Nonstationary Binary Choice. *Econometrica*, 68(5):1249–1280.

Park, J. Y. and Phillips, P. C. B. (2001). Nonlinear regressions with integrated time series. *Econometrica*, 69(1):117–161.

Tjøstheim, D. (2012). Modelling Nonlinear and Nonstationary Time Series. *Handbook of Statistics*, 30.

Wang, Q. and Phillips, P. C. B. (2009a). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory*, 25(03):1466–1497.

Wang, Q. and Phillips, P. C. B. (2009b). Structural nonparametric cointegrating regression. *Econometrica*, 77(6):1901–1948.

Wang, Q. and Phillips, P. C. B. (2011). Asymptotic Theory for Zero Energy Functionals with Nonparametric Regression Applications. *Econometric Theory*, 27(2):235–259.

Wang, Q. and Wang, Y. X. R. (2013). Non-parametric cointegrating regression with NNH errors. *Econometric Theory*, 29(1):1–27.

Xiong, X. and Lin, Z. (2012). Empirical likelihood inference for probability density functions under association. *Journal of statistical Planning and Inference*, 142(4):986–992.

Xu, K.-L. (2009). Empirical likelihood-based inference for nonparametric recurrent diffusions. *Journal of Econometrics*, 153(1):65–82.

Appendix

The following results are the auxiliary lemmas used for deriving the main results.

Lemma 1. *Under Assumptions 1–5, for any $h \to 0$, $nh^2 \to \infty$ and $nh^{2+4\gamma} \to 0$, the following asymptotic results (a), (b) and (c) hold.*

(a)

$$\frac{1}{(\sqrt{nh^2})^{1/2}} \sum_{t=1}^{n} g_t(a_0) \Rightarrow \sigma(x) \left(\frac{E[u_{m_0}^2]}{|\phi|} \right)^{1/2} N L_\kappa^{1/2}(1,0) \left(\int K^2(x) dx \right)^{1/2}.$$

(b)

$$\frac{1}{\sqrt{nh^2}} \sum_{t=1}^{n} g_t^2(a_0) \Rightarrow \sigma^2(x) \frac{E[u_{m_0}^2]}{|\phi|} \int K^2(x) dx L_\kappa(1,0).$$

(c)

$$\max |g_t(a_0)| = o_p((\sqrt{nh^2})^{1/4}).$$

Proof of (a)

Split $g_t(a_0)$ into the two terms

$$\sum_{t=1}^{n} g_t(a) = \sum_{t=1}^{n} (f(x_t) - f(x)) K[(x_t - x)/h] + \sum_{t=1}^{n} (\sigma(x_t) - \sigma(x)) u_t K[(x_t - x)/h] + \sigma(x) \sum_{t=1}^{n} u_t K[(x_t - x)/h]$$
$$:= \Theta_1 + \Theta_2 + \Theta_3.$$

Equation (3.6) of Wang and Phillips (2009b) implies that $\Theta_1 = O_p((nh^{2+2\gamma})^{1/2}) = o_p((nh^2)^{1/4})$, with the first term being asymptotically negligible. If we show that $\Theta_2 = o_p((nh^2)^{1/4})$, applying (3.8) of Wang and Phillips (2009b) to Θ_3 yields the required result. Therefore, it is sufficient to show that $\mathrm{E}[\Theta_2^2] = o_p((nh^2)^{1/2})$.

We split Θ_2^2 and x_t into

$$\Theta_2^2 = \sum_{t=1}^{n} (\sigma(x_t) - \sigma(x))^2 u_t^2 K^2[(x_t - x)/h] + 2 \sum_{1 \le s < t \le n} (\sigma(x_s) - \sigma(x))(\sigma(x_t) - \sigma(x)) u_s u_t K[(x_s - x)/h] K[(x_t - x)/h]$$
$$:= \Theta_4 + 2\Theta_5,$$

and $x_t := x^*_{s,t} + x'_{s,t}$, where

$$x^*_{s,t} = \rho^{t-s}x_s + \sum_{j=s+1}^{t} \rho^{t-j} \sum_{i=-\infty}^{s} \epsilon_i \phi_{j-i} \quad \text{and} \quad x'_{s,t} = \sum_{i=s+1}^{t} \epsilon_i \sum_{j=0}^{t-i} \rho^{t-j-i}\phi_j.$$

This representation of x_t is derived in (6.10) of Wang and Phillips (2009b). Note that $x^*_{s,t}$ depends only on $(\epsilon_s, \epsilon_{s-1}, \dots)$. By Assumption 3 and given the compactness domain of the kernel function in Assumption 4, there is a positive constant C such that

$$\Theta_4 < h^{2\gamma} \sum_{t=1}^{n} |(x_t - x)/h|^{2\gamma} u_t^2 K^2[(x_t - x)/h] < Ch^{2\gamma} \sum_{t=1}^{n} u_t^2 K^2[(x_t - x)/h].$$

Equation (7.6) of Wang and Phillips (2009b) implies that the right-hand side is $O_p(h^{2\gamma}\sqrt{nh^2}) = o_p((nh^2)^{1/2})$.

Because $x^*_{0,t}$ depends only on $(\epsilon_0, \epsilon_{-1}, \dots)$, we have almost surely

$$\begin{aligned} \mathrm{E}[\Theta_5|\epsilon_0, \epsilon_{-1}, \dots] &< \sum_{1\le s<t\le n} \sup_{y,z} \mathrm{E}[(\sigma(y + x'_{0,s}) - \sigma(x))(\sigma(z + x'_{0,t}) - \sigma(x))u_s u_t K[(y + x'_{0,s} - x)/h]K[(z + x'_{0,t} - x)/h]] \\ &= \sum_{1\le s<t\le n} \sup_{y,z} h^{2\gamma}\mathrm{E}[r(x'_{0,s}/h) r_1(x_{0,t}/h) g(u_s) g_1(u_t)], \end{aligned}$$

where

$$r(t) = \frac{(\sigma(x + h(y/h + t)) - \sigma(x))}{h^\gamma} K[y/h{+}t], \; r_1(t) = \frac{(\sigma(x + h(z/h + t)) - \sigma(x))}{h^\gamma} K[z/h{+}t] \quad \text{and} \quad g(t) = g_1(t) = t.$$

Because $|r(t)| < C|y/h{+}t|K[y/h{+}t]^\gamma$, we can apply Lemma 7.2 of Wang and Phillips (2009b) to the right-hand side. This reveals that $|\mathrm{E}\Theta_5|$ is bounded above for any $\epsilon > 0$ by

$$\begin{aligned} Ch^{2\gamma} \sum_{1\le s<t\le n} [\epsilon(t-s)^{-2} + h(t-s)^{-1}][(t-s)^{-2} + h/\sqrt{t-s}] &< Ch^{2\gamma}(\epsilon \sum_{t=1}^{n} t^{-2} + h \sum_{t=1}^{n} t^{-1}) \sum_{t=1}^{n} (t^{-2} + h/\sqrt{t}) \\ &< Ch^{2\gamma}(\epsilon + h\log n)(C + hn^{1/2}) \\ &= O(n^{1/2}h^{2(1+\gamma)}\log n) = o((nh^2)^{1/2}). \end{aligned}$$

Proof of (b)

$$\frac{1}{\sqrt{nh^2}}\sum_{t=1}^{n} g_t^2(a) = \frac{1}{\sqrt{nh^2}}\sum_{t=1}^{n}(f(x_t)-f(x))^2K^2[(x_t-x)/h] + 2\frac{1}{\sqrt{nh^2}}\sum_{t=1}^{n}\sigma(x_t)u_t(f(x_t)-f(x))K^2[(x_t-x)/h]$$
$$+\frac{1}{\sqrt{nh^2}}\sum_{t=1}^{n}\sigma^2(x_t)u_t^2K^2[(x_t-x)/h] := \Theta_6+2\Theta_7+\Theta_8.$$

First, we show that $\Theta_6 = o_p(1)$. Note that $x_{t,n} = x_t\big/d_t$ has a uniformly bounded density (say, $g_t(x_{t,n})$).

$$\begin{aligned}
\mathrm{E}[\Theta_6] &= \frac{1}{\sqrt{nh^2}}\sum_{t=1}^{n}\mathrm{E}[(f(d_tx_{t,n})-f(x))^2K^2[(d_tx_{t,n}-x)/h]] \\
&= \frac{1}{\sqrt{nh^2}}\sum_{t=1}^{n}\int \frac{h}{d_t}(f(x+hy)-f(x))^2K^2(y)g_t\left(\frac{x+hy}{x_t}\right)dy \\
&< C\sum_{t=1}^{n}\frac{h^{2\gamma}}{n}\int f_1^2(y,x)K^2(y)dy \\
&< Ch^{2\gamma}\left(\int f_1(y,x)K(y)dy\right)^2 = O\left(h^{2\gamma}\right) = o(1).
\end{aligned}$$

Because $\Theta_6 = o_p(1)$, it is sufficient to show that $\Theta_8 = O_p(1)$.

Second, we deal with Θ_7.

$$\begin{aligned}
|\Theta_7| &\le \frac{1}{\sqrt{nh^2}}\sum_{t=1}^{n}\sigma(x_t)|u_t||f(x_t)-f(x)|K^2[(x_t-x)/h] \\
&\le \left(\frac{1}{\sqrt{nh^2}}\sum_{t=1}^{n}\sigma^2(x_t)u_t^2K^2[(x_t-x)/h]\right)^{1/2}\left(\frac{1}{\sqrt{nh^2}}\sum_{t=1}^{n}(f(x_t)-f(x))^2K^2[(x_t-x)/h]\right)^{1/2} \quad \text{by Holder inequality} \\
&= \Theta_8^{1/2}\Theta_6^{1/2}.
\end{aligned}$$

Because $\Theta_6 = o_p(1)$, it is sufficient to show that $\Theta_8 = O_p(1)$.

From Proposition 7.2 of Wang and Phillips (2009b),

$$\frac{\sigma^2(x)}{\sqrt{nh^2}}\sum_{t=1}^{n} u_t^2 K^2[(x_t-x)/h] \Rightarrow \sigma^2(x)|\phi|^{-1}\mathrm{E}[u_{m0}^2]\int_{-\infty}^{\infty} K^2(s)dsL(1,0).$$

To complete the proof, we must show that

$$\frac{1}{\sqrt{nh^2}}\sum_{t=1}^{n} |\sigma^2(x_t)-\sigma^2(x)|u_t^2 K^2[(x_t-x)/h] = o_p(1).$$

From Assumption 3, $|\sigma^2(x_t)-\sigma^2(x)| \le |x_t-x|^{2\delta} + o_p(|x_t-x|^{2\delta})$. From the argument in the proof of (a), we obtain

$$\frac{h^{2\delta}}{\sqrt{nh^2}}\sum_{t=1}^{n} u_t^2|(x_t-x)/h|^{2\delta}K^2[(x_t-x)/h] \le C\frac{h^{2\delta}}{\sqrt{nh^2}}\sum_{t=1}^{n} u_t^2 K^2[(x_t-x)/h] = O(h^{2\delta}) = o(1).\square$$

Proof of (c)

$$\max_t |g_t(a_0)| \le \max_t |f(x_t)-f(x)|K[(x_t-x)/h] + \max_t |\sigma(x_t)u_t K[(x_t-x)/h]|$$

$$\begin{aligned}
\max_t |f(x_t)-f(x)|K[(x_t-x)/h] &= \max_t |f(x+h[(x_t-x)/h])-f(x)|K[(x_t-x)/h] \\
&\le \max_t h^{\gamma} f_1((x_t-x)/h,x)K[(x_t-x)/h] \\
&\le h^{\gamma}\int_{-\infty}^{\infty} f_1(y,x)K(y) < \infty.
\end{aligned}$$

Equation (7.26) of Wang and Phillips (2009b) shows that

$$\max_t |u_t K[(x_t-x)/h]| = o_p[(nh^2)^{1/4}].$$

Therefore,

$$\begin{aligned}\max_t |\sigma(x_t)u_t K[(x_t - x)/h]| &\le \max_t |(\sigma(x_t) - \sigma(x))u_t K[(x_t - x)/h]| + \max_t |\sigma(x)u_t K[(x_t - x)/h] \\ &< \max_t h^{\delta}|(x_t - x)/h|^{\delta}|u_t|K[(x_t - x)/h] + o_p[(nh^2)^{1/4}] \\ &\le Ch^{\delta} \max_t |u_t K[(x_t - x)/h]| + o_p((nh^2)^{1/4}) = o_p((nh^2)^{1/4}).\square\end{aligned}$$

Lemma 2. *Under Assumptions 1–5, for any $h \to 0$, $nh^2 \to \infty$ and $nh^{2+4\gamma} \to 0$, the following asymptotic results (a), (b) and (c) hold. (a)*

$$\frac{1}{(nh^2)^{1/4}} s_{n,1} = \frac{1}{(nh^2)^{1/4}} \sum_{t=1}^{n} [(x_t - x)/h] K[(x_t - x)/h] \Rightarrow |\phi|^{-1/2} N L^{1/2}(1,0) \left(\int s^2 K^2(s) ds \right)^{1/2}. \tag{6.1}$$

(b)

$$\frac{1}{(nh^2)^{1/2}} s_{n,2} = \frac{1}{(nh^2)^{1/2}} \sum_{t=1}^{n} [(x_t - x)/h]^2 K[(x_t - x)/h] \Rightarrow |\phi|^{-1} L(1,0) | \int_{-\infty}^{\infty} s^2 K(s) ds. \tag{6.2}$$

For any $h \to 0$, $nh^2 \to \infty$ and $nh^{10} \to 0$, the following asymptotic results hold.

(c)

$$\frac{1}{(nh^2)^{1/4}} \sum_{t=1}^{n} g_{t,L}(a_0) \Rightarrow \left(\frac{\mathrm{E}[u_{m_0}^2]}{|\phi|} \right)^{1/2} N L^{1/2}(1,0) \left(\int_{-\infty}^{\infty} K^2(x) dx \right)^{1/2}. \tag{6.3}$$

(d)

$$\frac{1}{(nh^2)^{1/2}} \sum_{t=1}^{n} g_{t,L}^2(a_0) \Rightarrow \frac{\mathrm{E}[u_{m_0}^2]}{|\phi|} L(1,0) \int_{-\infty}^{\infty} K^2(x) dx. \tag{6.4}$$

(e)

$$\max_t |g_{t,L}(a_0)| = o_p[(nh^2)^{1/4}]. \tag{6.5}$$

Proof of Lemma Results (a) and (b) are implied by Theorem 2.1 of Wang and Phillips (2011) and Corollary 2.2 of Wang and Phillips (2009a), respectively.

(c)

Given that

$$\frac{1}{(nh^2)^{1/4}}\sum_{t=1}^{n}(g_t(a_0)-g_{t,L}(a_0)) = -\frac{1}{(nh^2)^{1/4}}\frac{s_{n,1}}{s_{n,2}}\sum_{t=1}^{n}(y_t-a)[(x_t-x)/h]K[(x_t-x)/h],$$

we must show that the right-hand side is $o_p(1)$.

By (a) and (b) of Lemma 2, $s_{n,1}/s_{n,2} = O_p((nh^2)^{-1/4})$. Therefore, it is sufficient to show that

$$\frac{1}{\sqrt{nh^2}}\sum_{t=1}^{n}(f(x_t)-f(x))[(x_t-x)/h]K[(x_t-x)/h]+\frac{1}{\sqrt{nh^2}}\sum_{t=1}^{n}\sigma(x_t)u_t[(x_t-x)/h]K[(x_t-x)/h] := \Theta_8+\Theta_9 = o_p(1).$$

By the argument used in the proof of Theorem 2.2 of Wang and Phillips (2011), we can split Θ_8 into two terms as follows:

$$\Theta_8 = \frac{h}{\sqrt{nh^2}}f'(x)\sum_{t=1}^{n}[(x_t-x)/h]^2K[(x_t-x)/h]+\frac{1}{\sqrt{nh^2}}\sum_{t=1}^{n}[f(x_t)-f(x)-f'(x)(x_t-x)][(x_t-x)/h]K[(x_t-x)/h]. \tag{6.6}$$

The order of the first term is $O_p(h) = o_p(1)$. Given Assumptions 6 and 7, a Taylor-series expansion shows that there is a positive constant C such as

$$\sup_{y\in\Omega}|f(yh+x)-f(x)-f'(x)yh| \leq C|x_t-x|^2,$$

where Ω is the domain of the kernel function. The second term in equation (6.6) is bounded from above by

$$\frac{h^2}{\sqrt{nh^2}}\sum_{t=1}^{n}|(x_t-x)/h|^3K[(x_t-x)/h] = O_p(h^2) = o_p(1).$$

We have shown that Θ_8 is asymptotically negligible. The proof of Theorem 2.3. of Wang and Phillips (2011) shows that $\sum_{t=1}^{n}(y_t-a)[(x_t-x)/h]K[(x_t-x)/h] = o_p((nh^2)^{1/2})$.

(d)

$$\sum_{t=1}^{n} g_{t,L}^2(a_0) = \sum_{t=1}^{n} g_t^2(a_0) - 2\frac{s_{n,1}}{s_{n,2}} \sum_{t=1}^{n} g_t(a_0)(y_t - a)H_1[(x_t - x)/h] + \frac{s_{n,1}^2}{s_{n,2}^2} \sum_{t=1}^{n} (y_t - a)^2 H_1[(x_t - x)/h].$$

Because $K(\cdot)$ has a compact support,

$$\begin{aligned}\left|\sum_{t=1}^{n} g_t(a_0)(y_t - a)H_1[(x_t - x)/h]\right| &= \sum_{t=1}^{n} (y_t - a)^2 |(x_t - x)/h| K^2[(x_t - x)/h] \\ &< C\sum_{t=1}^{n} (y_t - a)^2 K^2[(x_t - x)/h] = C\sum_{t=1}^{n} g_t^2(a_0) = O_p((nh^2)^{1/2}).\end{aligned}$$

$$\left|\sum_{t=1}^{n} (y_t - a)^2 H_1[(x_t - x)/h]\right| < C\sum_{t=1}^{n} (y_t - a)^2 K^2[(x_t - x)/h] = O_p[(nh^2)^{1/2}].$$

(e) Lemma 1 (c) implies that

$$\max |g_{t,L}(a_0)| < \max_t |g_t(a_0)| + C\left|\frac{s_{n,1}}{s_{n,2}}\right| \max_t |g_t(a_0)| = o_p[(nh^2)^{1/4}].$$

$\square$.

Proof of Theorem 1 From equation (2.4), routine calculation yields the inequality

$$\frac{|\lambda|}{1 + |\lambda| \max_t |g_t(a_0)|} \leq \frac{|\sum_{t=1}^{n} g_t(a_0)|}{\sum_{t=1}^{n} g_t^2(a_0)}.$$

By virtue of (a) and (b) of Lemma 1, the order of the right-hand side is $O_p((nh^2)^{-1/4})$. Therefore, (c) of Lemma 1 implies that

$$|\lambda| = O_p((nh^2)^{-1/4}), \tag{6.7}$$

and

$$|\lambda| \max_t |g_t(a_0)| = o_p(1). \tag{6.8}$$

Result (b) of Lemma 1 and equations (6.7) and (6.8) reveal that

$$\begin{aligned}
0 &= \sum_{t=1}^n g_t(a_0) - \lambda \sum_{t=1}^n g_t^2(a_0) + \lambda^2 \sum_{t=1}^n \frac{g_t^3(a_0)}{1+\lambda g_t(a_0)} \\
&= \sum_{t=1}^n g_t(a_0) - \lambda \sum_{t=1}^n g_t^2(a_0) + O_p(|\lambda|^2 \max |g_t(a_0)| \sum_{t=1}^n g_t^2(a_0)) \\
&= \sum_{t=1}^n g_t(a_0) - \lambda \sum_{t=1}^n g_t^2(a_0) + o_p((nh^2)^{1/4}).
\end{aligned}$$

Therefore, the asymptotic distribution of λ is

$$(nh^2)^{1/4}\lambda = \frac{\sum_{t=1}^n g_t(a_0)/(nh^2)^{1/4}}{\sum_{t=1}^n g_t^2(a_0)/(nh^2)^{1/2}} + o_p(1) \Rightarrow NL^{-1/2}(1,0) \left(\int K^2(s)ds \right)^{-1/2} \left(\frac{\mathrm{E}[u_{m_0}^2]}{|\phi|} \right)^{-1/2}.$$

Expansion of the log-likelihood function in equation (2.3) yields the required result that

$$\begin{aligned}
2\sum_{t=1}^n \log(1+\lambda g_t(a_0)) &= 2\lambda \sum_{t=1}^n g_t(a_0) - \lambda^2 \sum_{t=1}^n g_t^2(a_0) + O_p(|\lambda|^3 \sum_{t=1}^n g_t^3(a_0)) \\
&= \lambda \sum_{t=1}^n g_t(a_0) + o_p(1) \\
&= \frac{\left(\sum_{t=1}^n g_t(a_0)\right)^2}{\sum_{t=1}^n g_t^2(a_0)} + o_p(1) \Rightarrow \chi_1^2. \square
\end{aligned}$$

Proof of Theorem 2

From (c) and (d) of Lemma 2, in the same way as in the proof of Theorem 1, we obtain

$$|\lambda| = O_p((nh^2)^{-1/4}),$$

and

$$|\lambda| \max_t |g_{c,t}(a_0)| = o_p(1).$$

The rest of the proof is omitted because it is so similar to that of Therorem1. $\square$

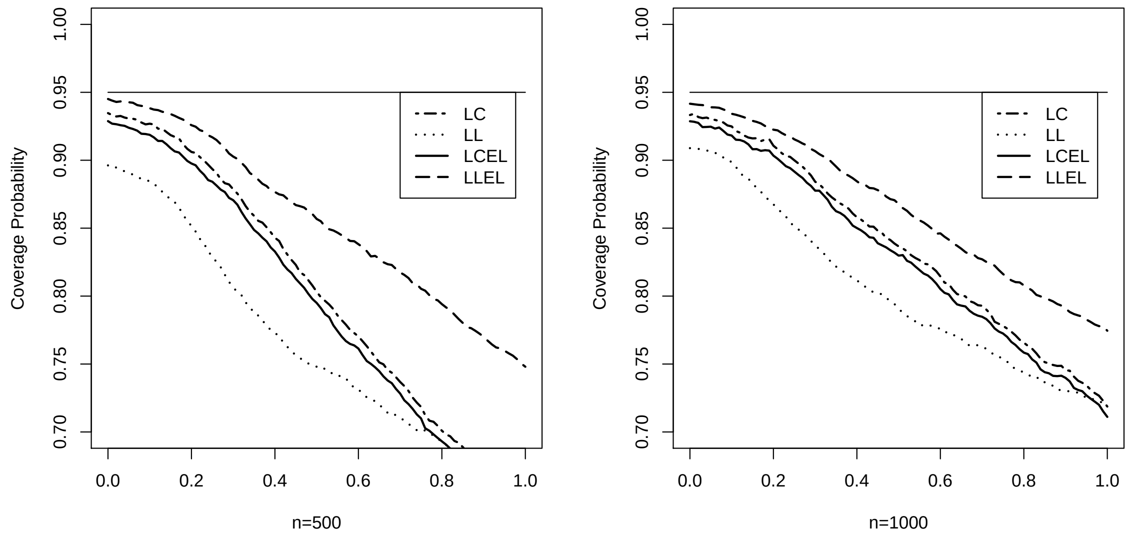

Figure 1: The empirical converge probabilities are shown for the local-constant and linear residual-based CIs (LC (point dashed line) and LL (dotted line)), and the local and linear EL-based CIs (LCEL (real line) and LLEL (dashed line)) for the case of $\sigma_\epsilon^2 = 0.01$.

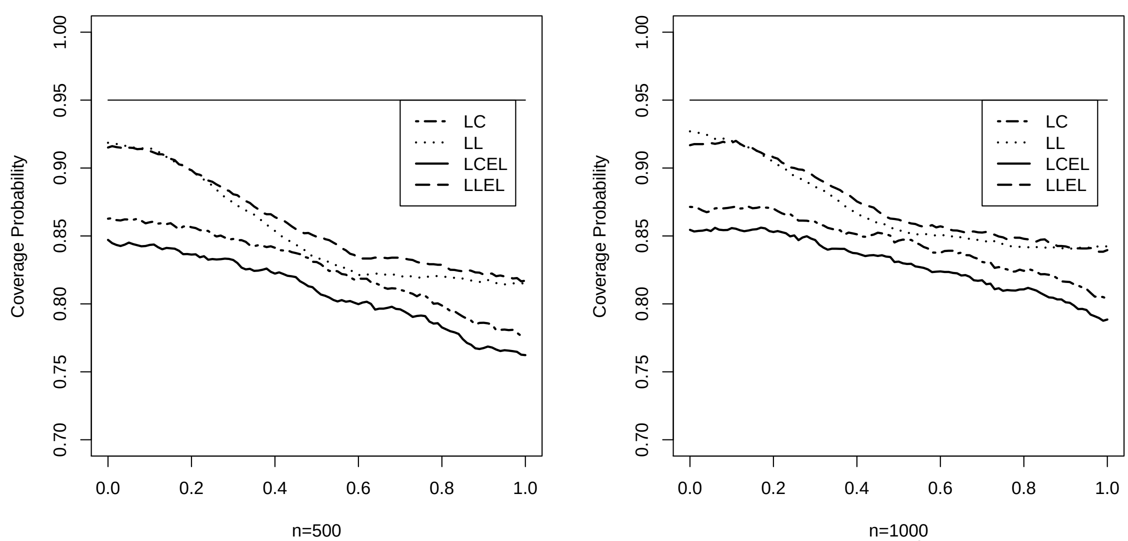

Figure 2: Empirical coverage probabilities for the case of $\sigma_\epsilon^2 = 0.1$.

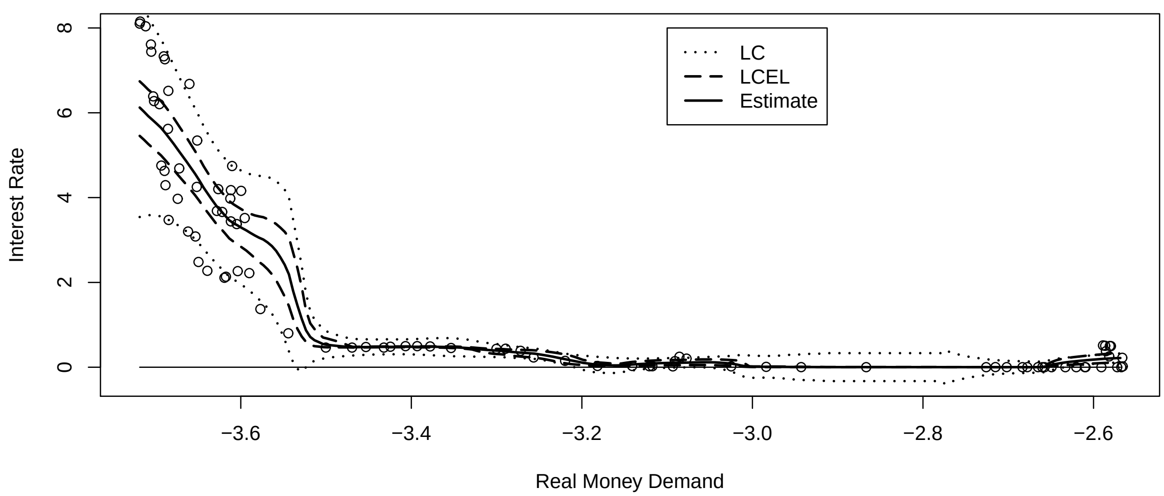

Figure 3: The local-constant-based 95% CIs for the inverse money demand function and the estimate from the local linear estimator. The EL-based CI (LCEL) is represented by the long dashed line. The residual-based CI (LC) is represented by the dashed line. The real line illustrates the local-constant estimator for the inverse money demand function.

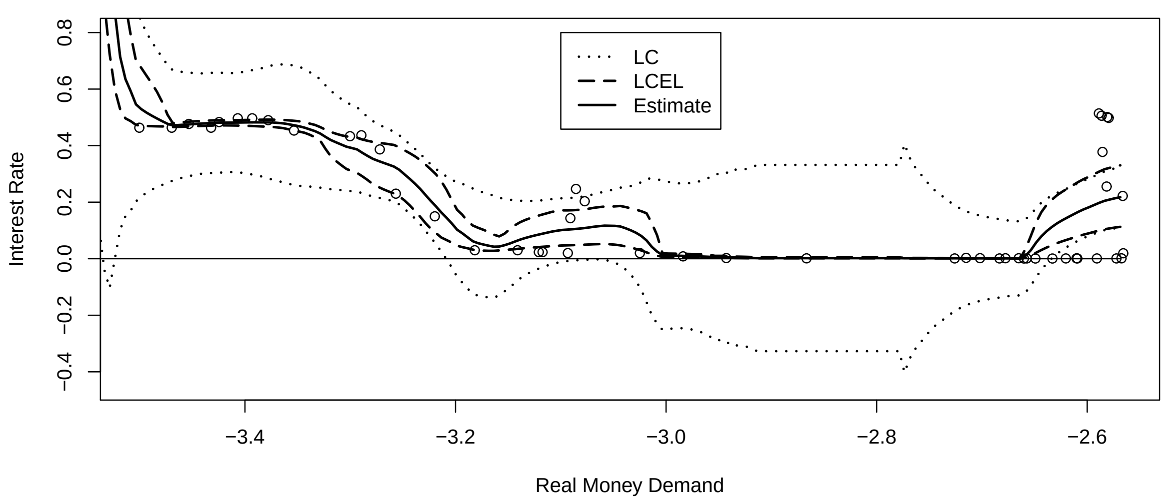

Figure 4: Figure 3 for when m_t changes from -3.6 to -2.55.

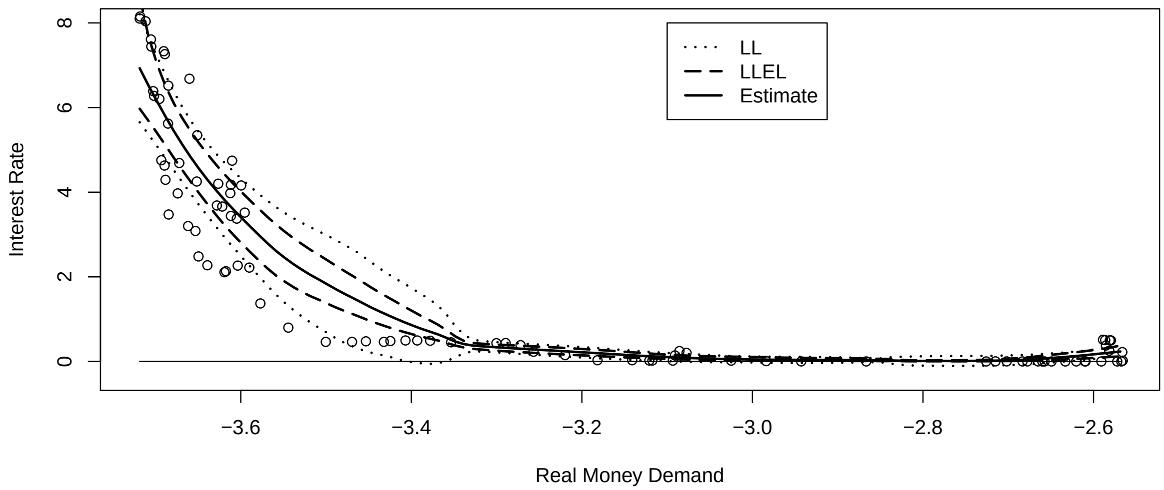

Figure 5: The local linear-based 95% CIs for the inverse money demand function and the estimate from the local linear estimator. The EL-based CI (LLEL) is represented by the long dashed line. The residual-based CI (LL) is represented by the dashed line. The real line illustrates the estimate from the local-constant estimator of the inverse money demand function.

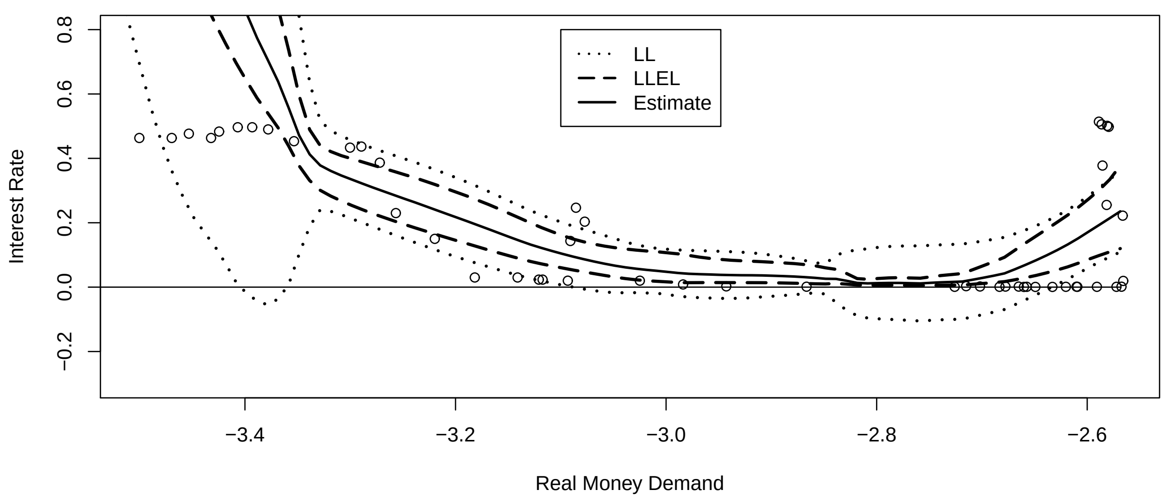

Figure 6: Figure 5 for when m_t changes from -3.6 to -2.55.