ASYMPTOTIC THEORY FOR REGRESSIONS WITH SMOOTHLY CHANGING PARAMETERS

ERIC HILLEBRAND, MARCELO C. MEDEIROS, AND JUNYUE XU

ABSTRACT. We derive the asymptotic properties of the quasi maximum likelihood estimator of smooth transition regressions when time is the transition variable. The consistency of the estimator and its asymptotic distribution are examined. It is shown that the estimator converges at the usual $\sqrt{T}$-rate and has an asymptotically normal distribution. The finite sample properties of the estimator are explored in simulations.
KEYWORDS: Regime switching; smooth transition regression; asymptotic theory.

1. INTRODUCTION

In this paper, we derive the asymptotic properties of the quasi maximum likelihood estimator (QMLE) of smooth transition regressions (STR) when time is the transition variable and the regressors are stationary. The consistency of the estimator and its asymptotic distribution are examined.

Nonlinear regression models have been widely used in practice for a variety of time series applications; see Granger and Teräsvirta (1993) for some examples in economics. In particular, STR models, initially proposed in its univariate form by Chan and Tong (1986), and further developed in Luukkonen, Saikkonen, and Teräsvirta (1988) and Teräsvirta (1994,1998), have been shown to be very useful for representing asymmetric behavior.[1] A comprehensive review of time series STR models is presented in van Dijk, Teräsvirta, and Franses (2002).

In most applications, stationarity, weak exogeneity,[2] and homoskedasticity have been imposed. In these cases, the standard method of estimation is nonlinear least squares (NLS), which is equivalent to quasi-maximum likelihood or, when the errors are Gaussian, to conditional maximum likelihood. The asymptotic properties of the NLS are discussed in Mira and Escribano (2000), Suarez-Fariñas, Pedreira, and Medeiros (2004), and Medeiros and Veiga (2005). Lundbergh and Teräsvirta (1998) and Li, Ling, and McAleer (2002) consider STR models with heteroskedastic errors. Chan, McAleer, and Medeiros (2005) study the properties of the QMLE when the errors follow a GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model. Saikkonen and Choi (2004) consider the case of STR models with cointegrated variables when the transition variable is integrated of order one, and

Date: February 11, 2010.

[1]The term "smooth transition" in its present meaning first appeared in Bacon and Watts (1971). They presented their smooth transition model as a generalization of models of two intersecting lines with an abrupt change from one linear regression to another at some unknown change point. Goldfeld and Quandt (1972, pp. 263–264) generalized the so-called two-regime switching regression model using the same idea.

[2]With respect to the parameters of interest.

Medeiros, Mendes, and Oxley (2009) analyze a similar model but with stationary transition variables. The case with endogenous regressors is considered in Areosa, McAleer, and Medeiros (in press).

An important case to consider is time as transition variable in STR models. Lin and Teräsvirta (1994) and Medeiros and Veiga (2003) consider this type of specification to construct models with parameters that change smoothly over time. Strikholm (2006) use this transition variable to determine the number of breaks in regression models. However, the asymptotic properties of the QMLE in this case have not been fully understood. If time is the transition variable, asymptotic theory of the QML estimator cannot be achieved in the standard way, because as the sample size T goes to infinity, the proportion of finite sub-samples goes to zero. Our solution to this problem is to scale the transition variable t so that the location of the transition is a certain fraction of the total sample rather than a fixed sample point. This modification allows asymptotic theory of the QML estimator. Andrews and McDermott (1995) and Saikkonen and Choi (2004) use similar transformations.

The outline of this paper is as follows. Section 2 describes the model and asymptotic properties of the QMLE. Monte Carlo simulations are presented in Section 3. Section 4 concludes the paper. All proofs are presented in the Appendix.

2. Model Definition and Estimation

2.1. **The Model.** We consider the following time series regression with time-varying parameters

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_0 + \sum_{m=1}^{M} \boldsymbol{x}_t'\boldsymbol{\beta}_m f[\gamma_m(t-c_m)] + \varepsilon_t,\ t = 1,\ 2,\ \ldots,\ T, \tag{1}$$

where ε_t is a martingale difference sequence with variance σ_ε^2. $\boldsymbol{x}_t$ is a vector of pre-determined regressors. The function f is the logistic transition function which has the form

$$f[\gamma(t-c)] = \frac{1}{1+e^{-\gamma(t-c)}}, t = 1,\ 2,\ \ldots,\ T. \tag{2}$$

where $\gamma > 0$ controls the smoothness of the transition and $c \in \{1,\ 2,\ \ldots, T\}$ is a location parameter. The loci $c_m \in \{1,\ 2,\ \ldots,\ T\}$ in (1) are change-points. Note that when $\gamma_m \longrightarrow \infty$, $m = 1, \ldots, M$, model (1) becomes a linear regression with M structural breaks occurring at the c_m.

2.2. **Embedding the Model in a Triangular Array.** Asymptotic theory for the QML estimator of the model defined above cannot be achieved the standard way. Consider model (1) with $M = 1$. As $T \to \infty$, the proportion of observations in the first regime goes to zero. Since for T large,

$$f\left[\gamma(t-c)\right] = f\left[T\gamma(T^{-1}t - T^{-1}c)\right] \approx \mathbb{1}_{\{T^{-1}t>0\}},$$

the parameter vector $\boldsymbol{\beta}_0$ that governs the first regime as well as the transition parameters γ and c vanish from the model and become unidentified. Figure 1 illustrates this. In the simulation, γ is set to be 0.2, c is equal to 50. In the upper plot of the figure, c is in the middle of the sample; in the lower plot ($T = 1000$), the second regime dominates. QML estimation of model (1) will be dominated by the second regime as the sample size increases. As the sample size goes to infinity, the first regime vanishes and its parameters become unidentified in the estimation. In order to obtain asymptotic theory for the estimator, the proportion of sub-samples in two regimes (before and after the transition) should remain constant as T goes to infinity. In other words, the shape of the plot of the time series should remain qualitatively the same as T grows. For this purpose, we scale the logistic transition function as

$$f\left[\gamma\left(\frac{T_0}{T}t-c\right)\right]=f\left[T^{-1}\gamma\left(T_0t-Tc\right)\right];\ t=1,\ldots,T;\ c\in\left[\frac{T_0}{T},T_0\right]. \tag{3}$$

where T_0 is the actual sample size in any given data situation. Accordingly,

$$y_t=\boldsymbol{x}_t'\boldsymbol{\beta}_0+\sum_{m=1}^{M}\boldsymbol{x}_t'\boldsymbol{\beta}_m f\left[\gamma_m\left(\frac{T_0}{T}t-c_m\right)\right]+\varepsilon_t. \tag{4}$$

Note that a given small-sample situation is embedded in this sequence of models at $T=T_0$. As can be seen in (3), with this scaling the slope of the logistic function is decreasing with T while the locus of the transition is increasing with T. The scaling of the time counter, T_0, remains constant. Therefore, the proportions of observations in the first regime, during the transition, and in the last regime remain the same as T grows, and the parameters in these groups of observations remain identified.

2.3. **Assumptions.** We denote the data-generating parameter vector as

$$\boldsymbol{\theta}_0=(\boldsymbol{\beta}_{0,0}',\boldsymbol{\beta}_{1,0}',\ldots,\boldsymbol{\beta}_{M,0}',\gamma_{1,0},\ldots,\gamma_{M,0},c_{1,0},\ldots,c_{M,0},\sigma^2_{\varepsilon,0})',$$

where the (second) 0-subscript indicates the data-generating character.

We write $\varepsilon_t(\boldsymbol{\theta})$ such that the notation can be used for both the residuals from the estimation and the data-generating errors:

$$\varepsilon_t\left(\boldsymbol{\theta}\right)=y_t-g(\boldsymbol{x}_t;\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{c})$$

where $\boldsymbol{\beta}=(\boldsymbol{\beta}_0,\ldots,\boldsymbol{\beta}_M)';\boldsymbol{\gamma}=(\gamma_1,\ldots,\gamma_M)';\boldsymbol{c}=(c_1,\ldots,c_M)'$ and

$$g(\boldsymbol{x}_t;\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{c})=\boldsymbol{x}_t'\boldsymbol{\beta}_0+\sum_{m=1}^{M}\boldsymbol{x}_t'\boldsymbol{\beta}_m f\left[\gamma_m\left(\frac{T_0}{T}t-c_m\right)\right].$$

We use the shorthand notation $\varepsilon_{t,0}:=\varepsilon_t(\boldsymbol{\theta}_0)$, for the data-generating errors and $\varepsilon_t=\varepsilon_t(\boldsymbol{\theta})$ for the residual evaluated at any $\boldsymbol{\theta}$.

We consider the following assumptions.

ASSUMPTION 1 (Parameter Space). *The parameter vector $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}$, a compact real parameter space.*

ASSUMPTION 2 (Errors).

(1) $\varepsilon_{t,0}$ *is a martingale difference sequence with constant variance* $\sigma^2_\varepsilon>c>0$.
(2) $\mathbb{E}|\varepsilon_{t,0}|^q<\infty$ *for* $q\leq 4$.
(3) $\boldsymbol{x}_t$ *and* $\varepsilon_{t,0}$ *are independent.*

ASSUMPTION 3 (Stationarity and Moments).

(1) $\boldsymbol{x}_t = (\boldsymbol{x}_{A,t}, \boldsymbol{x}_{B,t})'$, *where* $\boldsymbol{x}_{A,t}$ *consists of stationary and ergodic exogenous variables and* $\boldsymbol{x}_{B,t}$ *is a set of lagged values of* y_t. *The autoregressive polynomial in each regime associated to* $\boldsymbol{x}_{B,t}$ *has all roots outside the unit circle.*
(2) $\mathbb{E}\,\|\boldsymbol{x}_{A,t}\|^q < \infty$ *for* $q \leq 4$, *where* $\|\cdot\|$ *is the Euclidean vector norm.*
(3) $\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{x}_t\boldsymbol{x}_t')$ *converges in probability to* $\Omega = \mathbb{E}\,(\boldsymbol{x}_t\boldsymbol{x}_t')$, *which is symmetric positive definite.*

ASSUMPTION 4 (Transition Function). $g(\boldsymbol{x}_t;\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{c})$ *is parameterized such that the parameters are well defined.*

Assumption 1 is standard in the literature and is not too restrictive in the present case as we expect $\boldsymbol{\beta}_0$ to be finite, $\boldsymbol{\gamma}_0$ is positive and finite, and $\boldsymbol{c}_0 \in [0,1]$. Assumption 2 is also standard.

2.4. **Quasi Maximum Likelihood Estimator.** The quasi log-likelihood function is given by

$$\mathcal{L}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T}\ell_t(\boldsymbol{\theta}),$$

where

$$\ell_t(\boldsymbol{\theta}) = -\frac{1}{2}\left(\log 2\pi + \log\sigma_\varepsilon^2 + \varepsilon_t^2\sigma_\varepsilon^{-2}\right).$$

The parameter vector is estimated by quasi maximum likelihood as

$$\widehat{\boldsymbol{\theta}}_T = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmax}}\ \mathcal{L}_T(\boldsymbol{\theta}), \tag{5}$$

where $\boldsymbol{\Theta}$ is the parameter space.

THEOREM 1 (Consistency). *Under Assumptions 1 through 4, the quasi maximum likelihood estimator* $\widehat{\boldsymbol{\theta}}_T$ *is consistent:*

$$\widehat{\boldsymbol{\theta}}_T \xrightarrow{p} \boldsymbol{\theta}_0.$$

The proof is provided in the Appendix.

THEOREM 2 (Asymptotic Normality). *Under Assumptions 1 through 4, the quasi maximum likelihood estimator* $\widehat{\boldsymbol{\theta}}_T$ *is asymptotically normally distributed:*

$$\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathcal{N}\left[0,\ \boldsymbol{A}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{B}(\boldsymbol{\theta}_0)\boldsymbol{A}(\boldsymbol{\theta}_0)^{-1}\right], \tag{6}$$

where

$$\boldsymbol{A}(\boldsymbol{\theta}_0) = -\mathbb{E}\left(\left.\frac{\partial^2\ell_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}_0}\right),$$

$$\boldsymbol{B}(\boldsymbol{\theta}_0) = \mathbb{E}\left(\left.\frac{\partial\ell_t}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}_0}\left.\frac{\partial\ell_t}{\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}_0}\right).$$

PROPOSITION 1 (Covariance Matrix Estimation). *Under Assumptions 1 through 4,*

$$\boldsymbol{A}_T \xrightarrow{p} \boldsymbol{A}, \quad \boldsymbol{B}_T \xrightarrow{p} \boldsymbol{B},$$

where

$$\boldsymbol{A}_T(\boldsymbol{\theta}) = -\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \ell_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'},$$

and

$$\boldsymbol{B}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial \ell_t}{\partial\boldsymbol{\theta}}\frac{\partial \ell_t}{\partial\boldsymbol{\theta}'},$$

and $\boldsymbol{A}$, $\boldsymbol{B}$ *as defined in Theorem 2.*

3. Small Sample Simulations

We conduct a set of Monte Carlo simulations in order to evaluate both the small-sample properties and the asymptotic behavior of the QMLE. In particular, we consider the following models with three limiting regimes:

Model A – Independent and identically distributed (IID) regressors:

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_0 + \sum_{m=1}^{2}\boldsymbol{x}_t'\boldsymbol{\beta}_m f\left[\gamma_m\left(\frac{t}{T} - c_m\right)\right] + \varepsilon_t,$$
$$y_t = 1 + x + (-1 - 2x)f\left[30\left(\frac{t}{T} - \frac{1}{3}\right)\right] + (1 + 3x)f\left[30\left(\frac{t}{T} - \frac{2}{3}\right)\right] + \varepsilon_t,$$

where $\{x_t\}$ is a sequence of independent and normally distributed random variables with zero mean and unit variance, $x_t \sim \mathsf{NID}(0,1)$, and $\{\varepsilon_t\}$ is either a sequence of $\mathsf{NID}(0,1)$ or $\mathsf{Uniform}(-2,2)$ random variables.

Model B – Dependent regressors:

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_0 + \sum_{m=1}^{2}\boldsymbol{x}_t'\boldsymbol{\beta}_m f\left[\gamma_m\left(\frac{t}{T} - c_m\right)\right] + \varepsilon_t,$$
$$y_t = 0.5 + 0.4y_{t-1} + (-0.5 + 0.5y_{t-1})f\left[30\left(\frac{t}{T} - \frac{1}{3}\right)\right] + (0.5 - 1.7y_{t-1})f\left[30\left(\frac{t}{T} - \frac{2}{3}\right)\right] + \varepsilon_t,$$

where $\{\varepsilon_t\}$ is either a sequence of $\mathsf{NID}(0,1)$ or $\mathsf{Uniform}(-2,2)$ random variables.

Different values of T are used, ranging from 100 to 5000 observations. For each value of T, 1000 Monte Carlo simulations are repeated. When the errors are normally distributed, the estimators are maximum likelihood estimators. On the other hand, when the errors are uniformly distributed, the error distribution is misspecified and we have a quasi maximum likelihood estimation setup. For sample sizes up to 300 observations, the estimation procedure did not converge in less than 5% of the replications. These cases were discarded. The parameters γ are chosen in order to keep the transitions neither too smooth nor too sharp; see Figure 2.

The results are presented in Figures 3–14. Figures 3–6 show the average bias and the mean squared error (MSE) as a function of the sample size. Apart from the slope parameter, the average biases are rather small for all sample sizes, models, and error distributions. Furthermore, the MSE, as expected, converges to zero as the sample size increases. With respect to the slope parameter, the MSE is quite high for very small samples (100–300 observations) but also goes to zero as the sample size increases. The bias is also large in small sample, but turns to be negligible for larger sample sizes. The large biases and MSE are mainly caused by few very large estimates (less than 1% of the cases). For example, for Model A with Gaussian errors and 100 observations, the average bias and MSE for the first slope parameter ($\widehat{\gamma}_1$) are, respectively 908.82 and $106,447,280.55$. On the other hand, the median bias is just 13.00. For 500 observations and the same model, the average bias and MSE are 19.28 and $155,859.76$, respectively. The median bias is just 0.66 when $T = 500$. This pattern is somehow expected, as it is quite difficult to estimate the slope parameters in small samples. On the other hand, the location ($\boldsymbol{c}$) and the linear parameters ($\boldsymbol{\beta}$) are estimated quite precisely.

Figures 7–10 present the distribution the standardized QMLE of the linear parameters of the model ($\boldsymbol{\beta}$). Some interesting facts emerge from the graphs. First, even in very small samples, the estimate $\widehat{\boldsymbol{\beta}}_0$ has a distribution close to normal for all models and error distributions. Second, the distributions of $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$ have some outliers in small samples, but, as expected, they are close to normal for very large samples ($T = 5,000$).

Turning to the location parameter, Figures 11–14 show the distribution of the standardized QMLE for $\boldsymbol{c}$. It is quite remarkable that even for $T = 100$, the empirical distributions are close to normal.

4. CONCLUSION

In this paper, we propose asymptotic theory for the QML estimator of a logistic smooth transition regression model with time as the transition variable. Although asymptotic theory cannot be achieved in the standard way as the transition variable is not stationary, after proper scaling, we show that the QML estimator is consistent and asymptotically normal. The estimator is shown to converge to the true value of the parameter at the speed of $\sqrt{T}$. We explore the small sample behavior in simulations.

APPENDIX A. PROOF OF CONSISTENCY

Proof of Theorem 1. We establish the conditions for consistency according to Theorem 4.1.1 of Amemiya (1985). We have $\widehat{\boldsymbol{\theta}}_T \xrightarrow{p} \boldsymbol{\theta}_0$ if the following conditions hold:

(1) $\boldsymbol{\Theta}$ is a compact parameter set.
(2) $\mathcal{L}_T(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_t)$ is continuous in $\boldsymbol{\theta}$ and measurable in $\boldsymbol{\varepsilon}_t$.
(3) $\mathcal{L}_T(\boldsymbol{\theta})$ converges to a deterministic function $\mathcal{L}(\boldsymbol{\theta})$ in probability uniformly on $\boldsymbol{\Theta}$ as $T \to \infty$.
(4) $\mathcal{L}(\boldsymbol{\theta})$ attains a unique global maximum at $\boldsymbol{\theta}_0$.

Item (1) is given by Assumption 1. Item (2) holds by definition of the quasi-maximum likelihood estimator (5) from the definition of the normal density. For item (3) we refer to Theorem 4.2.1 of Amemiya (1985): This holds for i.i.d. data if $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\ell_t(\boldsymbol{\theta})| < \infty$ and $\ell_t(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$ for each ε_t. The extension to stationary and ergodic data using the same set of assumptions is achieved in Ling and McAleer (2003, Theorem 3.1). We have $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\ell_t(\boldsymbol{\theta})| < \infty$ by Jensen's inequality and $\mathbb{E}\sup |\phi(\boldsymbol{\varepsilon}_t, \boldsymbol{\theta})| < \infty$, where ϕ denotes the normal density function. The finiteness of the last expression follows from the assumption that $\sigma_\varepsilon^2 > c > 0$ for some constant c. The log density $\log \phi(\varepsilon_t, \boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$ for every ε_t.

Consider Item (4). By the Ergodic Theorem, $\mathbb{E}\ell_t(\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta})$. Rewrite the maximization problem as

$$\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathbb{E}\left[\ell_t\left(\boldsymbol{\theta}\right) - \ell_t(\boldsymbol{\theta}_0)\right].$$

Now, for a given number σ_ε^2,

$$\begin{aligned}
\mathbb{E}\left[\ell_t\left(\boldsymbol{\theta}\right) - \ell_t\left(\boldsymbol{\theta}_0\right)\right] &= \mathbb{E}\log\left[\frac{\phi(\boldsymbol{\varepsilon}_t, \boldsymbol{\theta})}{\phi(\boldsymbol{\varepsilon}_t, \boldsymbol{\theta}_0)}\right], \\
&= \mathbb{E}\left[-\frac{1}{2}\log\frac{\sigma_\varepsilon^2}{\sigma_{\varepsilon,0}^2} - \frac{1}{2}\left(\frac{\varepsilon_t^2}{\sigma_\varepsilon^2} - \frac{\varepsilon_{t,0}^2}{\sigma_{\varepsilon,0}^2}\right)\right], \\
&= -\frac{1}{2}\log\frac{\sigma_\varepsilon^2}{\sigma_{\varepsilon,0}^2} - \frac{1}{2}\left[\mathbb{E}(\varepsilon_t^2\sigma_\varepsilon^{-2}) - 1\right].
\end{aligned} \tag{7}$$

We show that $\mathbb{E}\varepsilon_t^2(\boldsymbol{\theta}) \geq \mathbb{E}\varepsilon_{t,0}^2 = \sigma_{\varepsilon,0}^2$ and that (7) attains an upper bound at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ uniquely. Consider

$$\mathbb{E}\varepsilon_t^2(\boldsymbol{\theta}) = \mathbb{E}\left[y_t - g(\boldsymbol{x}_t; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{c})\right]^2.$$

Substituting for $y_t = g(\boldsymbol{x}_t; \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \boldsymbol{c}_0) + \varepsilon_{t,0}$ and rearranging, we obtain

$$\begin{aligned}
\mathbb{E}\varepsilon_t^2(\boldsymbol{\theta}) &= \mathbb{E}\left[g(\boldsymbol{x}_t; \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \boldsymbol{c}_0) + \varepsilon_{t,0} - g(\boldsymbol{x}_t; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{c})\right]^2, \\
&\geq \mathbb{E}\varepsilon_{t,0}^2 = \sigma_{\varepsilon,0}^2.
\end{aligned}$$

The inequality holds from Assumption 2 (3). We have established that for any given σ_ε^2, the objective function (7) attains its maximum of

$$-\frac{1}{2}\left(\log\frac{\sigma_\varepsilon^2}{\sigma_{\varepsilon,0}^2} + \frac{\sigma_{\varepsilon,0}^2}{\sigma_\varepsilon^2} - 1\right)$$

at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$, $\boldsymbol{c} = \boldsymbol{c}_0$. Define $x = \sigma_\varepsilon^2/\sigma_{\varepsilon,0}^2$, then

$$f(x) = -\frac{1}{2}\left(\log x + \frac{1}{x} - 1\right)$$

attains its maximum of 0 at $x = 1$, therefore the maximizer is $\sigma_\varepsilon^2 = \sigma_{\varepsilon,0}^2$. This shows that $\mathbb{E}(\ell_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta}_0))$ is uniquely maximized at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. □

Appendix B. Proof of Asymptotic Normality

Remark 1.

(1) *In this proof, terms will sometimes involve expectations of cross-products of the type* $\mathbb{E}(XY)$*, where* X *and* Y *are correlated random variables. Note that by the Cauchy-Schwarz inequality, we have*

$$\mathbb{E}XY \leq \left(\mathbb{E}X^2\right)^{\frac{1}{2}} \left(\mathbb{E}Y^2\right)^{\frac{1}{2}},$$

and thus in order to show that the cross-product has finite expectation, it suffices to show that both random variables have finite second moments.

(2) *By the same token, if both* X *and* Y *have finite second moments,*

$$\begin{aligned}\mathbb{E}(X+Y)^2 &\leq \mathbb{E}X^2 + \mathbb{E}Y^2 + 2\left(\mathbb{E}X^2\right)^{\frac{1}{2}} \left(\mathbb{E}Y^2\right)^{\frac{1}{2}}, \\ &\leq K(\mathbb{E}X^2 + \mathbb{E}Y^2),\end{aligned}$$

for some $K < \infty$.

In the outline of the proof we follow Theorem 4.1.3 of Amemiya (1985). Therefore we have to establish the conditions

(1) $\frac{\partial^2 \ell_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$ exists and is continuous in an open neighborhood of $\boldsymbol{\theta}_0$.
(2) $\boldsymbol{A}_T(\boldsymbol{\theta}_T^*) \xrightarrow{p} \boldsymbol{A}(\boldsymbol{\theta}_0)$ for all sequences $\boldsymbol{\theta}_T^* \xrightarrow{p} \boldsymbol{\theta}_0$.
(3)

$$\boldsymbol{B}(\boldsymbol{\theta}_0)^{-\frac{1}{2}} \frac{1}{\sqrt{T}} \sum_{t=1}^{[rT]} \left.\frac{\partial \ell_t}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}_0} \xrightarrow{d} W(s),\ s \in [0,1],$$

where W is standard Brownian motion on the unit interval.

Item (1) is shown in Lemma 3. Item (2) needs consistency of $\widehat{\boldsymbol{\theta}}_T$ for $\boldsymbol{\theta}_0$, which we established in Theorem 1. It further needs uniform convergence of $\boldsymbol{A}_T$ to $\boldsymbol{A}$, i.e.

$$\sup_{\boldsymbol{\theta} \in \Theta} |\boldsymbol{A}_T(\boldsymbol{\theta}) - \boldsymbol{A}(\boldsymbol{\theta})| \xrightarrow{p} 0.$$

We use Ling and McAleer (2003, Theorem 3.1) to establish this, which achieved invocation of the Ergodic Theorem without having to show finiteness of third order derivative information. We show the uniform convergence in Lemma 4.

Item (3) uses Billingsley (1999, Theorem 18.3) and needs (a) that $\{\partial\ell_t/\partial\boldsymbol{\theta}|\boldsymbol{\theta}_0, \mathcal{F}_t\}$ is a stationary martingale difference sequence and (b) that $\boldsymbol{B}(\boldsymbol{\theta}_0)$ exists. Both with be proved in Lemma 3. The first two lemmas show a few technical properties of $g(\boldsymbol{x}_t; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{c})$ that are needed in the following.

LEMMA 1. *The transition function given by Equation* (3) *is bounded, and so are its first and second derivatives with respect to* γ_m *and* c_m, $\forall m = 1, 2, \ldots M$.

Proof. We will use shorthand notation f for $f\left[\gamma_m\left(\frac{T_0}{T}t - c_m\right)\right]$ below unless otherwise stated. Defining $a_m(t) := \frac{T_0}{T}t - c_m, t = 1, 2\ldots, T$, it is easy to verify that $-\infty < -c_m < a_m(t) \le T_0 - c_m < \infty$. Since the transition function has the range $(0, 1)$, it is clearly bounded. For the first derivative of f with respect to γ_m, $\forall m = 1, 2, \ldots M$,

$$\left|\frac{\partial f}{\partial \gamma_m}\right| = \left|\frac{a_m(t)e^{-\gamma_m a_m(t)}}{(1 + e^{-\gamma_m a_m(t)})^2}\right| \le |a_m(t)f| < \infty.$$

The first inequality follows from the fact that $1 + e^{-\gamma_m a_m(t)} > e^{-\gamma_m a_m(t)} > 0$. The second inequality holds because both $a_m(t)$ and f are bounded. For the second derivative of f with respect to c_m, $\forall m = 1, 2, \ldots M$,

$$\begin{aligned}
\left|\frac{\partial^2 f}{\partial \gamma_m^2}\right| &= \left|\frac{2a_m(t)^2 e^{-2\gamma_m a_m(t)}}{\left(1 + e^{-\gamma_m a_m(t)}\right)^3} + \frac{a_m(t)^2 e^{-\gamma_m a_m(t)}}{\left(1 + e^{-\gamma_m a_m(t)}\right)^2}\right|, \\
&\le \left|\frac{2a_m(t)^2 e^{-2\gamma_m a_m(t)}}{\left(1 + e^{-\gamma_m a_m(t)}\right)^3}\right| + \left|\frac{a_m(t)^2 e^{-\gamma_m a_m(t)}}{\left(1 + e^{-\gamma_m a_m(t)}\right)^2}\right|, \\
&\le \left|\frac{2a_m(t)^2}{1 + e^{-\gamma_m a_m(t)}}\right| + \left|\frac{a_m(t)^2}{1 + e^{-\gamma_m a_m(t)}}\right|, \\
&= \left|3a_m(t)^2 f\right| < \infty.
\end{aligned}$$

The second inequality follows from the fact that $1 + e^{-\gamma_m a_m(t)} > e^{-\gamma_m a_m(t)} > 0$, the last inequality holds because both $a_m(t)$ and f are bounded. The proof of the boundedness of the first and second derivatives of f with respect to c_m is almost identical to the one above and is omitted for brevity. $\square$

LEMMA 2.

Let $\boldsymbol{\xi} := (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{c})$, *then*

(1) $\mathbb{E}\left\|\frac{\partial}{\partial\boldsymbol{\xi}} g(\boldsymbol{x}_t; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{c})\right\|^2 < \infty.$

(2) $\mathbb{E}\left\|\frac{\partial^2}{\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}'} g(\boldsymbol{x}_t; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{c})\right\|^2 < \infty$, *where* $\|\cdot\|$ *denotes the standard vector and matrix norms.*

Proof. We will prove the statements element by element. For statement (1),

$$\mathbb{E}\left\|\frac{\partial}{\partial\boldsymbol{\beta}_0} g(\boldsymbol{x}_t; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{c})\right\|^2 = \mathbb{E}\left\|\boldsymbol{x}_t\right\|^2 < \infty$$

by Assumption 3 (2).

$$\mathbb{E}\left\|\frac{\partial}{\partial\boldsymbol{\beta}_m} g\left(\boldsymbol{x}_t; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{c}\right)\right\|^2 = \mathbb{E}\left\|\boldsymbol{x}_t f\right\|^2 \le \mathbb{E}\left\|\boldsymbol{x}_t\right\|^2 < \infty,$$

by the fact that $|f| < 1$.

$$\mathbb{E}\left\|\frac{\partial}{\partial\gamma_m}g(\boldsymbol{x}_t;\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{c})\right\|^2 = \mathbb{E}\left\|\boldsymbol{x}_t'\boldsymbol{\beta}_m\frac{\partial f}{\partial\gamma_m}\right\|^2,$$
$$\leq \mathbb{E}\left\|\boldsymbol{x}_t\right\|^2\left\|\boldsymbol{\beta}_m\right\|^2\left|\frac{\partial f}{\partial\gamma_m}\right|^2 < \infty$$

by Lemma 1, Assumption 1, and Assumption 3 (2). Similarly,

$$\mathbb{E}\left\|\frac{\partial}{\partial c_m}g(\boldsymbol{x}_t;\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{c})\right\|^2 = \mathbb{E}\left\|\boldsymbol{x}_t'\boldsymbol{\beta}_m\frac{\partial f}{\partial c_m}\right\|^2,$$
$$\leq \mathbb{E}\left\|\boldsymbol{x}_t\right\|^2\left\|\boldsymbol{\beta}_m\right\|^2\left|\frac{\partial f}{\partial c_m}\right|^2 < \infty.$$

For statement (2),

$$\mathbb{E}\left\|\frac{\partial^2}{\partial\boldsymbol{\beta}_0\partial\boldsymbol{\beta}_0'}g(\boldsymbol{x}_t;\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{c})\right\|^2 = \mathbf{0},$$
$$\mathbb{E}\left\|\frac{\partial^2}{\partial\boldsymbol{\beta}_m\partial\boldsymbol{\beta}_m'}g(\boldsymbol{x}_t;\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{c})\right\|^2 = \mathbf{0},$$

$$\mathbb{E}\left\|\frac{\partial^2}{\partial\gamma_m^2}g(\boldsymbol{x}_t;\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{c})\right\|^2 = \mathbb{E}\left\|\boldsymbol{x}_t'\boldsymbol{\beta}_m\frac{\partial^2 f}{\partial^2\gamma_m^2}\right\|^2,$$
$$\leq \mathbb{E}\left\|\boldsymbol{x}_t\right\|^2\left\|\boldsymbol{\beta}_m\right\|^2\left|\frac{\partial^2 f}{\partial\gamma_m^2}\right|^2 < \infty.$$

For the second inequality, we use the fact that $\left|\frac{\partial^2 f}{\partial\gamma_m^2}\right|$ is bounded from Lemma 1.

Similarly,

$$\mathbb{E}\left\|\frac{\partial^2}{\partial c_m^2}g(\boldsymbol{x}_t;\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{c})\right\|^2 = \mathbb{E}\left\|\boldsymbol{x}_t'\boldsymbol{\beta}_m\frac{\partial f}{\partial^2 c_m^2}\right\|^2,$$
$$\leq \mathbb{E}\left\|\boldsymbol{x}_t\right\|^2\left\|\boldsymbol{\beta}_m\right\|^2\left|\frac{\partial^2 f}{\partial c_m^2}\right|^2 < \infty.$$

□

LEMMA 3.

(1) *The sequence* $\left\{\frac{\partial\ell_t}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}_0}, \mathcal{F}_t\right\}_{t=1,\dots,T}$ *is a stationary martingale difference sequence.* $\mathcal{F}_t$ *is the sigma-algebra given by all information up to time* t.

(2)
$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\mathbb{E}\left\|\frac{\partial\ell_t}{\partial\boldsymbol{\theta}}\right\| < \infty,$$

(3)
$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\mathbb{E}\left\|\frac{\partial\ell_t}{\partial\boldsymbol{\theta}}\frac{\partial\ell_t}{\partial\boldsymbol{\theta}'}\right\| < \infty.$$

Proof. For part (1) of the proof, all derivatives are evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. The nought-subscript is suppressed to reduce notational clutter. Let $\boldsymbol{\xi} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{c})$, as before.

$$\mathbb{E}\left(\frac{\partial \ell_t}{\partial \boldsymbol{\xi}}\bigg|\, \mathcal{F}_{t-1}\right) = \mathbb{E}\left(-\frac{\varepsilon_t}{\sigma_\varepsilon^2}\frac{\partial \varepsilon_t}{\partial \boldsymbol{\xi}}\bigg|\, \mathcal{F}_{t-1}\right) = \mathbb{E}\left(\frac{\varepsilon_t}{\sigma_\varepsilon^2}\frac{\partial}{\partial \boldsymbol{\xi}} g(\boldsymbol{x}_t; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{c})\bigg|\, \mathcal{F}_{t-1}\right) = 0,$$

since $g(\boldsymbol{x}_t; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{c})$ is independent of ε_t and its derivatives are bounded (Lemma 2).

$$\mathbb{E}\left(\frac{\partial \ell_t}{\partial \sigma_\varepsilon^2}\bigg|\, \mathcal{F}_{t-1}\right) = \mathbb{E}\left(-\frac{1}{2\sigma_\varepsilon^2} + \frac{1}{2}\frac{\varepsilon_t^2}{\sigma_\varepsilon^4}\bigg|\, \mathcal{F}_{t-1}\right) = 0,$$

since ε_t has mean zero and variance σ_ε^2.

For part (2) and (3) of the proof, the expressions are evaluated at any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ if not otherwise stated. The data-generating parameters will be explicitly denoted by a nought-subscript. The process y_t is *data* and thus evaluated at $\boldsymbol{\theta}_0$ throughout.

We first consider the gradient vectors of $\boldsymbol{\xi}$,

$$\begin{aligned}
\mathbb{E}\left\|\frac{\partial \ell_t}{\partial \boldsymbol{\xi}}\right\| &= \mathbb{E}\left\|\frac{\varepsilon_t}{\sigma_\varepsilon^2}\frac{\partial}{\partial \boldsymbol{\xi}} g(\boldsymbol{x}_t; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{c})\right\|, \\
&\leq \left(\mathbb{E}\left|\frac{\varepsilon_t}{\sigma_\varepsilon^2}\right|^2\right)^{\frac{1}{2}} \left(\mathbb{E}\left\|\frac{\partial}{\partial \boldsymbol{\xi}} g(\boldsymbol{x}_t; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{c})\right\|^2\right)^{\frac{1}{2}}, \\
&\leq \left(\frac{\mathbb{E}\varepsilon_t^2}{c}\right)^{\frac{1}{2}} \left(\mathbb{E}\left\|\frac{\partial}{\partial \boldsymbol{\xi}} g(\boldsymbol{x}_t; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{c})\right\|^2\right)^{\frac{1}{2}} < \infty.
\end{aligned}$$

The finiteness of the second factor follows from Lemma 2 (1). For the first factor, note that

$$\begin{aligned}
\varepsilon_t^2 &= \left(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}_0 - \sum_{m=1}^{M} \boldsymbol{x}_t'\boldsymbol{\beta}_m f[\gamma_m(t - c_m)]\right)^2, \\
&= \left(\boldsymbol{x}_t'(\boldsymbol{\beta}_{0,0} - \boldsymbol{\beta}_0) + \sum_{m=1}^{M} \boldsymbol{x}_t'\left[\boldsymbol{\beta}_{m,0} f\left(\gamma_{m,0}(t - c_{m,0})\right) - \boldsymbol{\beta}_m f\left(\gamma_m(t - c_m)\right)\right]\right)^2.
\end{aligned}$$

Therefore, there exists $K \in \mathbb{N}$ such that

$$\begin{aligned}
\varepsilon_t^2 &\leq K\left|\boldsymbol{x}_t'(\boldsymbol{\beta}_{0,0} - \boldsymbol{\beta}_0)\right|^2 + K\sum_{m=1}^{M}\left|\boldsymbol{x}_t'\left(\boldsymbol{\beta}_{m,0} f[\gamma_{m,0}(t - c_{m,0})] - \boldsymbol{\beta}_m f[\gamma_m(t - c_m)]\right)\right|^2, \\
&\leq KL\left\|\boldsymbol{x}_t\right\|^2 + KL\sum_{m=1}^{M}\left\|\boldsymbol{x}_t\right\|^2, \\
&= KL(M+1)\left\|\boldsymbol{x}_t\right\|^2,
\end{aligned}$$

where L is some positive constant. The existence of such L is guaranteed by the compactness of the parameter space and the fact that f is bounded. Using Assumption 3 (2), it is clear that $\mathbb{E}\varepsilon_t^2$ is bounded.

For σ_ε^2,

$$\begin{aligned}\mathbb{E}\left|\frac{\partial \ell_t}{\partial \sigma_\varepsilon^2}\right| &= \mathbb{E}\left|\frac{1}{2\sigma_\varepsilon^2} - \frac{1}{2}\frac{\varepsilon_t^2}{\sigma_\varepsilon^4}\right|, \\ &\leq \frac{1}{2\sigma_\varepsilon^2} + \frac{1}{2}\mathbb{E}\left|\frac{\varepsilon_t^2}{\sigma_\varepsilon^4}\right|, \\ &= \frac{1}{\sigma_\varepsilon^2} < \infty.\end{aligned}$$

This shows statement (2) of Lemma 3. Statement (3) use similar techniques in the proof. We will only show the case of γ_m, which requires most work. The rest of the proof will be omitted for brevity.

$$\begin{aligned}\mathbb{E}\left|\frac{\partial \ell_t}{\partial \gamma_m}\frac{\partial \ell_t}{\partial \gamma_m'}\right| &= \mathbb{E}\left|\frac{\varepsilon_t^2}{\sigma_\varepsilon^4}\left(\frac{\partial f}{\partial \gamma_m}\right)^2 \boldsymbol{x}_t'\boldsymbol{\beta}_m\boldsymbol{\beta}_m'\boldsymbol{x}_t\right|, \\ &\leq \left(\mathbb{E}\left|\frac{\varepsilon_t^2}{\sigma_\varepsilon^4}\right|^2\right)^{\frac{1}{2}}\left(\mathbb{E}\left|\boldsymbol{x}_t'\boldsymbol{\beta}_m\boldsymbol{\beta}_m'\boldsymbol{x}_t\right|^2\right)^{\frac{1}{2}}\left|\frac{\partial f}{\partial \gamma_m}\right|^2, \\ &\leq \left(\frac{\mathbb{E}\varepsilon_t^4}{c^3}\right)^{\frac{1}{2}}\left(\mathbb{E}\left\|\boldsymbol{x}_t\right\|^4\left\|\boldsymbol{\beta}_m\right\|^4\right)^{\frac{1}{2}}\left|\frac{\partial f}{\partial \gamma_m}\right|^2 < \infty.\end{aligned}$$

The finiteness of $\mathbb{E}\left\|\boldsymbol{x}_t\right\|^4$ follows from Assumption 3 (2). $\left\|\boldsymbol{\beta}_m\right\|^4$ is finite due to Assumption 1. Lemma 1 ensures that the last factor is bounded. To see the finiteness of the first factor, recall in part (2) we have shown that

$$\varepsilon_t^2 \leq KL(M+1)\left\|\boldsymbol{x}_t\right\|^2.$$

It follows that

$$\varepsilon_t^4 \leq (KL)^2(M+1)^2\left\|\boldsymbol{x}_t\right\|^4.$$

Therefore,

$$\mathbb{E}\varepsilon_t^4 \leq (KL)^2(M+1)^2\mathbb{E}\left\|\boldsymbol{x}_t\right\|^4 < \infty$$

by Assumption 3 □

LEMMA 4. *The function*

$$g_t(\boldsymbol{\theta}) := -\frac{\partial^2 \ell_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \boldsymbol{A}(\boldsymbol{\theta})$$

where

$$\boldsymbol{A}(\boldsymbol{\theta}) = -\mathbb{E}\frac{\partial^2 \ell_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}$$

is absolutely uniformly integrable:

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|g_t(\boldsymbol{\theta})\right\| < \infty;$$

it is continuous in $\boldsymbol{\theta}$ and has zero mean: $\mathbb{E}g_t(\boldsymbol{\theta}) = \mathbf{0}$.

Proof. ¿From the triangular inequality,

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|g_t(\boldsymbol{\theta})\right\| \leq \mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{\partial^2 \ell_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\| + \mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{A}(\boldsymbol{\theta})\right\|.$$

If $\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\partial^2 \ell_t/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'\right\| < \infty$, $\boldsymbol{A}(\boldsymbol{\theta})$ exists and by the Ergodic Theorem, there is pointwise convergence. Thus showing absolute uniform integrability reduces to showing that

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\frac{\partial^2 \ell_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\| < \infty.$$

Proving finiteness of the expected value of the supremum consists of repeated application of the Lebesgue Dominated Convergence Theorem (Shiryaev (1996, p. 187), Ling and McAleer (2003), Lemmas 5.3 and 5.4). We will show the statement for second derivatives element by element, starting with $\boldsymbol{\beta}_{\mathbf{0}}$,

$$\frac{\partial^2 \ell_t}{\partial\boldsymbol{\beta}_0\partial\boldsymbol{\beta}_0'} = -\frac{\boldsymbol{x}_t\boldsymbol{x}_t'}{\sigma_\varepsilon^2}.$$

According to Assumption 2 (1) there exists a constant c such that $\sigma_\varepsilon^2 > c > 0$, therefore

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\frac{\partial^2 \ell_t}{\partial\boldsymbol{\beta}_0\partial\boldsymbol{\beta}_0'}\right\| \le \left\|\frac{\boldsymbol{x}_t\boldsymbol{x}_t'}{c}\right\|.$$

By Assumption 3 (3),

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\frac{\partial^2 \ell_t}{\partial\boldsymbol{\beta}_0\partial\boldsymbol{\beta}_0'}\right\| \le \mathbb{E}\left\|\frac{\boldsymbol{x}_t\boldsymbol{x}_t'}{c}\right\| < \infty.$$

For $\boldsymbol{\beta}_m, m = 1, 2, \ldots, M$,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\frac{\partial^2 \ell_t}{\partial\boldsymbol{\beta}_m\partial\boldsymbol{\beta}_m'}\right\| = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\frac{\boldsymbol{x}_t\boldsymbol{x}_t' f^2}{\sigma_\varepsilon^2}\right\| \le \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\frac{\boldsymbol{x}_t\boldsymbol{x}_t' f^2}{c}\right\| \le \left\|\frac{\boldsymbol{x}_t\boldsymbol{x}_t'}{c}\right\|.$$

The last inequality follows from the fact that $|f| \le 1$. Therefore,

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\frac{\partial^2 \ell_t}{\partial\boldsymbol{\beta}_m\partial\boldsymbol{\beta}_m'}\right\| \le \frac{\mathbb{E}\left\|\boldsymbol{x}_t\boldsymbol{x}_t'\right\|}{c} < \infty.$$

We next examine the second derivatives of the log likelihood with respect to σ_ε^2,

$$\left|\frac{\partial^2 \ell_t}{\partial(\sigma_\varepsilon^2)^2}\right| = \left|\frac{1}{2\sigma_\varepsilon^4} - \frac{\varepsilon_t^2}{\sigma_\varepsilon^6}\right| \le \left|\frac{1}{2\sigma_\varepsilon^4}\right| + \left|\frac{\varepsilon_t^2}{\sigma_\varepsilon^6}\right|,$$

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\frac{\partial^2 \ell_t}{\partial(\sigma_\varepsilon^2)^2}\right| \le \frac{1}{2c^2} + \frac{1}{c^3}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \varepsilon_t^2.$$

In order to show $\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\frac{\partial^2 \ell_t}{\partial(\sigma_\varepsilon^2)^2}\right| < \infty$, it is sufficient to show that $\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}(\varepsilon_t^2) < \infty$. Recall we have already proved in Lemma 3 (2) that

$$\varepsilon_t^2 \le KL(M+1)\left\|\boldsymbol{x}_t\right\|^2.$$

It follows that

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}(\varepsilon_t^2) \le KL(M+1)\mathbb{E}\left\|\boldsymbol{x}_t\right\|^2 < \infty.$$

To show that $\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{\partial^2\ell_t}{\partial\gamma_i^2}\right| < \infty$, consider

$$\begin{aligned}
\left|\frac{\partial^2\ell_t}{\partial\gamma_m^2}\right| &= \left|\frac{-\left(\boldsymbol{x}_t'\boldsymbol{\beta}_m\frac{\partial f}{\partial\gamma_m}\right)^2 + \varepsilon_t\left(\boldsymbol{x}_t'\boldsymbol{\beta}_m\frac{\partial^2 f}{\partial\gamma_m^2}\right)}{\sigma_\varepsilon^2}\right|, \\
&\leq \frac{1}{c}\left(\frac{\partial f}{\partial\gamma_m}\right)^2\left|\boldsymbol{x}_t'\boldsymbol{\beta}_m\right|^2 + \frac{1}{c}\left|\frac{\partial^2 f}{\partial\gamma_m^2}\right|\left|\varepsilon_t\right|\left|\boldsymbol{x}_t'\boldsymbol{\beta}_m\right|, \\
&\leq \frac{L}{c}\left(\frac{\partial f}{\partial\gamma_m}\right)^2\left\|\boldsymbol{x}_t\right\|^2 + \frac{1}{c}\left|\frac{\partial^2 f}{\partial\gamma_m^2}\right|\left|\varepsilon_t\right|\left|\boldsymbol{x}_t'\boldsymbol{\beta}_m\right|,
\end{aligned}$$

where L is some positive constant. The second term on the right side can be written as

$$\begin{aligned}
\frac{1}{c}\left|\frac{\partial^2 f}{\partial\gamma_m^2}\right|\left|\varepsilon_t\right|\left|(\boldsymbol{x}_t'\boldsymbol{\beta}_m)\right| &= \frac{1}{c}\left|\frac{\partial^2 f}{\partial\gamma_m^2}\right|\left|\boldsymbol{x}_t'(\boldsymbol{\beta}_{0,0}-\boldsymbol{\beta}_0) + \sum_{m=1}^{M}\boldsymbol{x}_t'(\boldsymbol{\beta}_{m,0}f_{m,0}-\boldsymbol{\beta}_m f_m)\right|\left|\boldsymbol{x}_t'\boldsymbol{\beta}_m\right|, \\
&= \frac{1}{c}\left|\frac{\partial^2 f}{\partial\gamma_m^2}\right|\left|\boldsymbol{x}_t'(\boldsymbol{\beta}_{0,0}-\boldsymbol{\beta}_0)\right|\left|\boldsymbol{x}_t'\boldsymbol{\beta}_m\right| + \left|\sum_{m=1}^{M}\boldsymbol{x}_t'(\boldsymbol{\beta}_{m,0}f_{m,0}-\boldsymbol{\beta}_m f_m)\right|\left|\boldsymbol{x}_t'\boldsymbol{\beta}_m\right|, \\
&\leq \frac{1}{c}\left|\frac{\partial^2 f}{\partial\gamma_m^2}\right| K\left\|\boldsymbol{x}_t\right\|^2,
\end{aligned}$$

where K is some positive constant. Again, the compactness of the parameter space, boundedness of f, and stationarity of $\boldsymbol{x}_t$ ensures the existence of K and L. It follows that

$$\left|\frac{\partial^2\ell_t}{\partial\gamma_i^2}\right| \leq \left(\frac{L}{c}\left(\frac{\partial f}{\partial\gamma_i}\right)^2 + \frac{1}{c}\left|\frac{\partial^2 f}{\partial\gamma_i^2}\right| K\right)\left\|\boldsymbol{x}_t\right\|^2.$$

The finiteness of the derivatives of f was shown in Lemma 1. Thus,

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{\partial^2\ell_t}{\partial\gamma_i^2}\right| \leq \left(\frac{L}{c}\left(\frac{\partial f}{\partial\gamma_i}\right)^2 + \frac{1}{c}\left|\frac{\partial^2 f}{\partial\gamma_i^2}\right| K\right)\mathbb{E}\left\|\boldsymbol{x}_t\right\|^2 < \infty.$$

The proof that $\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{\partial^2\ell_t}{\partial c_i^2}\right| < \infty$ closely resembles the proof above and is omitted for brevity.

□

Proof of Theorem 2. The proof establishes the conditions of Theorem 4.1.3 of Amemiya (1985) with a generalization due to Ling and McAleer (2003, Theorem 3.1). We need consistency of $\widehat{\boldsymbol{\theta}}_T$ for $\boldsymbol{\theta}_0$, which was shown in Theorem 1. Then we show

$$\boldsymbol{B}(\boldsymbol{\theta}_0)^{-\frac{1}{2}}\frac{1}{\sqrt{T}}\sum_{t=1}^{[rT]}\left.\frac{\partial\ell_t}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}_0} \xrightarrow{d} W(s),\ s\in[0,1],$$

where $W(r)$ is N-dimensional standard Brownian motion on the unit interval. This is condition (C) in Theorem 4.1.3 of Amemiya (1985). The convergence follows from Theorem 18.3 in Billingsley (1999) if (a) $\left\{\left.\frac{\partial\ell_t}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}_0}, \mathcal{F}_t\right\}$ is a stationary martingale difference, and (b) $\boldsymbol{B}(\boldsymbol{\theta}_0)$ exists. Both conditions were shown in Lemma 3.

To satisfy condition (B) of Theorem 4.1.3 of Amemiya (1985), we have to establish

$$\boldsymbol{A}_T(\boldsymbol{\theta}_T^*) \xrightarrow{p} \boldsymbol{A}(\boldsymbol{\theta}_0)$$

for any sequence $\boldsymbol{\theta}_T^* \xrightarrow{p} \boldsymbol{\theta}_0$,

$$\boldsymbol{A}_T(\boldsymbol{\theta}_T^*) = -\frac{1}{T}\sum_{t=1}^{T} \left.\frac{\partial^2 \ell_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}_T^*},$$

and

$$\boldsymbol{A}(\boldsymbol{\theta}_0) = -\mathbb{E} \left.\frac{\partial^2 \ell_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}_0}$$

is non-singular. Conditions for the double stochastic convergence can be found in Theorem 21.6 of Davidson (1994). We need to show

(1) consistency of $\widehat{\boldsymbol{\theta}}_T$ for $\boldsymbol{\theta}_0$ (Theorem 1), and
(2) uniform convergence of $\boldsymbol{A}_T$ to $\boldsymbol{A}$ in probability, i.e.

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\boldsymbol{A}_T(\boldsymbol{\theta}) - \boldsymbol{A}(\boldsymbol{\theta})| \xrightarrow{p} 0.$$

We prove uniform convergence of $\boldsymbol{A}_T$ using Theorem 3.1 of Ling and McAleer (2003), who generalize Theorem 4.2.1 of Amemiya (1985) from i.i.d. data to stationary and ergodic data. This allows the immediate invocation of the Ergodic Theorem without having to check finiteness of third derivatives of ℓ_t as in Andrews (1992, Theorem 2). To apply Theorem 3.1 of Ling and McAleer (2003) we need that

$$g_t(\boldsymbol{\theta}) = -\frac{\partial^2 \ell_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \boldsymbol{A}(\boldsymbol{\theta})$$

is continuous in $\boldsymbol{\theta}$ (this also establishes condition (A) of Theorem 4.1.3. of Amemiya (1985) along the way), has expected value $\mathbb{E}g_t(\boldsymbol{\theta}) = 0$ and is absolutely uniformly integrable:

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |g_t(\boldsymbol{\theta})| < \infty.$$

This was shown in Lemma 4. Thus, we have established all conditions for asymptotic normality according to Theorem 4.1.3 of Amemiya (1985). □

Proof of Proposition 1. The proof of uniform convergence in probability of $\boldsymbol{A}_T$ to $\boldsymbol{A}$ is given in Lemma 4 and Theorem 2. We need to show uniform convergence of $\boldsymbol{B}_T$ to $\boldsymbol{B}$. We employ Theorem 3.1 of Ling and McAleer (2003) again and show that

$$h_t(\boldsymbol{\theta}) := \frac{\partial \ell_t}{\partial\boldsymbol{\theta}}\frac{\partial \ell_t}{\partial\boldsymbol{\theta}'} - \boldsymbol{B}(\boldsymbol{\theta}),$$

is absolutely uniformly integrable, continuous in $\boldsymbol{\theta}$, and has expected value $\mathbb{E}h_t(\boldsymbol{\theta}) = 0$. The detailed proof is in complete analogy to Lemma 4 and is omitted for brevity. □

REFERENCES

AMEMIYA, T. (1985): *Advanced Econometrics*. Harvard University Press, Cambridge.

ANDREWS, D. (1992): "Generic Uniform Convergence," *Econometric Theory*, 8, 241–257.

ANDREWS, D. W. K., AND C. J. MCDERMOTT (1995): "Nonlinear Econometric Models with Deterministically Trending Variables," *Review of Economic Studies*, 62, 343–360.

AREOSA, W., M. MCALEER, AND M. MEDEIROS (in press): "Moment-Based Estimation of Smooth Transition Regression Models with Endogenous Variables," *Journal of Econometrics*.

BACON, D. W., AND D. G. WATTS (1971): "Estimating the Transition Between Two Intersecting Lines," *Biometrika*, 58, 525–534.

BILLINGSLEY, P. (1999): *Convergence of Probability Measures*. Wiley, New York, second edn.

CHAN, F., M. MCALEER, AND M. C. MEDEIROS (2005): "Structure and Asymptotic Theory for Nonlinear Models with GARCH Errors," Discussion Paper 506, Department of Economics, Pontifical Catholic University of Rio de Janeiro.

CHAN, K. S., AND H. TONG (1986): "On Estimating Thresholds in Autoregressive Models," *Journal of Time Series Analysis*, 7, 179–190.

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

GOLDFELD, S. M., AND R. QUANDT (1972): *Nonlinear Methods in Econometrics*. North Holland, Amsterdam.

GRANGER, C. W. J., AND T. TERÄSVIRTA (1993): *Modelling Nonlinear Economic Relationships*. Oxford University Press, Oxford.

LI, W. K., S. LING, AND M. MCALEER (2002): "Recent Theoretical Results for Time Series Models with GARCH Errors," *Journal of Economic Surveys*, 16, 245–269.

LIN, C. F. J., AND T. TERÄSVIRTA (1994): "Testing the Constancy of Regression Parameters Against Continuous Structural Change," *Journal of Econometrics*, 62, 211–228.

LING, S., AND M. MCALEER (2003): "Asymptotic Theory for a Vector ARMA-GARCH Model," *Econometric Theory*, 19, 280–310.

LUNDBERGH, S., AND T. TERÄSVIRTA (1998): "Modelling Economic High-Frequency Time Series with STAR-STGARCH Models," Working Paper Series in Economics and Finance 291, Stockholm School of Economics.

LUUKKONEN, R., P. SAIKKONEN, AND T. TERÄSVIRTA (1988): "Testing Linearity Against Smooth Transition Autoregressive Models," *Biometrika*, 75, 491–499.

MEDEIROS, M., E. MENDES, AND L. OXLEY (2009): "Cointegrating Smooth Transition Regressions with a Stationary Transition Variable," Discussion paper, Department of Economics, Pontifical Catholic University of Rio de Janeiro.

MEDEIROS, M. C., AND A. VEIGA (2003): "Diagnostic Checking in a Flexible Nonlinear Time Series Model," *Journal of Time Series Analysis*, 24, 461–482.

——— (2005): "A Flexible Coefficient Smooth Transition Time Series Model," *IEEE Transactions on Neural Networks*, 16, 97–113.

MIRA, S., AND A. ESCRIBANO (2000): "Nonlinear Time Series Models: Consistency and Asymptotic Normality of NLS Under New Conditions," in *Nonlinear Econometric Modeling in Time Series Analysis*, ed. by W. A. Barnett, D. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjøsthein, and A. Würtz, pp. 119–164. Cambridge University Press.

SAIKKONEN, P., AND I. CHOI (2004): "Cointegrating Smooth Transition Regressions," *Econometric Theory*, 20, 301–340.

SHIRYAEV, A. (1996): *Probability*. Springer, New York, second edn.

STRIKHOLM, B. (2006): "Determining the number of breaks in a piecewise linear regression model," Working Paper Series in Economics and Finance 648, Stockholm School of Economics.

SUAREZ-FARIÑAS, M., C. E. PEDREIRA, AND M. C. MEDEIROS (2004): "Local Global Neural Networks: A New Approach for Nonlinear Time Series Modeling," *Journal of the American Statistical Association*, 99, 1092–1107.

TERÄSVIRTA, T. (1994): "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association*, 89, 208–218.

——— (1998): "Modelling Economic Relationships with Smooth Transition Regressions," in *Handbook of Applied Economic Statistics*, ed. by A. Ullah, and D. E. A. Giles, pp. 507–552. Dekker.

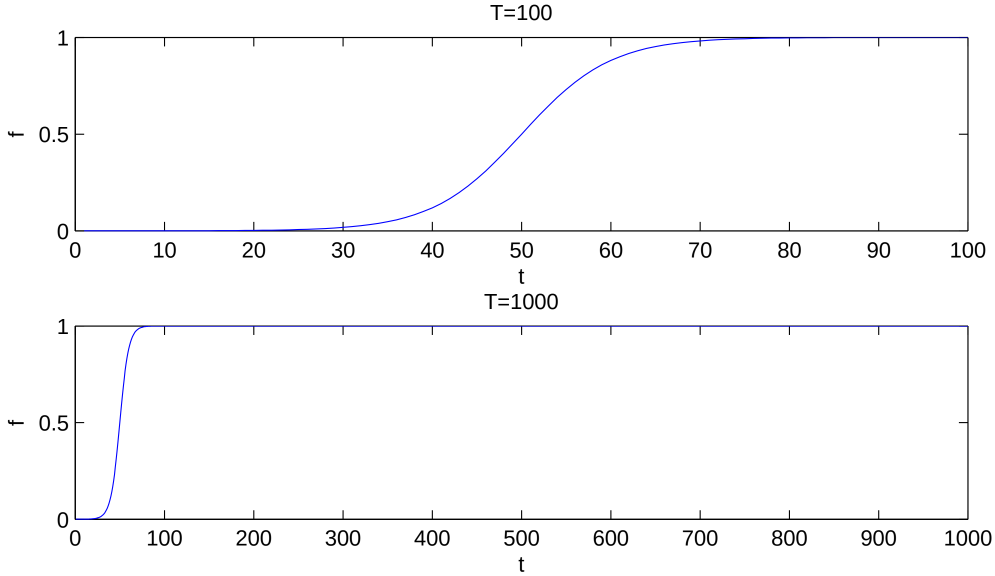

FIGURE 1. Same unscaled logistic transition functions with different sample sizes $T = 100$ & 1000. $\gamma = 0.2; c = 50$.

VAN DIJK, D., T. TERÄSVIRTA, AND P. H. FRANSES (2002): "Smooth Transition Autoregressive Models - A Survey of Recent Developments," *Econometric Reviews*, 21, 1–47.

(E. Hillebrand) DEPARTMENT OF ECONOMICS, LOUISIANA STATE UNIVERSITY, BATON ROUGE, USA.
E-mail address: erhil@lsu.edu

(M. C. Medeiros) DEPARTMENT OF ECONOMICS, PONTIFICAL CATHOLIC UNIVERSITY OF RIO DE JANEIRO, RIO DE JANEIRO, RJ, BRAZIL.
E-mail address: mcm@econ.puc-rio.br

(J. Xu) DEPARTMENT OF ECONOMICS, LOUISIANA STATE UNIVERSITY, BATON ROUGE, USA.
E-mail address: jxu2@lsu.edu

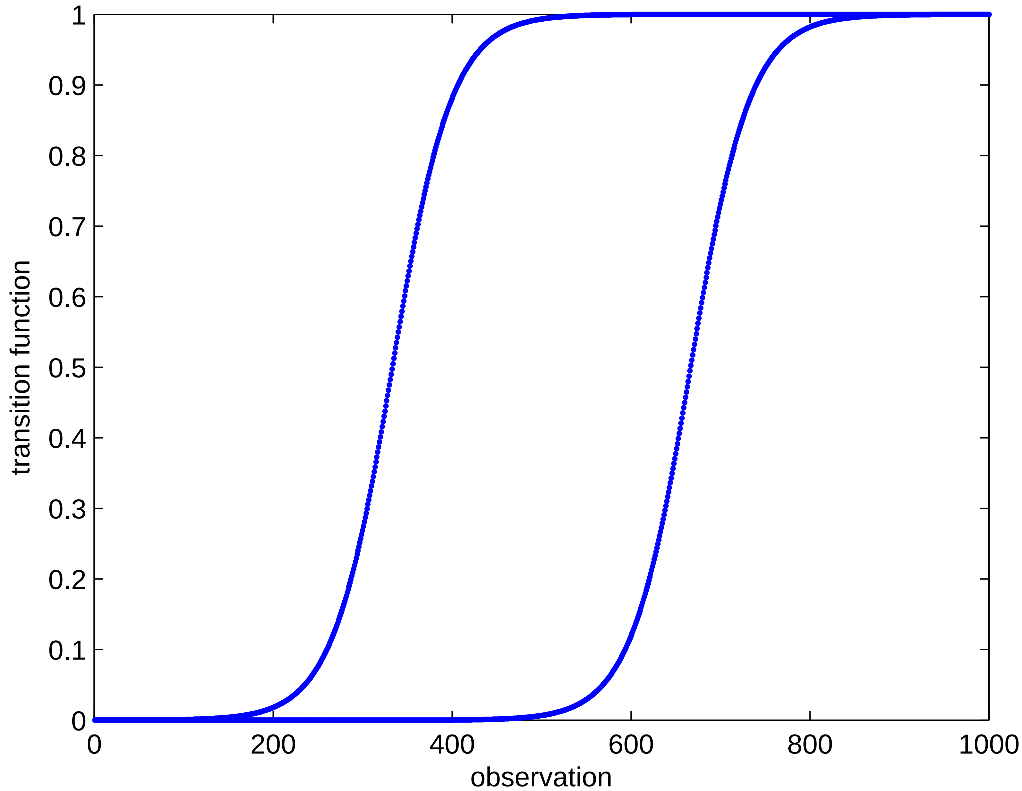

FIGURE 2. Transition function for Models A and B with 1000 observations.

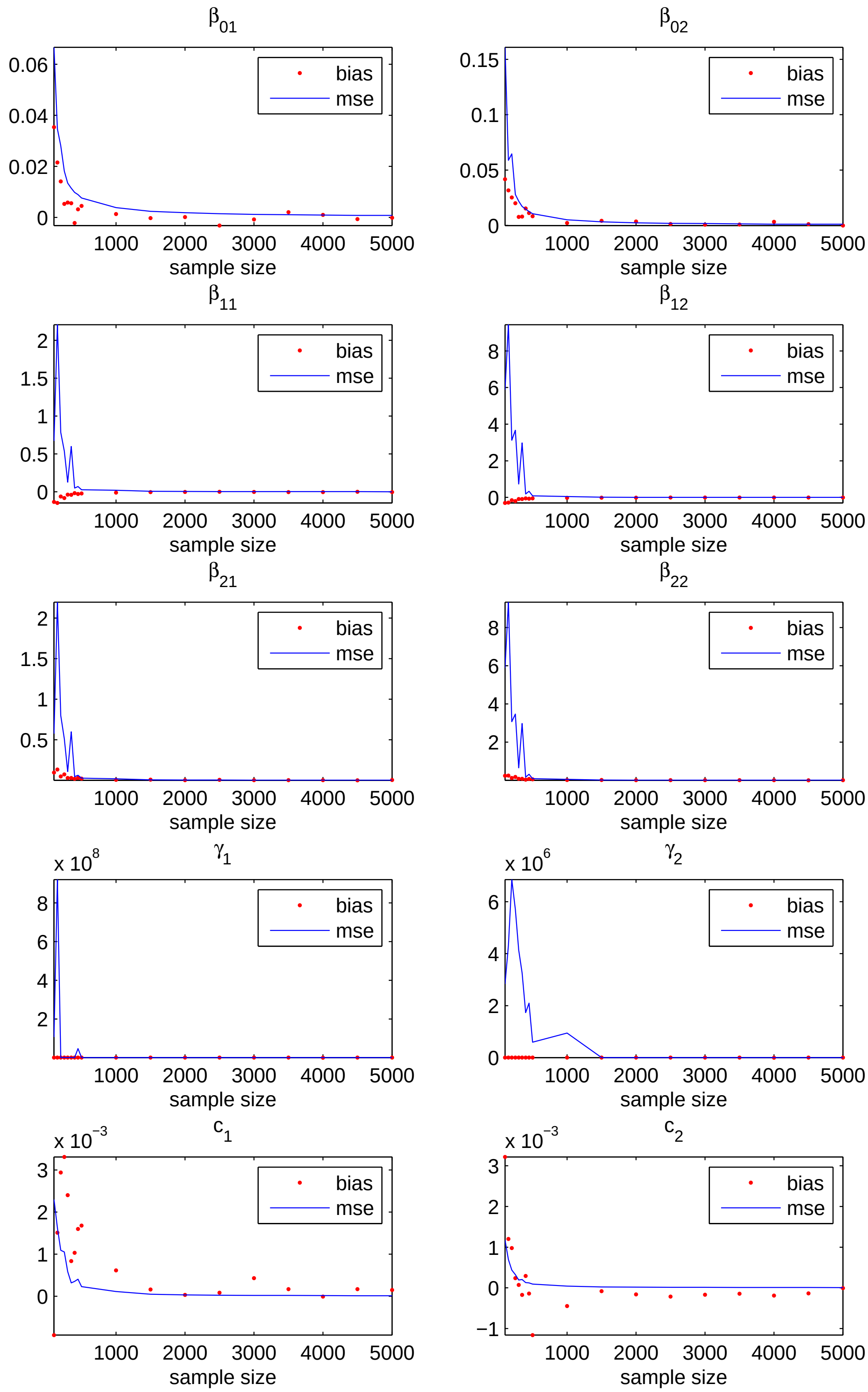

FIGURE 3. Bias and mean squared error (MSE) of the quasi-maximum likelihood estimator of the parameters of Model A with gaussian errors.

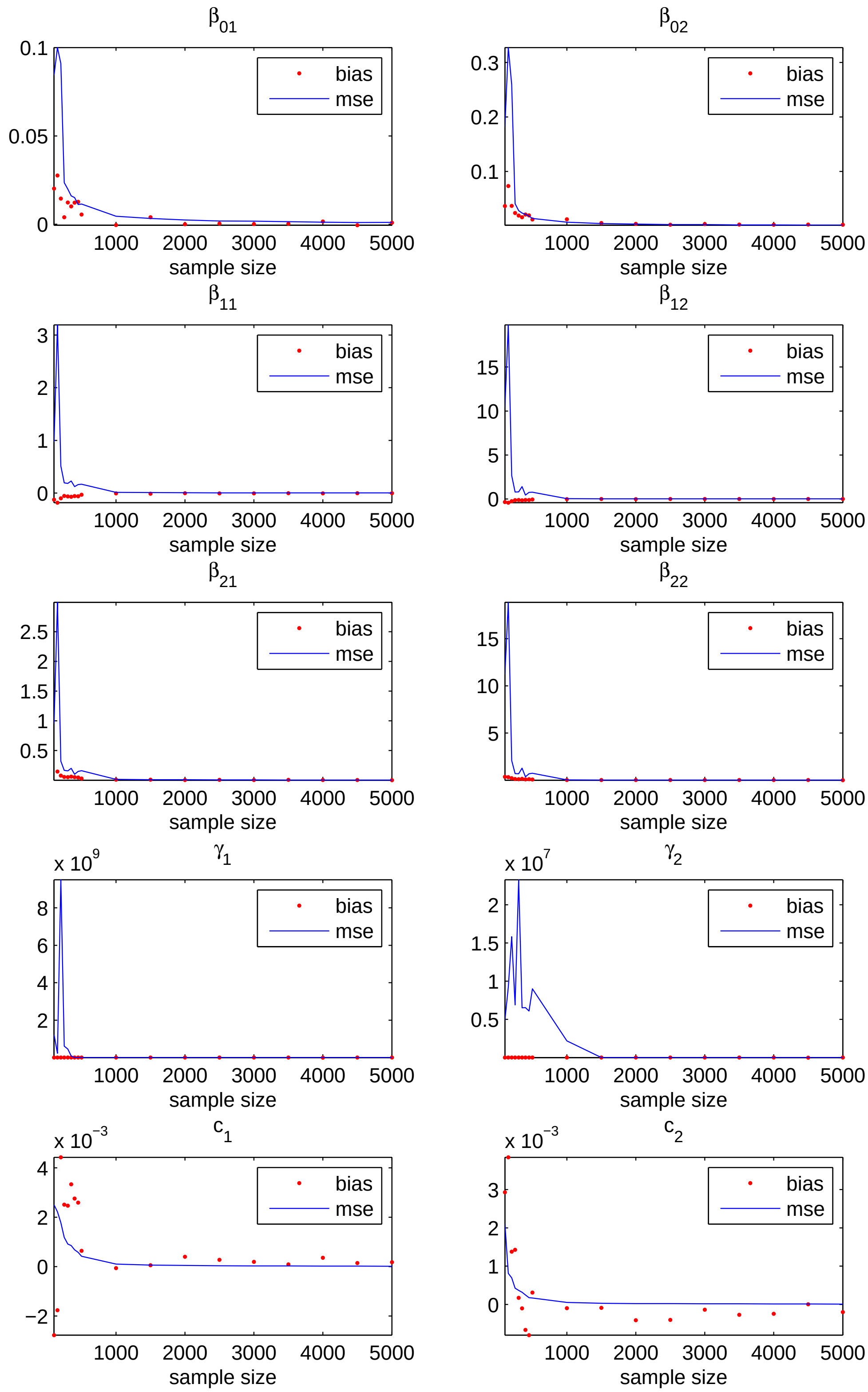

FIGURE 4. Bias and mean squared error (MSE) of the quasi-maximum likelihood estimator of the parameters of Model A with uniform errors.

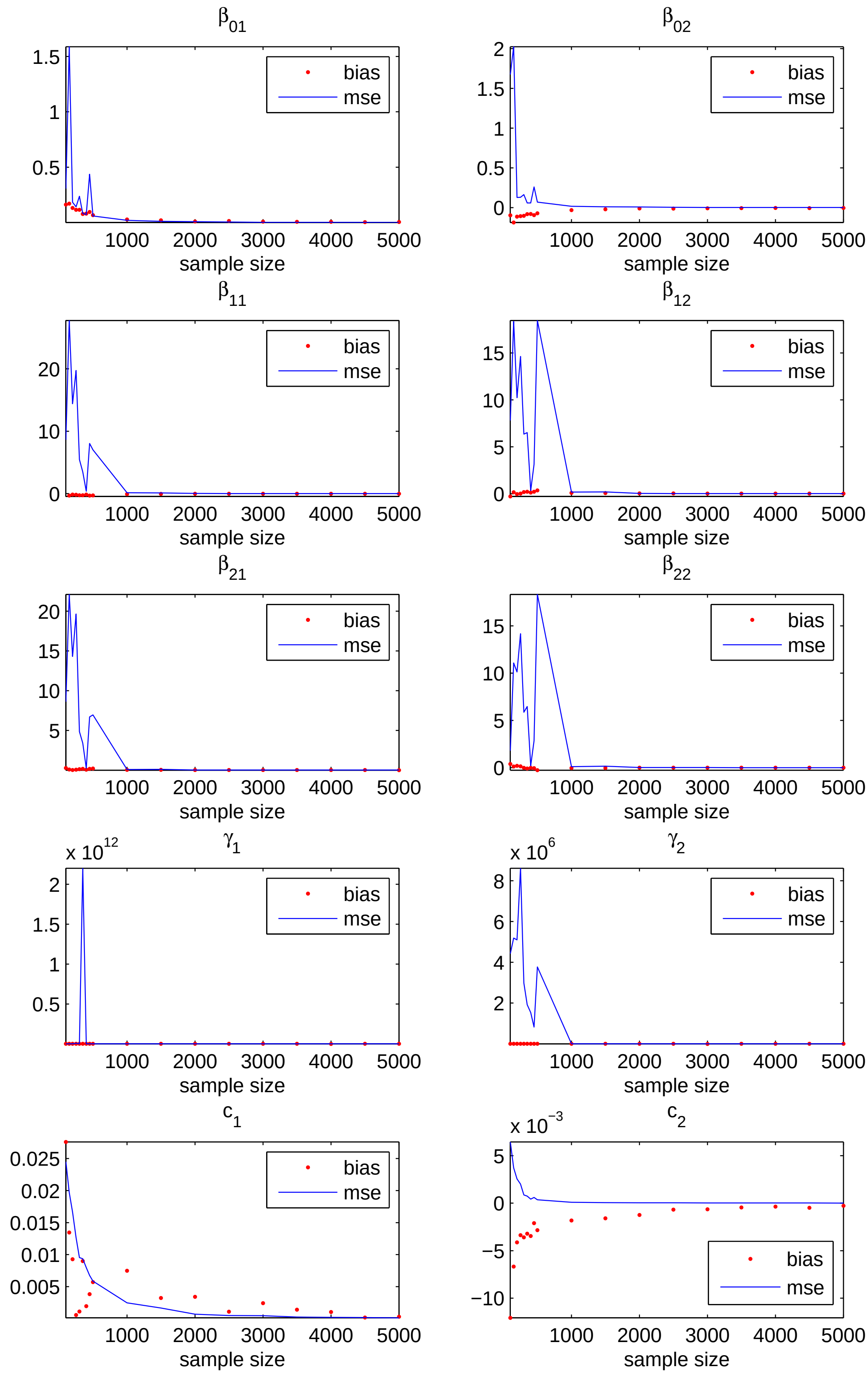

FIGURE 5. Bias and mean squared error (MSE) of the quasi-maximum likelihood estimator of the parameters of Model B with gaussian errors.

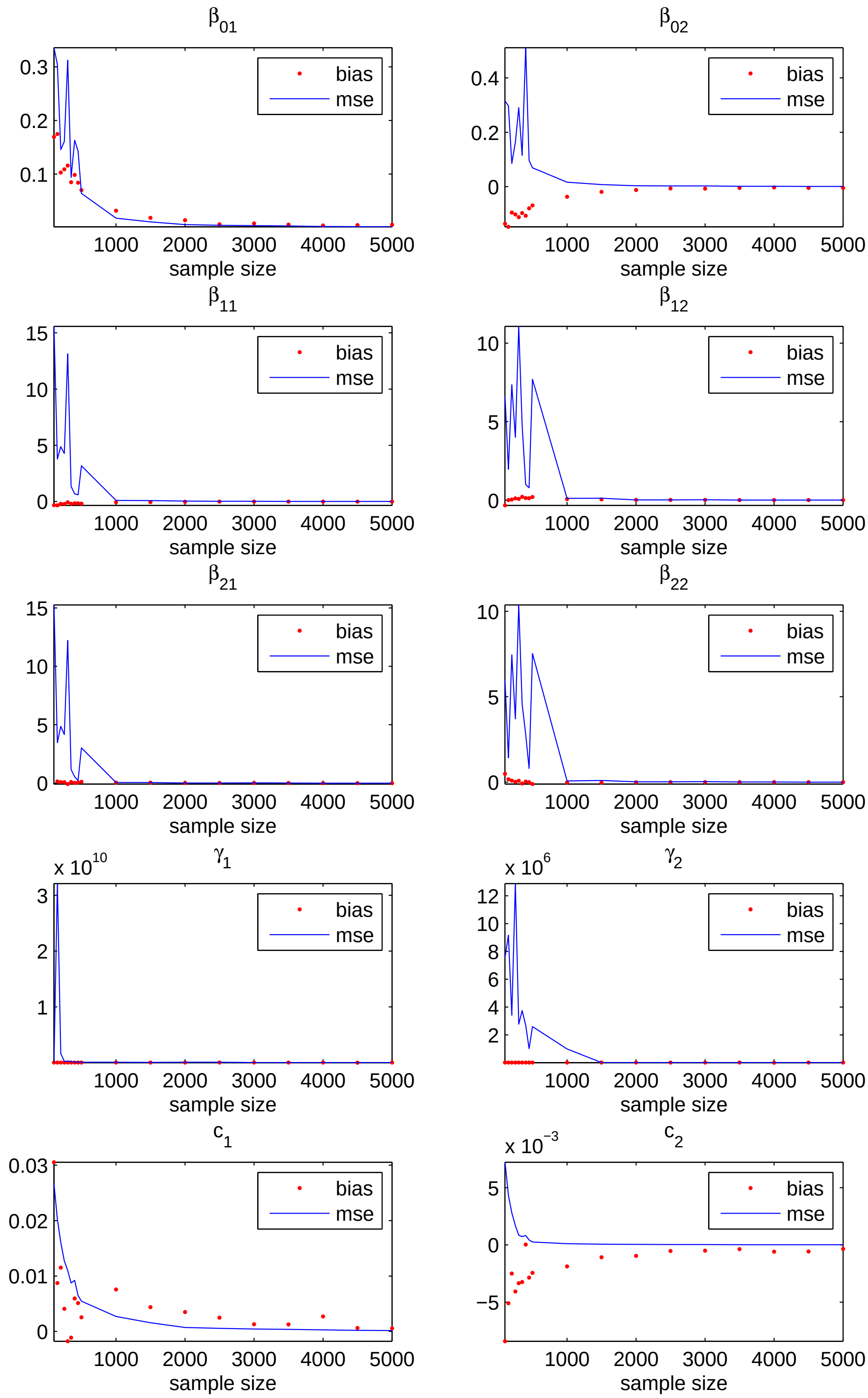

FIGURE 6. Bias and mean squared error (MSE) of the quasi-maximum likelihood estimator of the parameters of Model B with uniform errors.

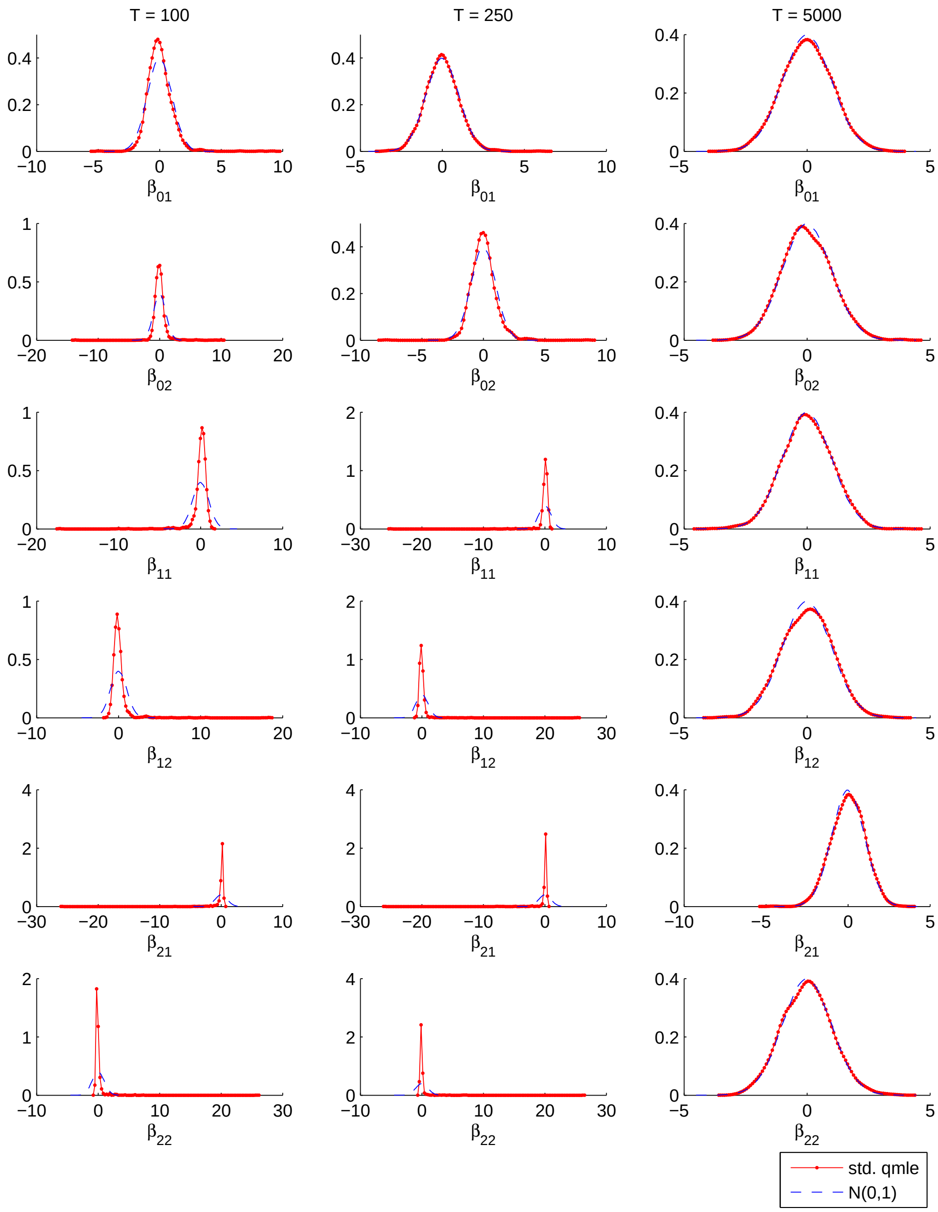

FIGURE 7. Distribution of the standardized QMLE of the linear parameters of Model A with gaussian errors.

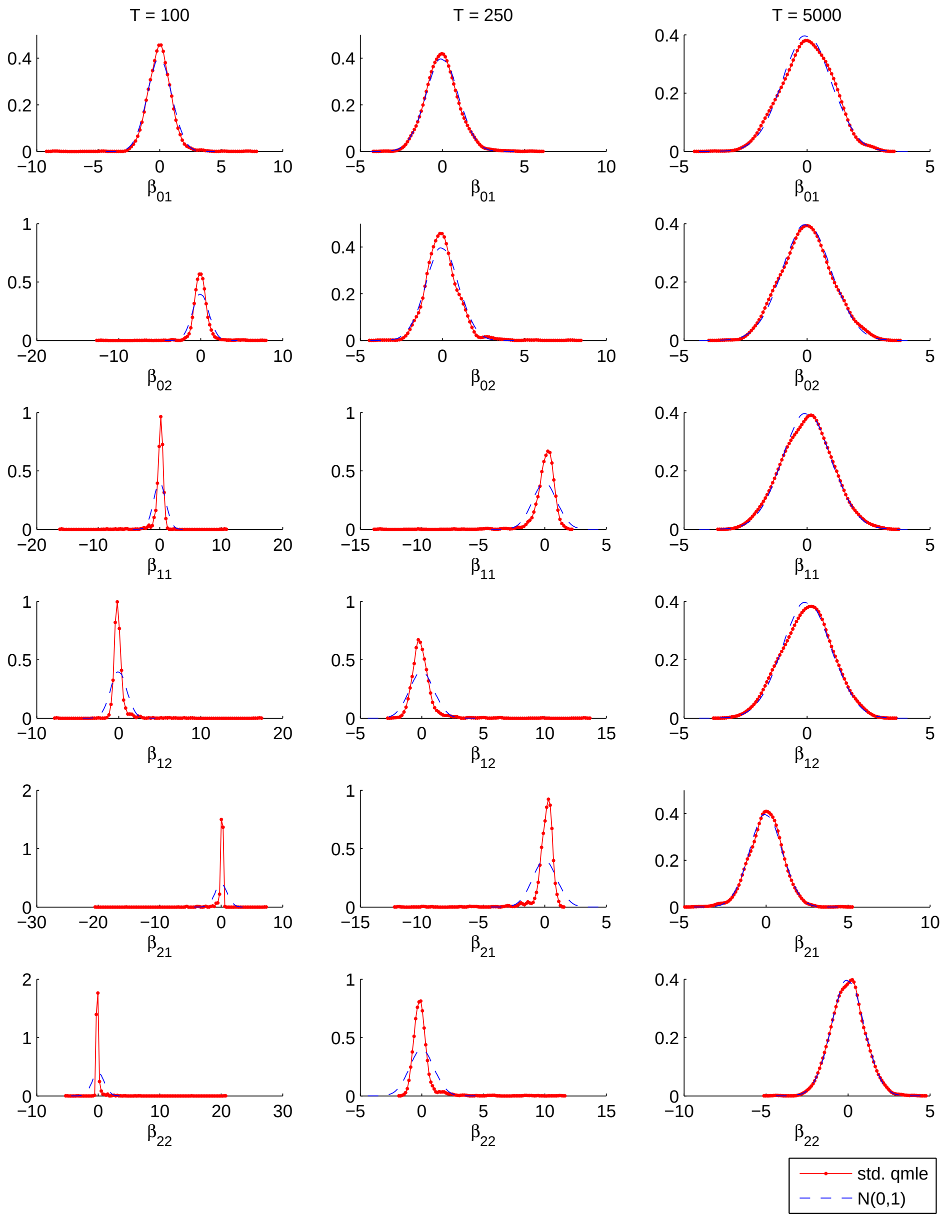

FIGURE 8. Distribution of the standardized QMLE of the linear parameters of Model A with uniform errors.

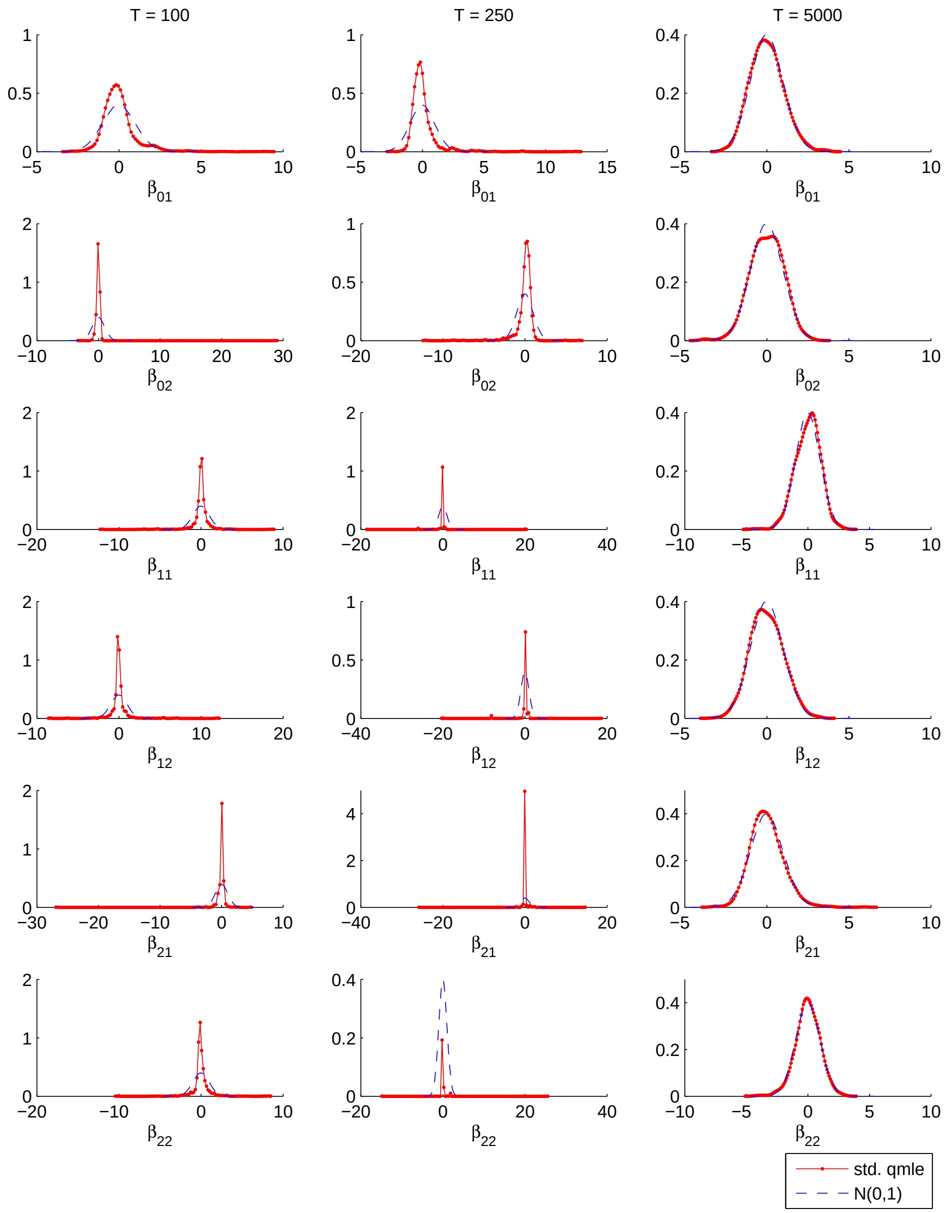

FIGURE 9. Distribution of the standardized QMLE of the linear parameters of Model B with gaussian errors.

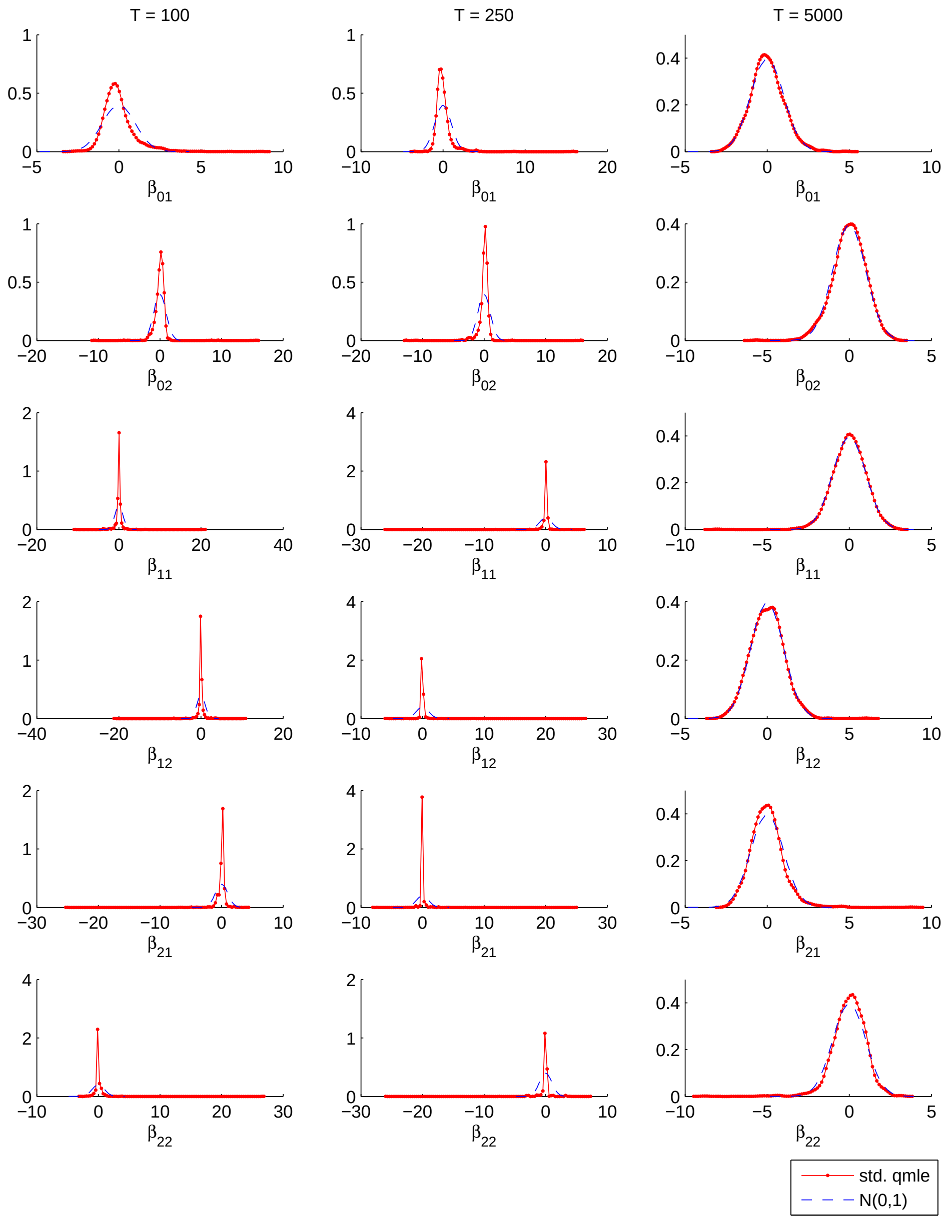

FIGURE 10. Distribution of the standardized QMLE of the linear parameters of Model B with uniform errors.

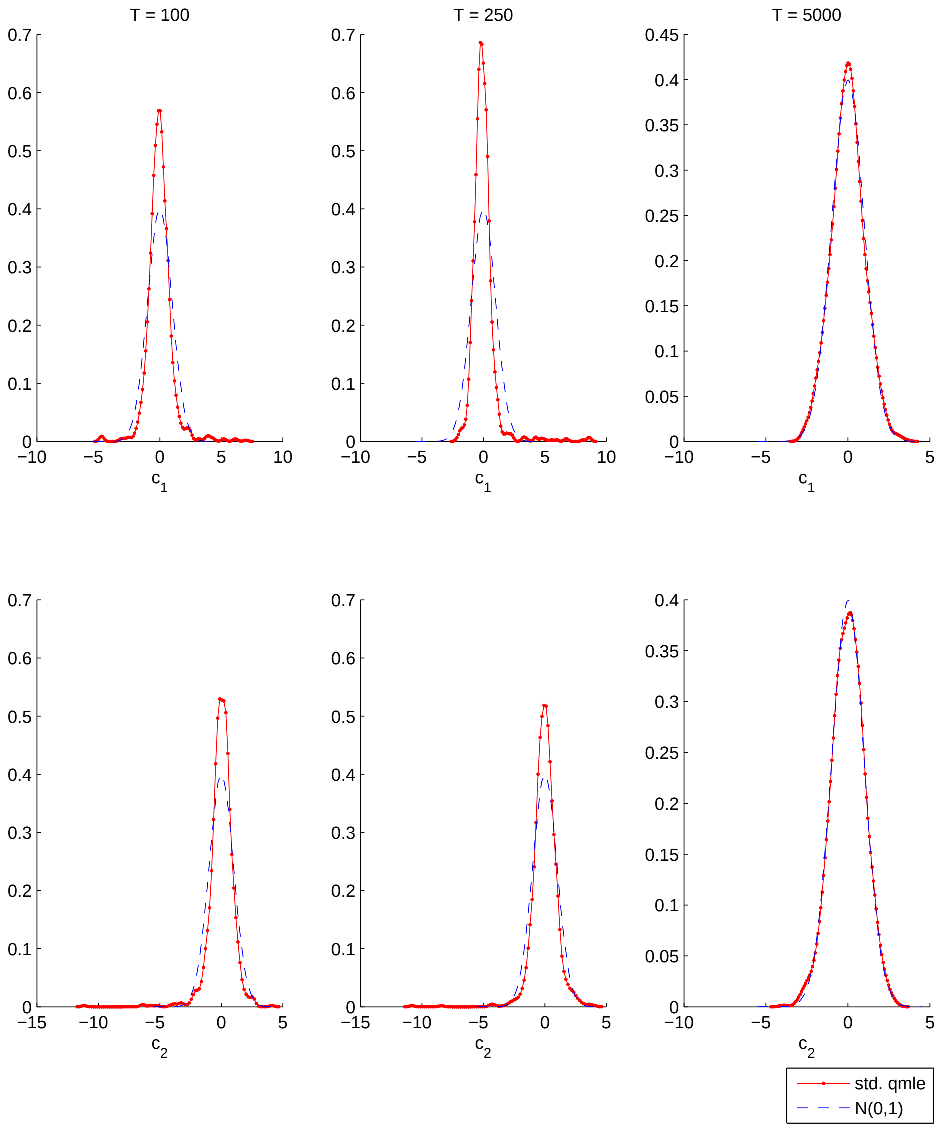

FIGURE 11. Distribution of the standardized QMLE of the location parameters for Model A with gaussian errors.

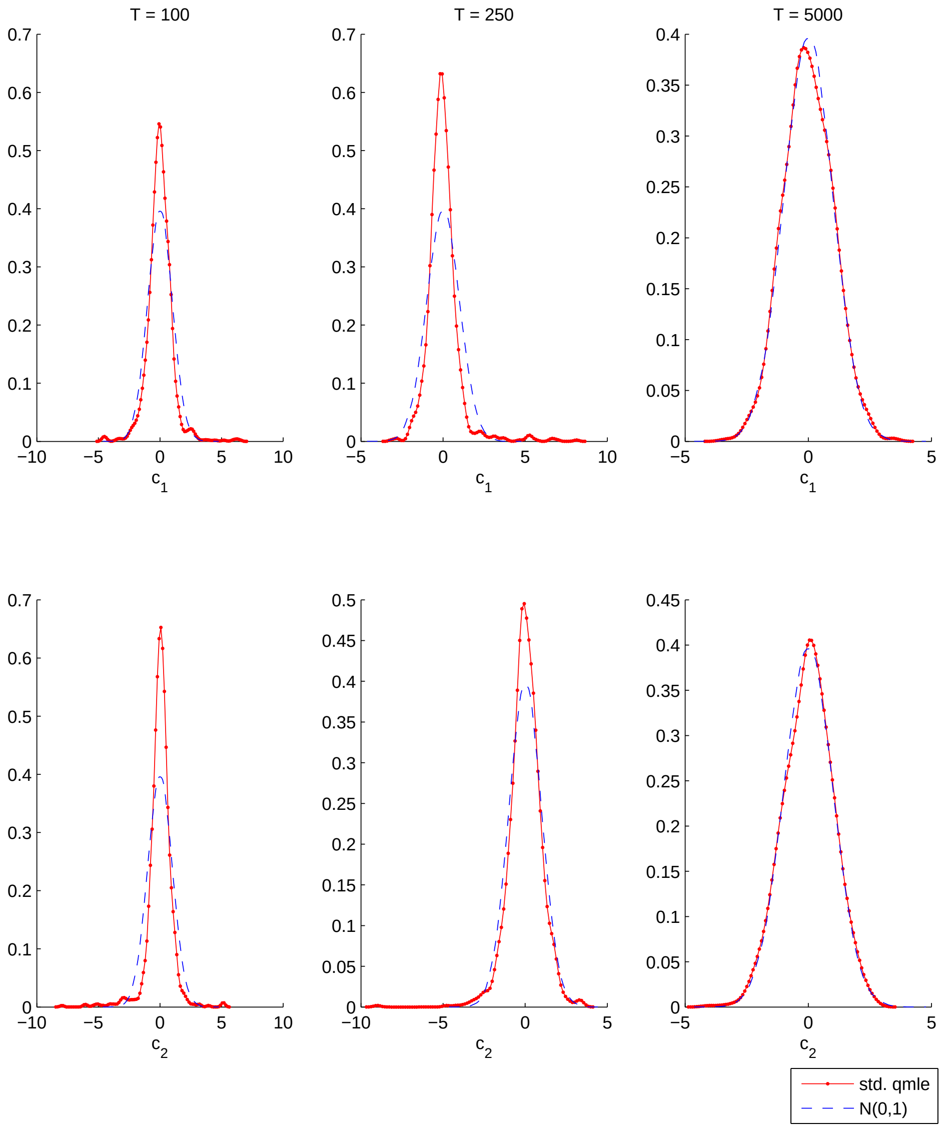

FIGURE 12. Distribution of the standardized QMLE of the location parameters for Model A with uniform errors.

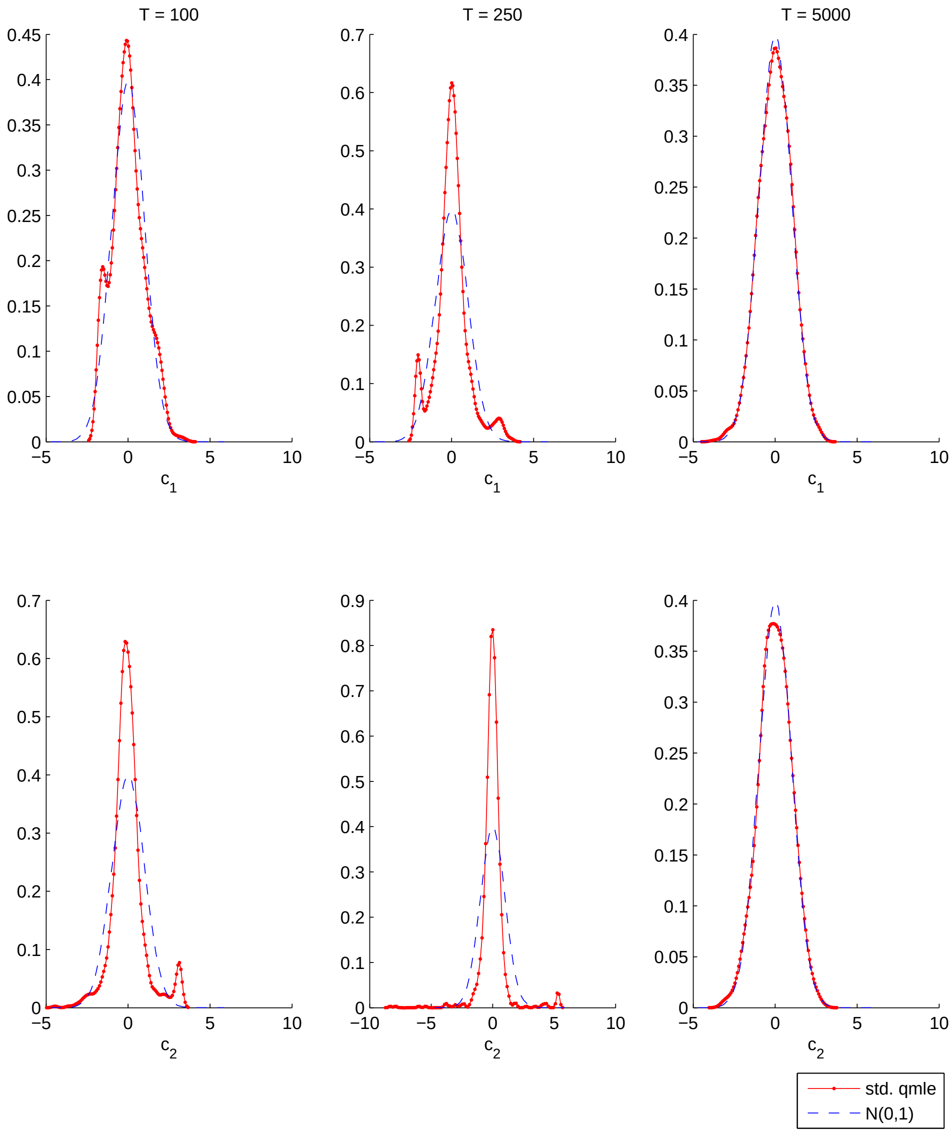

FIGURE 13. Distribution of the standardized QMLE of the location parameters for Model B with gaussian errors.

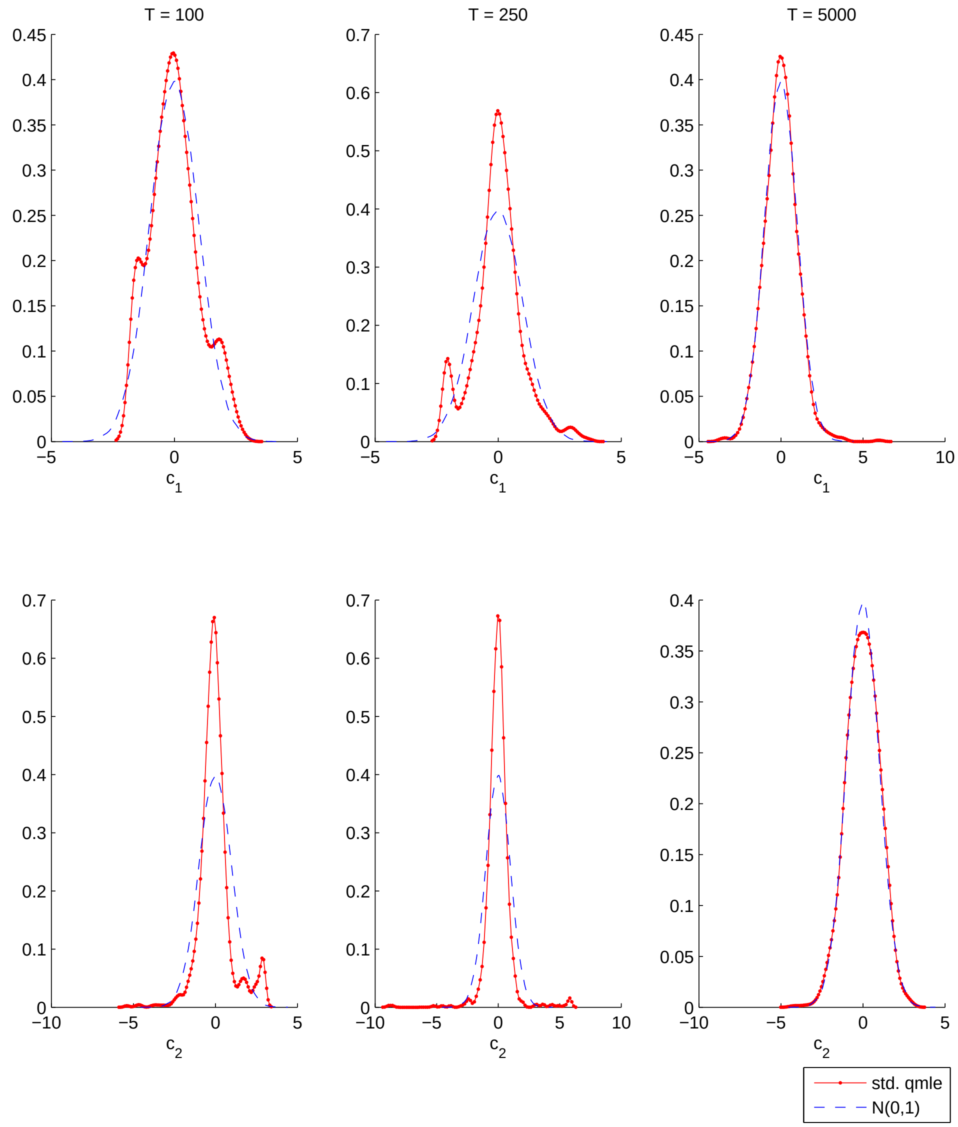

FIGURE 14. Distribution of the standardized QMLE of the location parameters for Model B with uniform errors.