Estimation and Inference under Weak Identification and Persistence: An Application on Forecast-Based Monetary Policy Reaction Function

Ke-Li Xu*and Jui-Chung Yang†

Texas A&M University

November, 2013

Abstract

The reaction coefficients of expected inflations and output gaps in the forecast-based monetary policy reaction function may be merely weakly identified when the smoothing coefficient is close to unity and the nominal interest rates are highly persistent. In this paper we modify the method of Andrews and Cheng (2012, *Econometrica*) on inference under weak / semi-strong identification to accommodate the persistence issue. Our modification involves the employment of asymptotic theories for near unit root processes and novel drifting sequence approaches. Large sample properties with a desired smooth transition with respect to the true values of parameters are developed for the nonlinear least squares (NLS) estimator and its corresponding t / Wald statistics of a general class of models.

Despite the not-consistent-estimability, the conservative confidence sets of weakly-identified parameters of interest can be obtained by inverting the t / Wald tests. We show that the null-imposed least-favorable confidence sets will have correct asymptotic sizes, and the projection-based method may lead to asymptotic over-coverage. Our empirical application suggests that the NLS estimates for the reaction coefficients in U.S.'s forecast-based monetary policy reaction function for 1987:3–2007:4 are not accurate sufficiently to rule out the possibility of indeterminacy.

*Department of Economics, Texas A&M University, 4228 TAMU, College Station, TX 77843-4228, USA. *Email:* `keli.xu@econmail.tamu.edu`

†Department of Economics, Texas A&M University, 4228 TAMU, College Station, TX 77843-4228, USA. *Email:* `jyang@econmail.tamu.edu`

Keywords: asymptotic size, confidence set, monetary policy reaction function, near unit root, real-time data, weak identification.

JEL classification: C12, C22, E58.

1 Introduction

Consider the monetary policy reaction function ($MPRF$, Clarida, Galí and Gertler, 2000):

$$i_t = \rho i_{t-1} + (1-\rho)\left(\pi_\alpha + \pi_{\dot{p}}\mathbb{E}_t\dot{p}_{t,k} + \pi_x\mathbb{E}_t x_{t,k}\right) + \varepsilon_t, \tag{1.1}$$

where the nominal interest rate i_t is modeled as a weighted average of nominal interest rate in the previous period i_{t-1} and the monetary authority's target rate i_t^*. The target rate i_t^* follows a forward-looking Taylor monetary policy rule (Taylor, 1993; Clarida *et al.*, 2000):

$$i_t^* = \pi_\alpha + \pi_{\dot{p}}\mathbb{E}_t\dot{p}_{t,k} + \pi_x\mathbb{E}_t x_{t,k},$$

where $\dot{p}_{t,k}$ and $x_{t,k}$ denote the annualized inflation and the average output gap between periods t and $t+k$. $\mathbb{E}_t(\cdot)$ denotes the expectation of the monetary authority at time t. $\rho \in [0,1)$ is known as the smoothing coefficient. $\{\pi_{\dot{p}}, \pi_x\}$ are the reaction coefficients. The model is called the forecast-based $MPRF$ when the real-time data, *i.e.*, the historical *ex ante* forecasts ($\{\mathbb{E}_t\dot{p}_{t,k}, \mathbb{E}_t x_{t,k}\}$) are used. Throughout this paper, the region $\mathcal{DR} = \{\pi_{\dot{p}} > 1, \pi_x > 0\}$ is called the determinacy region. When $\pi_{\dot{p}} > 1$ and $\pi_x > 0$, regardless of the values of other unknown parameters, the $MPRF$ sufficiently satisfies the determinacy condition, *i.e.*, the monetary authority adjusts the nominal interest rates with 'sufficient strength' in response to inflations and output gaps (Woodford, 2003; Galí, 2008)[1].

In this paper we are interested in the nonlinear least squares (NLS) estimation and inference of the forecast-based $MPRF$ when the smoothing coefficient ρ is close to unity. When $\rho \approx 1$, the NLS objective function is relatively flat with respect to $\pi = \{\pi_\alpha, \pi_{\dot{p}}, \pi_x\}$ and π may not be consistently estimated. The inference about π based on the standard asymptotic theory (Newey and McFadden, 1994) may also be spurious because of a twofold reason. First, the Hessian of the NLS objective function is near singular when the objective

[1]According to Woodford (2003, Proposition 4.6), the determinacy condition of the $MPRF$ is:

$$\pi_{\dot{p}} + \frac{1-\beta_{discount}}{\lambda_{slope}}\pi_x - 1 > 0,$$

where $\beta_{discount} \in (0,1)$ and $\lambda_{slope} > 0$ are the discount factor and the slope parameter in the forward-looking Phillips curve. The definitions for the determinacy region in this paper is the same as Mavroeidis (2010).

function is relatively flat, and the standard asymptotic approximations involve the inverse of the Hessian. Second, when $\rho \approx 1$, the nominal interest rates $\{i_t\}$ will be highly persistent, and the NLS estimator will have a nonstandard asymptotic distribution. Lately close-to-one estimates for ρ had been found by Bunzel and Enders (2010), Nikolsko-Rzhevskyy (2011) and Nikolsko-Rzhevskyy and Papell (2012), but the identification failure of the reaction coefficients $\{\pi_{\dot{p}}, \pi_x\}$ when $\rho \approx 1$ has not been well studied. To the best of our knowledge, the identification failure of the $MPRF$ when $\rho \approx 1$ has only been noticed by Urquiza (2010) and Guerron-Quintana *et al.* (2009). Neither of them established the large sample properties of the estimators.

Three main contributions of this paper are as follows. First, our paper is the first in the literature establishing the large sample properties of the estimator / tests for the forecast-based $MPRF$ with a close-to-unity smoothing coefficient ρ. In this paper we modify the method of Andrews and Cheng (2012) on inference under weak / semi-strong identification to accommodate the persistence issue. Our modification involves the employment of asymptotic theories for near unit root processes (Phillips, 1987; Giraitis and Phillips, 2006) and novel drifting sequence approaches, which match the nonstandard convergence / divergence rates of the NLS estimator in the extreme case when $\rho = 1$. Large sample properties with a desired smooth transition with respect to the true values of parameters are developed for the NLS estimator and its corresponding t / Wald statistics of a general class of models.

Second, despite the not-consistent-estimability, the conservative confidence sets (CS) of weakly-identified parameters of interest ($\pi = \{\pi_\alpha, \pi_{\dot{p}}, \pi_x\}$) can be obtained by inverting the t / Wald tests. We show that the null-imposed least-favorable CS ($NILF$, Andrews and Cheng, 2012) will have correct asymptotic sizes, and the projection-based method (Dufour, 1997) may lead to asymptotic over-coverage.

Third, we obtain the conservative CS of the reaction coefficients $\{\pi_{\dot{p}}, \pi_x\}$ in U.S.'s forecast-based $MPRF$ for 1987:3–2007:4 with confidence coefficients $1 - \alpha = 0.8$, 0.9 and 0.95. The obtained CS contain many values of $\{\pi_{\dot{p}}, \pi_x\}$ not in the determinacy region $\mathcal{DR} = \{\pi_{\dot{p}} > 1, \pi_x > 0\}$. Our empirical application suggests that the NLS estimates for $\{\pi_{\dot{p}}, \pi_x\}$ are not accurate sufficiently to rule out the possibility of indeterminacy.

In the last decade there have been concerns over the identifiability of the monetary policy reaction function (*e.g.*, Cochrane, 2011; Inoue and Rossi, 2011; Mavroeidis, 2004, 2010). However, many were focus on the issue of weak instruments (weak IV). In a seminal paper Clarida, Galí and Gertler (2000) estimated the monetary policy reaction function of U.S.

for the pre-Volcker (1960:1 – 1979:2) / Volcker-Greenspan period (1979:3 – 1996:4)[2]. Since the expectations of the inflation and the output gap of the Federal Reserve ($\{\mathbb{E}_t\dot{p}_{t,k}, \mathbb{E}_t x_{t,k}\}$) were unobservable to the public, Clarida *et al.* (2000) replaced the *ex ante* expectations by the observable *ex post* realizations ($\{\dot{p}_{t,k}, x_{t,k}\}$).

$$\begin{aligned} i_t &= \rho i_{t-1} + (1-\rho)\left(\pi_\alpha + \pi_{\dot{p}}\dot{p}_{t,k} + \pi_x x_{t,k}\right) + \varepsilon_t^*, \\ \varepsilon_t^* &= \varepsilon_t - (1-\rho)\left[\pi_{\dot{p}}\left(\dot{p}_{t,k} - \mathbb{E}_t\dot{p}_{t,k}\right) + \pi_x\left(x_{t,k} - \mathbb{E}_t x_{t,k}\right)\right]. \end{aligned}$$

Because $\{\dot{p}_{t,k}, x_{t,k}\}$ would be correlated with ε_t^* (when $\rho \neq 1$ and $\pi_{\dot{p}} \neq 0$ / $\pi_x \neq 0$), Clarida *et al.* (2000) used the lags of $\{i_t, \dot{p}_{t,k}, x_{t,k}\}$ as IV and estimated the $MPRF$ by the generalized method of moments (GMM, Hansen, 1982). Their estimates for the reaction coefficients $\{\pi_{\dot{p}}, \pi_x\}$ for the pre-Volcker / Volcker-Greenspan period were respectively not in / in the determinacy region $\mathcal{DR} = \{\pi_{\dot{p}} > 1, \pi_x > 0\}$[3]. However, many empirical studies (*e.g.*, Inoue and Rossi, 2011; Mavroeidis, 2004, 2010) suggested that the lags of $\{i_t, \dot{p}_{t,k}, x_{t,k}\}$ were merely weakly correlated to $\{\dot{p}_{t,k}, x_{t,k}\}$. Recently Inoue and Rossi (2011) and Mavroeidis (2010) re-examined the empirical findings of Clarida *et al.* (2000). Inoue and Rossi (2011) developed a novel technique to test the strong identification of GMM estimation and rejected the null hypothesis of the strong identification of $\{\pi_{\dot{p}}, \pi_x\}$ for the Volcker-Greenspan period. Mavroeidis (2010) obtained the confidence set robust to weak IV and found the 90% robust confidence set of $\{\pi_{\dot{p}}, \pi_x\}$ for the Volcker-Greenspan period contains many values of parameters not in $\mathcal{DR} = \{\pi_{\dot{p}} > 1, \pi_x > 0\}$. Their findings suggested that the GMM estimates of $\{\pi_{\dot{p}}, \pi_x\}$ for the Volcker-Greenspan period were not accurate sufficiently to conclude the determinacy.

To prevent the identification failure due to weak IV, as in Orphanides (2001, 2004), we use the real-time data, *i.e.*, the historical *ex ante* forecasts of inflations and output gaps ($\{\mathbb{E}_t\dot{p}_{t,k}, \mathbb{E}_t x_{t,k}\}$) of the Federal Reserve. Orphanides (2004) collected the historical real-time data and estimated U.S.'s forecast-based $MPRF$ for the Volcker-Greenspan period (1979:3–1995:4) by NLS without any IV. His estimates for the reaction coefficients $\{\pi_{\dot{p}}, \pi_x\}$ were in the determinacy region[4]. Since 2008, the Greenbook projections of many macroeconomic variables have been open to the public (after a five-year declassification period) in the Federal

[2]The pre-Volcker period is the tenures of W. M. Martin, A. Burns and G. W. Miller as Federal Reserve chairmen. The Volcker-Greenspan period is the terms of P. Volcker and A. Greenspan.

[3]Instead of only one lag, Clarida *et al.* (2000) considered two lags of interest rates. Their estimates of $\{\pi_{\dot{p}}, \pi_x\}$ for the pre-Volcker / Volcker-Greenspan period ($k = 1$) were respectively $\{0.83, 0.27\}$ and $\{2.15, 0.93\}$.

[4]Orphanides (2004) collected the historical forecasts from the Greenbooks of Federal Reserve, the Council of Economic Advisers, the Department of Commerce and the internal Federal Reserve staff estimates. The estimates of Orphanides (2004) of $\{\pi_{\dot{p}}, \pi_x\}$ for the Volcker-Greenspan period ($k = 1, 2, 3, 4$) were respectively around $1.89 - 2.12$ and $0.14 - 0.18$.

Reserve Bank of Philadelphia[5]. For details about the real-time data, see Croushore and Stark (2001).

Lately close-to-unity estimates for the smoothing coefficient ρ had been found empirically, especially when more recent data was used. For example, Bunzel and Enders (2010) and Nikolsko-Rzhevskyy (2011) estimated the forecast-based $MPRF$ of U.S. with data up to 2007. Many of their estimates for ρ were around $0.88 - 0.98$[6]. Nikolsko-Rzhevskyy and Papell (2012) also found estimates for ρ around $0.88 - 0.94$ for the sample period 1966:1 – 1979:2 when using the Hodrick-Prescott (1997) filter in computing output gaps[7]. However, to the best of our knowledge, the identification failure of π when $\rho \approx 1$ had only been noticed by Urquiza (2010) and Guerron-Quintana *et al.* (2009). Urquiza (2010) found that when ρ approaches one, the zero-information-limit condition ($ZILC$, Nelson and Startz, 2007) is satisfied and the asymptotic variance of the NLS estimator of π become infinite. His Monte-Carlo simulations further showed that when the sample size is realistically small ($n = 100$), even if ρ is fairly below one (*e.g.*, $\rho = 0.8$), the inference for π based on the standard normal / χ^2 distribution is still spurious. Guerron-Quintana *et al.* (2009) suggested to reparameterize $(1-\rho)\,\pi$ to prevent the identification failure of π. Neither of them established the asymptotic properties of the estimators.

In this paper we modify the method of Andrews and Cheng (2012) on weak / semi-strong identification. In their seminal paper, Andrews and Cheng (2012) provided a unified treatment for a general class of models in which the parameters of interest are $\{\beta, \zeta, \pi\}$. β and ζ are always identified and can be $\sqrt{n}$-consistently estimated regardless of the value of π. π is identified if and only if $\beta \neq 0$ and the estimator for π may weakly converge to a nondegenerate random variable when $\beta \approx 0$. The problem considered in this paper looks similar to Andrews and Cheng (2012) if we reparameterize $\rho = 1 - \beta$ in the $MPRF$, and consider the following data generating process (DGP):

$$\begin{aligned} y_t &= \rho y_{t-1} + (1-\rho)\, X_t^\top \pi + \varepsilon_t \\ &= (1-\beta)\, y_{t-1} + \beta X_t^\top \pi + \varepsilon_t, \quad t = 1, \ldots, n. \end{aligned} \tag{1.2}$$

Like Andrews and Cheng (2012), π can be identified if and only if $\beta = 1 - \rho \neq 0$. However,

[5] http://www.philadelphiafed.org/research-and-data/real-time-center/

[6] Bunzel and Enders (2010) estimated the $MPRF$ with Taylor (1993)'s original backward looking rule for different subsample periods in 1965:3 – 2007:3. Most their estimates for ρ were in $0.894 - 0.974$. Nikolsko-Rzhevskyy (2011) estimated the forecast-based $MPRF$ using Greenbook projections. For different forecast horizons (k) in 1982:1 – 2007:1, his estimates for ρ when $k = 0$ or 1 were respectively 0.91 and 0.88.

[7] Nikolsko-Rzhevskyy and Papell (2012) considered different forecast horizons ($k = 1$ or 4) in 1966:1 – 1979:2 (with $p = 1$).

when Equation (1.2) contains a close-to-zero β (close-to-one ρ), $\{y_t\}$ will be highly persistent, and the NLS estimator for β will be super-consistent with a convergence rate n, and the NLS estimator for π will not possess limiting distributions but actually diverge as $n \to \infty$ with a divergence rate $\sqrt{n}$. Due to the different convergence rates of the estimators, the problem considered in this paper, despite the similarity, does not belong to the class of models considered by Andrews and Cheng (2012, 2013a, 2013b).

Two modifications were made to the existing method of Andrews and Cheng (2012). First, we propose novel and simple drifting sequence approaches in approximating the finite-sample behaviors of the NLS estimator. To study the weakly-identified π, Andrew and Cheng (2012) approximated the true value of β as a sequence drifting to zero with a standardization factor $\sqrt{n}$, which matched the convergence rate of the estimator for β in their models when $\beta = 0$. In this paper, however, to accommodate the persistence of $\{i_t\}$ when $\beta \approx 0$, drifting sequences different from Andrews and Cheng (2012) are selected to match the nonstandard convergence / divergence rates of NLS estimators when $\beta = 0$. Specifically, two different scenarios are considered. In the first scenario, 'local-to-zero β', $\beta = \beta_n$ drifts to zero with a standardization factor n^{-1}, and $\pi = \pi_n$ drifts to $\pm\infty$ with a standardization factor $n^{1/2}$. And the second scenario 'distant-from-zero β' bridges the 'local-to-zero β' scenario and the case when β is fixed and strictly above zero. As in Stock (1991), the drifting sequences in this paper are assumed to be simple linear functions of the unknown localization parameters. Divergent drifting sequences for parameter values have never appeared in the literature and may seem not intuitive. However, rather than any arbitrary artificial choice, the drifting-to-infinity sequences are logical outcomes of the nonstandard convergence / divergence rates in the NLS estimation when $\beta = 0$. Intuitively, the drifting-to-infinity π_n assumption is made simultaneously with the drifting-to-zero β_n assumption. We made this assumption to ensure the desired smooth transition in the asymptotic approximation to mimic the finite-sample behavior (Anatolyev and Gospodinov, 2011).

Second, by virtue of the linearity of drifting sequences, we are able to employ the asymptotic theories for near unit root processes (Phillips, 1987; Stock, 1991; Giraitis and Phillips, 2006) to establish the large sample properties with a desired smooth transition with respect to the true values of $\{\beta, \pi\}$ for the NLS estimator and its corresponding t / Wald test statistics. Specifically, when β is distant from zero, the t / Wald statistics will be asymptotically Gaussian / χ^2 distributed. However, when β is local to zero, the t / Wald statistics will have nonstandard and non-pivotal asymptotic distributions. Our Monte Carlo simulation shows that with the correctly specified values of the unknown and not-consistently-estimable localization parameters, our asymptotic approximations fit the finite-sample densities very well. Despite the drifting to infinity assumption for π, our asymptotic results provide good approximations even when π is small in magnitude (*e.g.*, $\pi = \mathbf{0}$). Our simulation also show that when β is close to zero, the finite-sample densities can not be approximated by ordinary bootstrapping procedure (*e.g.* resampling). It is not surprising since when β is local to zero, the asymptotic distributions will depend on the true values of the localization parameters.

The confidence sets (CS) for any linear functions of parameters are obtained by inverting the t / Wald tests. When β is local to zero, the CS will depend on the values of unknown and not-consistently-estimable localization parameters. Accordingly, we consider the null-imposed least-favorable method ($NILF$, Andrews and Cheng, 2012) and the projection-based method (Dufour, 1997). The $NILF$ method takes the supremum of the critical values of tests with respect to all possible values of the localization parameters under the null hypothesis corresponding to the tests to be inverted. The projection-based method projects the CS for all parameters to the codomain of the function of interest. Though both the $NILF$ method and the projection-based method are conservative, we show that the $NILF$ CS will have correct asymptotic sizes and the projection-based method may lead to asymptotic over-coverage. However, the projection-based method uses the information from the estimates for all parameters of interest, and is possible to obtain a more informative CS compared to the $NILF$ CS under certain circumstances. Both the $NILF$ method and the projection-based method require the computation of the test statistics for possible values of parameters. In practice, the CS can be obtained by grid methods.

According to our asymptotic theory, we construct the conservative CS for the reaction coefficients $\{\pi_{\dot{p}}, \pi_x\}$ in U.S.'s forecast-based $MPRF$ for 1987:3–2007:4. In the NLS estimation we use the Greenbook projections, *i.e.*, the real-time data for expected inflations and the expected output gaps ($\{\mathbb{E}_t \dot{p}_{t,k}, \mathbb{E}_t x_{t,k}\}$) from the Federal Reserve Bank of Philadelphia. As in Nikolsko-Rzhevskyy (2011), we consider the case with $k = 0$ or 1. For confidence coefficients $1 - \alpha = 0.8$, 0.9 and 0.95, both the $NILF$ and the projection-based CS contain many values not in the determinacy region $\mathcal{DR} = \{\pi_{\dot{p}} > 1, \pi_x > 0\}$. Our empirical results show that the NLS estimates for the reaction coefficients are not accurate sufficiently to rule out the possibility of indeterminacy.

The remainder of the paper is organized as follows. Section 2 provides the asymptotic theory for the NLS estimator when $\beta \approx 0$. Section 3 establishes the limiting properties of the t / Wald test statistics and introduces the procedure to obtain the CS for linear functions of parameters of interest. Section 4 gives the empirical results for U.S.'s forecast-based $MPRF$ for 1987:3–2007:4. Section 5 concludes. Proofs of the main results are collected in Appendix.

2 Assumptions and Asymptotic Theory

Consider the following data generating process (DGP) as Equation (1.2):

$$
\begin{aligned}
y_t &= \rho_n y_{t-1} + (1-\rho_n) X_t^\top \pi_n + \varepsilon_t \\
&= (1-\beta_n) y_{t-1} + \beta_n X_t^\top \pi_n + \varepsilon_t, \quad t = 1, \ldots, n.
\end{aligned}
$$

The DGP is known as the forecast-based monetary policy reaction function (forecast-based $MPRF$) when $\{y_t\}$ denotes the nominal interest rate and $\{X_t\}$ represents the expected inflation ($\mathbb{E}_t \dot{p}_{t,k}$), the expected output gap ($\mathbb{E}_t x_{t,k}$) and a constant one as in equation (1.1). We reparameterize $\rho_n = 1 - \beta_n$ so the notations are consistent with Andrews and Cheng (2012).

Assumption 1 *(Data generating process)* $y_t = (1-\beta_n) y_{t-1} + \beta_n X_t^\top \pi_n + \varepsilon_t$ *for* $t = 1, \ldots, n$, *where* $\theta_n = \{\beta_n, \pi_n\}$ *denote the true values of the parameters when the sample size equal to* $n \in \mathbb{N}$. $\theta_n \in \Theta_n^* \subset (0,1] \times \mathbb{R}^{d_\pi}$.

Assumption 2 $\{X_t\}$ *is a* d_π*-dimensional stationary ergodic sequence with* $\mathbb{E}(X_t) = \mu_X \in \mathbb{R}^{d_\pi}$ *and* $\mathbb{E}\left|X_{t,l}^2\right|^2 < \infty$ *for all* $l = 1, \ldots, d_\pi$ *and* $t = 1, \ldots, n$, *where* $X_{t,l}$ *denotes the* l*-th element of* X_t. $\mathbf{M}_X = \mathbb{E}\left(X_t X_t^\top\right) \in \mathbb{R}^{d_\pi \times d_\pi}$ *is positive definite.* $\mathbf{\Sigma}_X = \mathrm{var}(X_t) = \mathbf{M}_X - \mu_X \mu_X^\top$.

Assumption 3 $\{\varepsilon_t\}$ *is a sequence of independent and identically distributed (i.i.d.) random variables independent of* $\{y_{t-1}, X_t\}$ *with* $\mathbb{E}(\varepsilon_t) = 0$, $\mathbb{E}|\varepsilon_t|^2 < \infty$ *and* $\mathrm{var}(\varepsilon_t) = \sigma_\varepsilon^2 > 0$ *for all* $t = 1, \ldots, T$.

$\{\varepsilon_t\}$ is assumed to be serially uncorrelated because if $\{\varepsilon_t\}$ is also persistent, then in general $\mathrm{Cov}(y_{t-1}, \varepsilon_t) \neq 0$, *i.e.*, y_{t-1} will be endogenous. In that case when β_n is strictly greater than zero, the parameters of interest can not be consistently estimated by nonlinear least squares. We further assume $\{\varepsilon_t\}$ to be *i.i.d.* for simplicity.

For notational simplicity, let $\varphi_0 = \{\mu_X, \mathbf{M}_X, \sigma_\varepsilon^2\}$ denote the nuisance parameters, where $\varphi_0 \in \Phi_0 \subset \mathbb{R}^{d_\pi} \times \mathbb{R}^{d_\pi \times d_\pi} \times (0, \infty)$. Also let $\gamma_n = \{\theta_n, \varphi_0\} \in \Gamma_n = \Theta_n^* \times \Phi_0$ denote all the parameters in the model, including the parameters of interest $\theta_n = \{\beta_n, \pi_n\}$ and the nuisance parameters $\varphi_0 = \{\mu_X, \mathbf{M}_X, \sigma_\varepsilon^2\}$.

In this section we establish the asymptotic properties of the nonlinear least squares (NLS) estimator. $\theta_n = \{\beta_n, \pi_n\}$ belongs to the 'true parameter space' Θ_n^*. For any 'optimization parameter space' $\Theta_n \subset \mathbb{R}^{d_\pi+1}$ containing Θ_n^* (*i.e.*, $\Theta_n^* \subset \Theta_n$), the NLS estimator $\widehat{\theta}_n = \left\{\widehat{\beta}_n, \widehat{\pi}_n\right\}$ is defined as the minimizer of the objective function $Q_n(\theta;\gamma_n)$.

$$Q_n\left(\widehat{\theta}_n;\gamma_n\right) = \min_{\theta\in\Theta_n} Q_n(\theta;\gamma_n) = \min_{\theta\in\Theta_n} \frac{1}{2n}\sum_{t=1}^{n}\left[y_t - (1-\beta)\,y_{t-1} - \beta X_t^\top \pi\right]^2. \tag{2.1}$$

In practice, the optimization parameter space Θ_n can be selected as a large set to prevent the misspecification of the parameter space.

When $\beta_n = \beta_0$ and $\pi_n = \pi_0$, *i.e.*, when θ_n is fixed at the constant vector $\theta_0 = \{\beta_0, \pi_0\} \in \Theta_n^*$, by the standard asymptotic theory (Newey and McFadden, 1994), $\widehat{\theta}_n$ is $\sqrt{n}$-consistent and asymptotically normally distributed.

Theorem 1 *Suppose that Assumptions 1, 2 and 3 hold and $\theta_n = \theta_0 \in \Theta_n^*$, i.e., $\beta_n = \beta_0$ and $\pi_n = \pi_0$ for any $n \in \mathbb{N}$. Then $\widehat{\theta}_n \xrightarrow{p} \theta_n$, and*

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_n\right) \overset{A}{\sim} \mathcal{N}\left(\mathbf{0}_{(d_\pi+1)\times 1}, \sigma_\varepsilon^2 \mathcal{V}_0^{-1}(\gamma_n)\right),$$

where $\mathcal{V}_0(\gamma_n)$ is the probability limit of the Hessian of the NLS objective function,

$$\mathcal{V}_0(\gamma_n) = \operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\begin{bmatrix} \left(y_{t-1} - X_t^\top\pi_0\right)^2 & -\beta_0\left(y_{t-1} - X_t^\top\pi_0\right)X_t^\top \\ -\beta_0 X_t\left(y_{t-1} - X_t^\top\pi_0\right) & \beta_0^2 X_t X_t^\top \end{bmatrix}.$$

However, when $\beta = 0$, the NLS objective function $Q_n(\theta;\gamma_n)$ does not depend on π and therefore π is not identifiable. And when $\beta \approx 0$, the NLS objective function is relatively flat with respect to π and therefore π may not be consistently estimated. The inference about π based on the standard asymptotic results (Theorem 1) may also be spurious because of a twofold reason. First, the Hessian of the NLS objective function $\mathcal{V}_0(\gamma_n)$ is near singular when the objective function is relatively flat, and the standard asymptotic approximations involve the inverse of the Hessian $\mathcal{V}_0(\gamma_n)$. Second, when $\beta \approx 0$, the sequence $\{y_t\}$ will be highly persistent, and the NLS estimator $\widehat{\theta}_n$ will have a nonstandard asymptotic distribution.

To study the case when $\beta \approx 0$, first we consider the extreme case when $\beta_n = 0$. For simplicity, we assume $y_0 = o_p\left(n^{1/2}\right)$ to prevent the the effect from the initial observation. This assumption is similar to the conditional case assumption in the unit root literature (Elliott *et al*, 1996).

Lemma 1 *Suppose that Assumptions 1, 2 and 3 hold except that β_n is assumed to be* 0 *for any $n \in \mathbb{N}$. If $y_0 = o_p\left(n^{1/2}\right)$, then $\widehat{\beta}_n = O_p\left(n^{-1}\right)$, and $\widehat{\pi}_n = O_p\left(n^{1/2}\right)$.*

In Lemma 1 we show that when $\beta_n = 0$, $\widehat{\beta}_n$ will be super-consistent with a convergence rate n, and $\widehat{\pi}_n$ does not possess limiting distribution but actually diverge as $n \to \infty$ with a divergence rate $\sqrt{n}$. Accordingly, in this paper we consider the following two different asymptotic approaches, $\Gamma_n(1, b, \mathbf{c})$ and $\Gamma_n(h, b, \mathbf{c})$, to mimic the finite sample behaviors of $\widehat{\theta}_n = \left\{\widehat{\beta}_n, \widehat{\pi}_n\right\}$. Through out this paper, the two classes $\Gamma_n(1, b, \mathbf{c})$ and $\Gamma_n(h, b, \mathbf{c})$ are called the 'local-to-zero β_n' and 'distant-from-zero β_n' scenarios.

Definition 1 *($\Gamma_n(1, b, \mathbf{c})$ and $\Gamma_n(h, b, \mathbf{c})$) For any $b \in (0, +\infty)$, $\mathbf{c} \in \mathbb{R}^{d_\pi}$ and $h \in (0, 1)$,*

$$\begin{aligned}\Gamma_n(1, b, \mathbf{c}) &= \left\{\{\gamma_n\} \in \Gamma_n : \beta_n = \frac{b}{n}, \quad \pi_n = n^{1/2}\mathbf{c}\right\}, \\ \Gamma_n(h, b, \mathbf{c}) &= \left\{\{\gamma_n\} \in \Gamma_n : \beta_n = \frac{b}{n^h}, \quad \pi_n = n^{-1/2+h}\mathbf{c}\right\}.\end{aligned}$$

For $\gamma_n \in \Gamma_n(1, b, \mathbf{c})$, β_n and π_n are assumed to be sequences respectively drifting to zero / $\pm\infty$ when $n \to \infty$. The standardization factors n^{-1} and $n^{1/2}$ are selected to match the convergence / divergence rates of the NLS estimator when $\beta_n = 0$ (Lemma 1). The asymptotic approach $\Gamma_n(h, b, \mathbf{c})$ bridges the case $\{\theta_n = \theta_0 \in \Theta_n^*\}$ and the local-to-zero β_n class $\Gamma_n(1, b, \mathbf{c})$. The drifting sequences for β_n are exactly the same as the frequently used local-to-unity (Phillips, 1987; Stock, 1991) / neighborhood-of-unity (Giraitis and Phillips, 2006; Phillips and Magdalinos, 2007) asymptotic approaches if we reparameterize $\rho_n = 1 - \beta_n$. Divergent drifting sequences, to the best of our knowledge, have never appeared in the literature and may seem not intuitive. However, rather than any arbitrary artificial choice, the drifting-to-infinity sequences are logical outcomes of the convergence / divergence rates of the NLS estimators when $\beta_n = 0$. We will discuss the divergent drifting sequences in more details in Subsection 2.3.

In the following two subsections we establish the asymptotic results under $\Gamma_n(1, b, \mathbf{c})$ and $\Gamma_n(h, b, \mathbf{c})$. Our method is a modification of Andrews and Cheng (2012) on weak / semi-strong identification. In their seminal paper, Andrews and Cheng (2012) provided a unified treatment of a general class of models in which the parameters of interest are $\{\beta, \zeta, \pi\}$. β and ζ are always identified and can be $\sqrt{n}$-consistently estimated regardless of the value of π. π is identified if and only if $\beta \neq 0$ and the estimator for π may weakly converge to a nondegenerate random variable when $\beta \approx 0$. Despite the similarity, in Lemma 1 we have already shown that when $\beta_n = 0$, $\widehat{\beta}_n$ and $\widehat{\pi}_n$ are respectively $O_p(n^{-1})$ / $O_p(n^{1/2})$. Due to the different convergence / divergence rates of the estimators, the problem considered in this paper does not belong to the class of models considered by Andrews and Cheng (2012, 2013a, 2013b). Although different drifting sequences are used, we develop the asymptotic properties of the NLS estimator and its corresponding t / Wald test statistics with quadratic approximations for the objective function similar to Andrews and Cheng (2012).

In contrast to Andrew and Cheng (2012), who considered more general drifting sequences (*e.g.*, $n^{1/2}\beta_n \to b$), the drifting sequences in this paper are assumed to be simple linear functions of the unknown localization parameters ($\beta_n = n^{-1}b$, $\pi_n = n^{1/2}\mathbf{c}$, or $\beta_n = n^{-1}b$, $\pi_n = n^{1/2}\mathbf{c}$). The linear drifting sequences and the property of the exponential function $\lim_{n\to\infty}(1-n^{-1}b)^n = \exp(-b)$ allow us to employ the large sample theory for the time series with a local-to-unity root by Phillips (1987) and Stock (1991) in the establishment of the asymptotic approximations. When obtaining the confidence sets for linear functions of parameters by inverting the tests, as in Stock (1991), the linear drifting sequences also guarantee a surjective mapping from the values of localization parameters to the null hypotheses corresponding to the tests to be inverted, which is very useful in constructing a more informative but still conservative confidence set.

2.1 Estimation Results for Local-to-Zero β_n

In this subsection we determine the asymptotic distributions of the NLS estimator $\widehat{\theta}_n = \left\{\widehat{\beta}_n, \widehat{\pi}_n\right\}$ when $\gamma_n \in \Gamma_n(1, b, \mathbf{c})$, *i.e.*, $\beta_n = n^{-1}b$ and $\pi_n = n^{1/2}\mathbf{c}$. When $\gamma_n \in \Gamma_n(1, b, \mathbf{c})$, as in Andrews and Cheng (2012), we consider a quadratic approximation for $Q_n(\beta, \pi; \gamma_n)$ in β around $\beta = 0$.

$$Q_n(\beta, \pi; \gamma_n) - Q_n(0, \pi; \gamma_n) = \frac{\partial}{\partial\beta}Q_n(0, \pi; \gamma_n)\cdot\beta + \frac{1}{2}\frac{\partial^2}{\partial\beta^2}Q_n(\beta^*, \pi; \gamma_n)\cdot\beta^2, \tag{2.2}$$

where $0 < \beta^* < \beta$,

$$\begin{aligned}
\frac{\partial}{\partial\beta}Q_n(0, \pi; \gamma_n) &= n^{-1}\sum_{t=1}^{n}(y_t - y_{t-1})\left(y_{t-1} - X_t^\top\pi\right), \\
\frac{\partial^2}{\partial\beta^2}Q_n(\beta^*, \pi; \gamma_n) &= n^{-1}\sum_{t=1}^{n}\left(y_{t-1} - X_t^\top\pi\right)^2.
\end{aligned}$$

Since $\partial^2 Q_n(\beta, \pi; \gamma_n) / \partial \beta^2$ does not depend on β, $\partial^2 Q_n(\beta^*, \pi; \gamma_n) / \partial \beta^2 = \partial^2 Q_n(0, \pi; \gamma_n) / \partial \beta^2$. Therefore, equation (2.2) can be written as:

$$Q_n(\beta, \pi; \gamma_n) - Q_n(0, \pi; \gamma_n) = \frac{\partial}{\partial \beta} Q_n(0, \pi; \gamma_n) \cdot \beta + \frac{1}{2} \frac{\partial^2}{\partial \beta^2} Q_n(0, \pi; \gamma_n) \cdot \beta^2. \tag{2.3}$$

For any $\mathbb{R}^{d_\pi}$-valued π, when $n \to \infty$, let

$$n^{-1/2} \pi \Rightarrow \kappa_\pi. \tag{2.4}$$

Lemma 2 *Suppose that Assumptions 1, 2 and 3 hold, $\gamma_n \in \Gamma_n(1, b, \mathbf{c})$, and $y_0 = o_p(n^{1/2})$. Then for any $\mathbb{R}^{d_\pi}$-valued π with $n^{-1/2}\pi \Rightarrow \kappa_\pi$ as $n \to \infty$,*

$$\begin{aligned} \frac{\partial}{\partial \beta} Q_n(0, \pi; \gamma_n) &\Rightarrow \mathcal{G}(\kappa_\pi, b, \mathbf{c}; \varphi_0), \quad \textit{and} \\ n^{-1} \frac{\partial^2}{\partial \beta^2} Q_n(0, \pi; \gamma_n) &\Rightarrow \mathcal{H}(\kappa_\pi, b, \mathbf{c}; \varphi_0), \end{aligned}$$

where $\mathcal{G}(\kappa_\pi, b, \mathbf{c}; \varphi_0)$ and $\mathcal{H}(\kappa_\pi, b, \mathbf{c}; \varphi_0)$ are defined in Lemma 5 in Appendix A.

According to equation (2.3) and Lemma 2, let $q(\lambda_\beta, \kappa_\pi, b, \mathbf{c}; \varphi_0)$ be the asymptotic approximation of $Q_n(\beta, \pi; \gamma_n) - Q_n(0, \pi; \gamma_n)$,

$$q(\lambda_\beta, \kappa_\pi, b, \mathbf{c}; \varphi_0) = \mathcal{G}(\kappa_\pi, b, \mathbf{c}; \varphi_0) \cdot \lambda_\beta + \frac{1}{2} \mathcal{H}(\kappa_\pi, b, \mathbf{c}; \varphi_0) \cdot \lambda_\beta^2. \tag{2.5}$$

For any given κ_π, let $\widehat{\lambda}_\beta(\kappa_\pi, b, \mathbf{c}; \varphi_0)$ be the infimizer of $q(\lambda_\beta, \kappa_\pi, b, \mathbf{c}; \gamma_n)$:

$$q\left(\widehat{\lambda}_\beta(\kappa_\pi, b, \mathbf{c}; \varphi_0), \kappa_\pi, b, \mathbf{c}; \varphi_0\right) = \inf_{\lambda_\beta} q(\lambda_\beta, \kappa_\pi, b, \mathbf{c}; \varphi_0), \tag{2.6}$$

and $\widehat{\kappa}_\pi(b, \mathbf{c}; \varphi_0)$ be the infimizer of $q\left(\widehat{\lambda}_\beta(\kappa_\pi, b, \mathbf{c}; \varphi_0), \kappa_\pi, b, \mathbf{c}; \varphi_0\right)$:

$$\begin{aligned} & q\left(\widehat{\lambda}_\beta(\widehat{\kappa}_\pi(b, \mathbf{c}; \varphi_0), b, \mathbf{c}; \varphi_0), \widehat{\kappa}_\pi(b, \mathbf{c}; \varphi_0), b, \mathbf{c}; \varphi_0\right) \\ = \; & \inf_{\kappa_\pi} q\left(\widehat{\lambda}_\beta(\kappa_\pi, b, \mathbf{c}; \varphi_0), \kappa_\pi, b, \mathbf{c}; \varphi_0\right). \end{aligned} \tag{2.7}$$

Theorem 2 *Suppose that Assumptions 1, 2 and 3 hold, $\gamma_n \in \Gamma_n(1, b, \mathbf{c})$, and $y_0 = o_p\left(n^{1/2}\right)$. Then*

$$\begin{bmatrix} n\left(\widehat{\beta}_n - \beta_n\right) \\ n^{-1/2}\left(\widehat{\pi}_n - \pi_n\right) \end{bmatrix} \Rightarrow \widehat{\tau}(b, \mathbf{c}; \varphi_0) = \begin{bmatrix} \widehat{\lambda}_\beta\left(\widehat{\kappa}_\pi(b, \mathbf{c}; \varphi_0), b, \mathbf{c}; \varphi_0\right) - b \\ \widehat{\kappa}_\pi(b, \mathbf{c}; \varphi_0) - \mathbf{c} \end{bmatrix}.$$

Remark 1 *1. In Theorem 2 we show that when $\gamma_n \in \Gamma_n(1, b, \mathbf{c})$, $\widehat{\beta}_n$ is super-consistent with a convergence rate n, and $\widehat{\pi}_n$ does not possess limiting distribution but actually diverge as $n \to \infty$ with a divergence rate $\sqrt{n}$. The asymptotic distributions of $n\left(\widehat{\beta}_n - \beta_n\right) / n^{-1/2}\left(\widehat{\pi}_n - \pi_n\right)$ are nonstandard and depend on the values of unknown parameters, including nuisance parameters $\varphi_0 = \{\mu_X, \mathbf{M}_X, \sigma_\varepsilon^2\}$ and localization parameters $\{b, \mathbf{c}\}$. In Section 3 we will show that when $\gamma_n \in \Gamma_n(1, b, \mathbf{c})$, the t / Wald test statistics corresponding to the null hypothesis $H_0 : \mathbf{R}\theta_n = \upsilon$ and the confidence sets of $\mathbf{R}\theta_n$ will still depend on the values of $\{b, \mathbf{c}\}$. And it causes difficulties in testing H_0 and obtaining the confidence sets of $\mathbf{R}\theta_n$.*

2. The problem considered in this paper is not in the class of models in Andrews and Cheng (2012), and our drifting sequence approaches are different from theirs. However, our quadratic approximation of the NLS objective function, which is only with respect to β around $\beta = 0$, is similar to the corresponding weak-identification scenario in Andrews and Cheng (2012). Since π vanishes in $Q_n(\beta, \pi; \gamma_n)$ when $\beta = 0$, $Q_n(0, \pi; \gamma_n)$ does not depend on the values of both β and π. Therefore, the NLS estimator $\widehat{\theta}_n = \left\{\widehat{\beta}_n, \widehat{\pi}_n\right\}$ is also a minimizer for $Q_n(\beta, \pi; \gamma_n) - Q_n(0, \pi; \gamma_n)$, which has the quadratic expansion as in equation (2.3). Then the asymptotic properties of $\widehat{\theta}_n = \left\{\widehat{\beta}_n, \widehat{\pi}_n\right\}$ can be determined with Lemma 2, which employs the asymptotic theories for near unit root processes by Phillips (1987) and Stock (1991). Because of the persistence of $\{y_t\}$ when $\beta \approx 0$, the empirical process central limit theorems (e.g., Andrews, 1994) used by Andrews and Cheng (2012) in their corresponding weak-identification scenario can not be applied to the problem in the present paper.

According to Theorem 2, the asymptotic distributions of $n\left(\widehat{\beta}_n - \beta_n\right) / n^{-1/2}\left(\widehat{\pi}_n - \pi_n\right)$ depend on unknown nuisance parameters $\varphi_0 = \{\mu_X, \mathbf{M}_X, \sigma_\varepsilon^2\}$. Let $\{\widehat{\varepsilon}_t\}$ be the residuals of the NLS estimation, and $\widehat{\varphi}_n = \left\{\widehat{\mu}_{X,n}, \widehat{\mathbf{M}}_{X,n}, \widehat{\sigma}_n^2\right\}$ be the estimator for φ_0:

$$\begin{aligned} \widehat{\mu}_{X,n} &= n^{-1}\sum_{t=1}^n X_t, \quad \widehat{\mathbf{M}}_{X,n} = \frac{1}{n}\sum_{t=1}^n X_t X_t^\top, \quad \widehat{\sigma}_n^2 = n^{-1}\sum_{t=1}^n \widehat{\varepsilon}_t^2, \quad \text{where} \qquad (2.8) \\ \widehat{\varepsilon}_t &= y_t - \left(1 - \widehat{\beta}_n\right) y_{t-1} - \widehat{\beta}_n X_t^\top \widehat{\pi}_n, \quad t = 1\ldots, n. \end{aligned}$$

Lemma 3 *Suppose that all conditions of Theorem 2 are satisfied. Then* $\widehat{\varphi}_n \xrightarrow{p} \varphi_0$.

Lemma 3 shows that φ_0 can be consistently estimated by $\widehat{\varphi}_n$. Therefore, when the true values of the localization parameters $\{b, \mathbf{c}\}$ are known, we are able to replace the unknown nuisance parameters φ_0 with the estimates $\widehat{\varphi}_n$, and obtain the asymptotic distributions of $n\left(\widehat{\beta}_n - \beta_n\right) \;/\; n^{-1/2}\left(\widehat{\pi}_n - \pi_n\right)$ by Monte Carlo simulation. We omit the formal proof since it directly follows by the continuous mapping theorem. Our Monte Carlo simulation in Example 1 shows that our asymptotic approximations fit the finite-sample densities very well. Since the localization parameters $\{b, \mathbf{c}\}$ are unknown in practice, we propose to approximate the finite-sample behaviors of $n\left(\widehat{\beta}_n - \beta_n\right) \;/\; n^{-1/2}\left(\widehat{\pi}_n - \pi_n\right)$ by the grid method, *i.e.*, to generate grids over the parameter space Θ_n^* and to obtain the asymptotic approximations for every grid. Due to the not-consistent-estimability of $\{b, \mathbf{c}\}$, we are not able to determine the 'correct' grid. However, the asymptotic distributions for every grids can be used in the construction of the confidence sets with correct asymptotic sizes for specific linear functions of parameters θ_n, which we will discuss in more details in section 3. In Example 1 we also show that because the asymptotic distributions of $n\left(\widehat{\beta}_n - \beta_n\right) \;/\; n^{-1/2}\left(\widehat{\pi}_n - \pi_n\right)$ depend on the true values of the localization parameters $\{b, \mathbf{c}\}$, the bootstrapping procedures, *e.g.*, the simple resampling, will not provide valid approximations for the finite-sample behaviors.

Example 1 *Consider the following model as equation (2.9):*

$$y_t = (1 - \beta_n)\, y_{t-1} + \beta_n \left(\pi_{0,n} + \pi_{1,n} x_t\right) + \varepsilon_t, \quad t = 1, \ldots, n, \tag{2.9}$$

where $x_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, $\beta_n = b\,/n$, $\pi_{0,n} = n^{1/2} c_0$, $\pi_{1,n} = n^{1/2} c_1$, *and* $n = 100$.

Using Theorem 2, Figures 1 and 2 provide the simulated finite-sample and asymptotic densities of $n\left(\widehat{\beta}_n - \beta_n\right)$ *and* $n^{-1/2}\left(\widehat{\pi}_{1,n} - \pi_{1,n}\right)$ *given the true values of* $\{b, c_0, c_1\}$. *We consider* $\beta_n \in \{0.02, 0.05, 0.1\}$ *and* $\pi_{0,n} = \pi_{1,n} = 2$, *i.e.*, $b \in \{2, 5, 10\}$ *and* $c_0 = c_1 = 0.2$. *We do not report the densities of* $\widehat{\pi}_{0,n}$ *since the results are similar to* $\widehat{\pi}_{1,n}$. *The asymptotic approximations based on Theorem 2 fit the finite-sample densities very well. In contrast, Figures 3 and 4 provide the simulated finite-sample and bootstrapping (resampling) densities of* $n\left(\widehat{\beta}_n - \beta_n\right)$ *and* $n^{-1/2}\left(\widehat{\pi}_{1,n} - \pi_{1,n}\right)$. *The resampling densities do not fit the finite-sample densities.*

For all results $50{,}000$ *simulation repetitions are used. For the asymptotic densities, we first generate* 200 *different sets of data to estimate the unknown nuisance parameters* $\varphi_0 = \{\mu_X, \mathbf{M}_X, \sigma_\varepsilon^2\}$. *And with each estimates* $\widehat{\varphi}_n$ *we generate the asymptotic approximations with* 250 *repetitions. The Wiener process* $\mathcal{W}_\varepsilon(r)$ *and the Ornstein–Uhlenbeck* *process* $\mathcal{J}_{-b,\varepsilon}(r)$ *in the asymptotic distributions are approximated by* $T^{-1/2}\sum_{s=1}^{\lfloor Tr\rfloor}\eta_s$ *and* $T^{-1/2}\sum_{s=1}^{\lfloor Tr\rfloor}(1-b/T)^{\lfloor Tr\rfloor-s}\eta_s$ *with* $T=10,000$ *and* $\eta_t\overset{i.i.d.}{\sim}\mathcal{N}(0,1)$.

2.2 Estimation Results for Distant-from-Zero β_n

In this subsection we determine the asymptotic distributions of the NLS estimator $\widehat{\theta}_n=\left\{\widehat{\beta}_n,\widehat{\pi}_n\right\}$ when $\gamma_n\in\Gamma_n(h,b,\mathbf{c})$, i.e., $\beta_n=n^{-h}b$ and $\pi_n=n^{-1/2+h}\mathbf{c}$, where $h\in(0,1)$. When $\gamma_n\in\Gamma_n(h,b,\mathbf{c})$, we consider a quadratic approximation for $Q_n(\theta;\gamma_n)$ around θ_n as Newey and McFadden (1994) and Andrews and Cheng (2012):

$$\begin{aligned}&Q_n(\theta;\gamma_n)-Q_n(\theta_n;\gamma_n)\\=\;&D_\theta^\top Q_n(\theta_n;\gamma_n)(\theta-\theta_n)+\frac{1}{2}(\theta-\theta_n)^\top D_{\theta\theta^\top}Q_n(\theta_n;\gamma_n)(\theta-\theta_n)+R(\theta^*;\gamma_n),\end{aligned}\tag{2.10}$$

in which θ^* is in between of θ_n and θ,

$$\begin{aligned}D_\theta Q_n(\theta_n;\gamma_n)&=\begin{bmatrix}n^{-1}\sum_{t=1}^n\left(y_{t-1}-X_t^\top\pi_0\right)\varepsilon_t\\-\beta_n n^{-1}\sum_{t=1}^n X_t\varepsilon_t\end{bmatrix},\\D_{\theta\theta^\top}Q_n(\theta_n;\gamma_n)&=\begin{bmatrix}n^{-1}\sum_{t=1}^n\left(y_{t-1}-X_t^\top\pi_0\right)^2&\\-n^{-1}\sum_{t=1}^n X_t\left[\beta_n\left(y_{t-1}-X_t^\top\pi_0\right)+\varepsilon_t\right]&\beta_n^2n^{-1}\sum_{t=1}^n X_tX_t^\top\end{bmatrix}.\end{aligned}$$

Let

$$B(h)=\begin{bmatrix}n^{h/2}&\mathbf{0}_{1\times d_\pi}\\\mathbf{0}_{d_\pi\times1}&n^{-h}\mathbb{I}_{d_\pi}\end{bmatrix}.\tag{2.11}$$

Lemma 4 *Suppose that Assumptions 1, 2 and 3 hold,* $\gamma_n\in\Gamma_n(h,b,\mathbf{c})$ *and* $\gamma=\{\theta,\varphi_0\}\in\Gamma_n(h,b,\mathbf{c})$. *Then*

1. $n^{1/2}B^{-1}(h)D_\theta Q_n(\theta_n;\gamma_n)\Rightarrow\mathcal{G}^*(b;\varphi_0)\sim\mathcal{N}\left(\mathbf{0}_{(d_\pi+1)\times1},\sigma_\varepsilon^2\mathcal{V}_h(b;\varphi_0)\right)$, *where*

$$\mathcal{V}_h(b;\varphi_0)=\begin{bmatrix}(2b)^{-1}\sigma_\varepsilon^2&\mathbf{0}_{1\times d_\pi}\\\mathbf{0}_{d_\pi\times1}&b^2\mathbf{M}_X\end{bmatrix}.$$

2. $B^{-1}(h)D_{\theta\theta^\top}Q_n(\theta_n;\gamma_n)B^{-1}(h)\xrightarrow{p}\mathcal{V}_h(b;\varphi_0)$.

Theorem 3 *Suppose that Assumptions 1, 2 and 3 hold and $\gamma_n \in \Gamma_n(h, b, \mathbf{c})$. Then*

$$n^{1/2} B(h)\left(\widehat{\theta}_n - \theta_n\right) = \begin{bmatrix} n^{1/2+h/2}\left(\widehat{\beta}_n - \beta_n\right) \\ n^{1/2-h}\left(\widehat{\pi}_n - \pi_n\right) \end{bmatrix} \overset{A}{\sim} \mathcal{N}\left(\mathbf{0}_{(d_\pi+1)\times 1}, \sigma_\varepsilon^2 \mathcal{V}_h^{-1}(b;\varphi_0)\right).$$

Remark 2 1. *In Theorem 3 we show that when $\gamma_n \in \Gamma_n(h, b, \mathbf{c})$, $\widehat{\beta}_n - \beta_n = O_p\left(n^{-1/2-h/2}\right)$, and $\widehat{\pi}_n - \pi_n = O_p\left(n^{-1/2+h}\right)$. Despite the non-standard convergence / divergence rates, the asymptotic distributions of $n^{1/2+h/2}\left(\widehat{\beta}_n - \beta_n\right)$ / $n^{1/2-h}\left(\widehat{\pi}_n - \pi_n\right)$ are standard (Gaussian distributions). In the next section when we consider the tests for the null hypothesis $H_0 : \mathbf{R}\theta_n = \upsilon$ and the confidence sets of $\mathbf{R}\theta_n$, we will show that when $\gamma_n \in \Gamma_n(h, b, \mathbf{c})$, the asymptotic distributions of the t / Wald test statistics corresponding to H_0 will also be standard (Gaussian / χ^2 distributions) and pivotal (not depending on the values of $\{b, \mathbf{c}, h\}$). This result will be very useful in testing H_0 and obtaining the confidence sets of $\mathbf{R}\theta_n$.*

2. *Again, the problem considered in this paper is not in the class of models in Andrews and Cheng (2012), and our drifting sequence approaches are different from theirs. However, our quadratic approximation of the NLS objective function is similar to the corresponding semi-strong-identification scenario in Andrews and Cheng (2012). The asymptotic properties of $\widehat{\theta}_n = \left\{\widehat{\beta}_n, \widehat{\pi}_n\right\}$ are determined with Lemma 4, which employs the asymptotic theory for near unit root processes by Giraitis and Phillips (2006), who rescaled the statistics of interest to satisfy the central limit theorem. Andrews and Cheng (2012) also rescaled their statistics of interest for exactly the same reason in their semi-strong-identification case.*

2.3 Sequences Drifting to Infinity

When $\gamma_n \in \Gamma_n(1, b, \mathbf{c})$, or $\gamma_n \in \Gamma_n(h, b, \mathbf{c})$ with $h > 1/2$, to mimic the true value of π we use sequences drifting to $\pm\infty$ ($\pi_n = n^{1/2}\mathbf{c}$ when $\gamma_n \in \Gamma_n(1, b, \mathbf{c})$, and $\pi_n = n^{-1/2+h}\mathbf{c}$ when $\gamma_n \in \Gamma_n(h, b, \mathbf{c})$), which, to the best of our knowledge, have never appeared before in the literature.

We consider sequences drifting to infinity for two reasons. First, rather than any arbitrary artificial choice, the drifting-to-infinity sequences are logical outcomes of the convergence / divergence rates of the NLS estimators. Lemma 1 shows that when $\beta_n = 0$, $\widehat{\beta}_n$ will be super-consistent with a convergence rate n, and $\widehat{\pi}_n$ does not possess limiting distribution but actually diverge as $n \to \infty$ with a divergence rate $\sqrt{n}$. The class $\Gamma_n(1, b, \mathbf{c})$ ($\beta_n = n^{-1}b$ and $\pi_n = n^{1/2}\mathbf{c}$) matches the convergence / divergence rates. And the class $\Gamma_n(h, b, \mathbf{c})$ ($\beta_n = n^{-h}b$ and $\pi_n = n^{-1/2+h}\mathbf{c}$) bridges the two cases ($\{\theta_n = \theta_0 \in \Theta_n^*\}$ and $\Gamma_n(1, b, \mathbf{c})$). Intuitively, the drifting-to-infinity π_n assumption is made simultaneously with the drifting-to-zero β_n assumption. We made this assumption to ensure the desired smooth transition in the asymptotic approximation to mimic the finite-sample behavior (Anatolyev and Gospodinov, 2011).

Second, Theorems 2 and 3 show the necessity of the drifting-to-infinity assumption for inference about π_n. When β_n is approximated by a local-to-zero sequence ($\beta_n = b/n$), if we consider a drifting sequence for π_n with a standardization factor less than $n^{1/2}$, *e.g.*, if we assume that π_n is a non-zero constant vector, then by Theorem 2 with $\mathbf{c} \to \mathbf{0}$,

$$n^{-1/2}\widehat{\pi}_n = n^{-1/2}(\widehat{\pi}_n - \pi_n) + o(1) \Rightarrow \widehat{\kappa}_\pi(b, \mathbf{0}; \varphi_0).$$

That is, $n^{-1/2}\widehat{\pi}_n \overset{A}{=} n^{-1/2}(\widehat{\pi}_n - \pi_n)$, the asymptotic distribution of $\widehat{\pi}_n$ does not depend on the true value of π_n, and we are not able to make inference about π_n based on its estimate $\widehat{\pi}_n$. Similarly, when β_n is approximated by a neighborhood-of-zero sequence ($\beta_n = b/n^h$ with $h \in (0,1)$), if we consider a drifting sequence for π_n with a standardization factor less than $n^{-1/2+h}$, then again, by Theorem 3 with $\mathbf{c} \to \mathbf{0}$,

$$n^{1/2-h}\widehat{\pi}_n = n^{1/2-h}(\widehat{\pi}_n - \pi_n) + o(1) \overset{A}{\sim} \mathcal{N}\left(\mathbf{0}_{d_\pi \times 1}, \sigma_\varepsilon^2 b^{-2}\mathbf{M}_X^{-1}\right).$$

Again, $n^{1/2-h}\widehat{\pi}_n \overset{A}{=} n^{1/2-h}(\widehat{\pi}_n - \pi_n)$, the asymptotic distribution of $\widehat{\pi}_n$ does not depend on the true value of π_n, and we are not able to make inference about π_n based on its estimate $\widehat{\pi}_n$, either.

Although we reparameterize the true value of π as a sequence drifting to $\pm\infty$ as the sample size $n \to \infty$ when $\gamma_n \in \Gamma_n(1, b, \mathbf{c})$, or $\gamma_n \in \Gamma_n(h, b, \mathbf{c})$ with $h > 1/2$, the true value of π should not be viewed as an infinite number. π_n is always finite for any sample size $n \in \mathbb{N}$. In practice, the true value of π does not need to be large in magnitude. In Example 2 we show that even if $\pi = \mathbf{0}$, our asymptotic approximations based on Theorem 2 still fit the finite-sample densities very well.

Example 2 *Again, consider the following model as equation (2.9):*

$$y_t = (1 - \beta_n) y_{t-1} + \beta_n (\pi_{0,n} + \pi_{1,n} x_t) + \varepsilon_t, \quad t = 1, \ldots, n,$$

where $x_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, $\beta_n = b/n$, $\pi_{0,n} = n^{1/2}c_0$, $\pi_{1,n} = n^{1/2}c_1$, *and* $n = 100$. *Still, we consider* $\beta_n \in \{0.02, 0.05, 0.1\}$, *i.e.,* $b \in \{2, 5, 10\}$. *However, in this example we consider the case when* $\pi_{0,n} = \pi_{1,n} = 0$, *i.e.,* $c_0 = c_1 = 0$.

Using Theorem 2, Figures 5 and 6 provide the simulated finite-sample and asymptotic densities of $n\left(\widehat{\beta}_n - \beta_n\right)$ *and* $n^{-1/2}\left(\widehat{\pi}_{1,n} - \pi_{1,n}\right)$ *given the true values of* $\{b, c_0, c_1\}$. *The asymptotic approximations based on Theorem 2 fit the finite-sample densities very well.*

2.4 Drifting Sequence in Andrews and Cheng (2012)

We conclude this section by discussing the differences between the asymptotic approaches in this paper and Andrews and Cheng (2012). In the models considered by Andrews and Cheng (2012), the parameters of interest are $\{\beta, \zeta, \pi\}$, in which β and ζ are always identified and can be $\sqrt{n}$-consistently estimated regardless of the value of π, and π is identified if and only if $\beta \neq 0$ and the estimator for π may weakly converge to a nondegenerate random variable when $\beta \approx 0$. To match the convergence rate they employed the drifting sequence $n^{1/2}\beta_n \to b$ in their weak-identification scenario, and the sequence $n^{1/2}\beta_n \to \infty$ in their semi-strong-identification case. However, for the problem considered in this paper, in Lemma 1 we have already shown that when $\beta_n = 0$, $\widehat{\beta}_n - \beta_n$ and $\widehat{\pi}_n$ are respectively $O_p\left(n^{-1}\right) / O_p\left(n^{1/2}\right)$. Due to the difference in the convergence rates of estimators, we consider $\Gamma_n(1, b, \mathbf{c})$, in which $n\beta_n = b$ and $n^{-1/2}\pi_n = \mathbf{c}$ to match the convergence / divergence rates, and use $\Gamma_n(h, b, \mathbf{c})$, *i.e.*, $n^h\beta_n = b$ and $n^{1/2-h}\pi_n = \mathbf{c}$, to bridge $\{\theta_n = \theta_0 \in \Theta_n^*\}$ and $\Gamma_n(1, b, \mathbf{c})$. In Theorems 2 and 3 we have already shown the necessity of the drifting-to-infinity assumption for inference about π_n.

The drifting sequences considered in Andrews and Cheng (2012) in their weak-identification scenario reduces to the case when $\gamma_n \in \Gamma_n(h, b, \mathbf{c})$ with $h = 1/2$, in which β_n is a sequence drifting to zero with a standardization factor $n^{-1/2}$ ($n^{1/2}\beta_n = b$) and π_n is a constant vector ($\pi_n = \mathbf{c}$). We have already shown (in Theorem 3) that when $\gamma_n \in \Gamma_n(h, b, \mathbf{c})$, the NLS estimator $\widehat{\pi}_n$ is asymptotically Gaussian distributed when $b \neq 0$, and is unidentifiable when $b = 0$ since $\mathrm{Avar}(\widehat{\pi}_n) = \sigma_\varepsilon^2 b^{-2}\mathbf{M}_X^{-1} \to \infty$ when $b \to 0$. For the problem considered in this paper, if one consider the drifting sequences of Andrews and Cheng (2012), the desired smooth transition of the asymptotic approximation will be missing, due to the insufficient standardization factors not matching the convergence / divergence rates of the NLS estimator when $\beta_n = 0$.

3 Confidence Sets and Tests

In the section we establish the limiting properties of the t / Wald test statistics and introduce the procedure to obtain the confidence sets with correct asymptotic sizes for specific linear functions of parameters of interest. Consider a linear null statistical hypothesis:

$$H_0 : \mathbf{R}\theta_n = \upsilon, \tag{3.1}$$

where $\mathbf{R} \in \mathbb{R}^{d_r \times (d_\pi + 1)}$, $\upsilon \in \mathbb{R}^{d_r}$ where $d_r \leq d_\pi + 1$, and $\text{Rank}\left(\mathbf{R}\right) = d_r$.

3.1 t and Wald Test Statistics

Consider the t statistics $T_n\left(\upsilon\right)$ (when $d_r = 1$) and the Wald statistics $W_n\left(\upsilon\right)$ corresponding to the null (equation (3.1)):

$$\begin{aligned} T_n\left(\upsilon\right) &= \frac{n^{1/2}\left[\mathbf{R}\widehat{\theta}_n - \upsilon\right]}{\left[\widehat{\sigma}_n^2 \mathbf{R}\widehat{\mathbf{V}}_n^{-1}\mathbf{R}^\top\right]^{1/2}}, && (3.2) \\ W_n\left(\upsilon\right) &= n\left[\mathbf{R}\widehat{\theta}_n - \upsilon\right]^\top \left[\widehat{\sigma}_n^2 \mathbf{R}\widehat{\mathbf{V}}_n^{-1}\mathbf{R}^\top\right]^{-1}\left[\mathbf{R}\widehat{\theta}_n - \upsilon\right], && (3.3) \end{aligned}$$

where $\widehat{\sigma}_n^2 = n^{-1}\sum_{t=1}^n \widehat{\varepsilon}_t^2$ is defined in equation (2.8), and

$$\widehat{\mathbf{V}}_n = n^{-1}\sum_{t=1}^n \left[\begin{array}{cc} \left(y_{t-1} - X_t^\top \widehat{\pi}_n\right)^2 & -\widehat{\beta}_n\left(y_{t-1} - X_t^\top\widehat{\pi}_n\right)X_t^\top \\ -\widehat{\beta}_n X_t\left(y_{t-1} - X_t^\top\widehat{\pi}_n\right) & \widehat{\beta}_n^2 X_t X_t^\top \end{array}\right]. \tag{3.4}$$

In the results below we provide the asymptotic properties of the t / Wald test statistics under the null. Consider the sequence of the null hypotheses $H_0 : \mathbf{R}\theta_n = \upsilon_n$, where υ_n is the true value of $\mathbf{R}\theta_n$. $T_n\left(\upsilon_n\right)$ / $W_n\left(\upsilon_n\right)$ are denoted as T_n / W_n for notational simplicity.

Theorem 4 *Suppose that Assumptions 1, 2 and 3 hold.*

1. *When $\theta_n = \theta_0 \in \Theta_n^*$, i.e., $\beta_n = \beta_0$ and $\pi_n = \pi_0$ for any $n \in \mathbb{N}$, $T_n \overset{A}{\sim} \mathcal{N}\left(0,1\right)$, and $W_n \overset{A}{\sim} \chi^2\left(d_r\right)$.*

2. *When $\gamma_n \in \Gamma_n\left(1, b, \mathbf{c}\right)$ i.e., $\beta_n = b/n$ with $0 < b < \infty$ and $\pi_n = n^{1/2}\mathbf{c}$, and $y_0 =$*

$o_p\left(n^{1/2}\right)$,

$$\begin{aligned} T_n &\Rightarrow \mathcal{T}\left(b,\mathbf{c};\varphi_0\right)=\frac{\mathbf{R}\widehat{\tau}\left(b,\mathbf{c};\varphi_0\right)}{\left[\sigma_\varepsilon^2\mathbf{R}\mathcal{V}_1^{-1}\left(b,\mathbf{c};\varphi_0\right)\mathbf{R}^\top\right]^{1/2}}, \\ W_n &\Rightarrow \mathcal{W}\left(b,\mathbf{c};\varphi_0\right)=\left[\mathbf{R}\widehat{\tau}\left(b,\mathbf{c};\varphi_0\right)\right]^\top\left[\sigma_\varepsilon^2\mathbf{R}\mathcal{V}_1^{-1}\left(b,\mathbf{c};\varphi_0\right)\mathbf{R}^\top\right]^{-1}\mathbf{R}\widehat{\tau}\left(b,\mathbf{c};\varphi_0\right), \end{aligned}$$

where $\mathcal{V}_1\left(b,\mathbf{c};\varphi_0\right)$ is defined in Theorem 6 in Appendix A, and $\widehat{\tau}\left(b,\mathbf{c};\varphi_0\right)$ are defined in Theorem 2.

3. *When $\gamma_n\in\Gamma_n\left(h,b,\mathbf{c}\right)$, i.e., $\beta_n=b/n^h$ with $0<b<\infty$ and $\pi_n=n^{-1/2+h}\mathbf{c}$, where $0<h<1$, $T_n\overset{A}{\sim}\mathcal{N}\left(0,1\right)$, and $W_n\overset{A}{\sim}\chi^2\left(d_r\right)$.*

Remark 3 1. *In Theorem 4 we obtain the asymptotic distribution of the t / Wald statistics for all three cases we consider. When $\theta_n=\theta_0\in\Theta_n^*$ or $\gamma_n\in\Gamma_n\left(h,b,\mathbf{c}\right)$, T_n / W_n have the standard and pivotal asymptotic Gaussian / χ^2 distributions. However, when $\gamma_n\in\Gamma_n\left(1,b,\mathbf{c}\right)$, the asymptotic distribution of the T_n / W_n will depend on $\widehat{\tau}\left(b,\mathbf{c};\varphi_0\right)$ and $\mathcal{V}\left(b,\mathbf{c};\varphi_0\right)$, which themselves are functionals of the Ornstein–Uhlenbeck process we define in Lemma 2 and depend on the values of unknown nuisance parameters $\varphi_0=\{\mu_X,\mathbf{M}_X,\sigma_\varepsilon^2\}$ and localization parameters $\{b,\mathbf{c}\}$.*

2. *Similar to Mikusheva (2012), in this paper we only consider linear null hypotheses $H_0:\mathbf{R}\theta_n=\upsilon$. For the nonlinear null hypothesis, e.g., $H_0:r\left(\theta_n\right)=\upsilon$ with a differentiable function $r:\mathbb{R}^{(d_\pi+1)}\to\mathbb{R}^{d_r}$, econometricians usually use the delta method to approximate the asymptotic variance of $r\left(\theta_n\right)$ by $\widehat{\sigma}_n^2R^\top\left(\widehat{\theta}_n\right)\widehat{\mathbf{V}}_n^{-1}R\left(\widehat{\theta}_n\right)$, where $R\left(\theta\right)=D_\theta r\left(\theta\right)$ is the derivative of $r\left(\theta\right)$. When $\widehat{\theta}_n$ is a consistent estimator for θ_n, by the continuous mapping theorem, $R\left(\widehat{\theta}_n\right)\xrightarrow{p}R\left(\theta_n\right)$. For the problem we consider, however, we have shown that when $\gamma_n\in\Gamma_n\left(1,b,\mathbf{c}\right)$ or $\gamma_n\in\Gamma_n\left(h,b,\mathbf{c}\right)$ with $h\geq1/2$, $\widehat{\pi}_n$ is not a consistent estimator for π_n, and therefore the bias of $R\left(\widehat{\theta}_n\right)$ is not negligible. For inference of nonlinear functions, one may consider the parametric bootstrapping (Krinsky and Robb, 1986) or the confidence interval bootstrapping (Woutersen and Ham, 2013).*

Again, when $\gamma_n\in\Gamma_n\left(1,b,\mathbf{c}\right)$, the asymptotic distributions of T_n / W_n depend on unknown nuisance parameters $\varphi_0=\{\mu_X,\mathbf{M}_X,\sigma_\varepsilon^2\}$. In Lemma 3 we have already shown that the unknown nuisance parameters $\varphi_0=\{\mu_X,\mathbf{M}_X,\sigma_\varepsilon^2\}$ can be consistently estimated by $\widehat{\varphi}_n=\left\{\widehat{\mu}_{X,n},\widehat{\mathbf{M}}_{X,n},\widehat{\sigma}_n^2\right\}$. Therefore, for any given values of the localization parameters $\{b,\mathbf{c}\}$, the asymptotic distributions of T_n / W_n can be obtained by replacing the unknown nuisance parameters φ_0 with the estimates $\widehat{\varphi}_n$.

Example 3 (Example 1 continued) *Again, consider the following model as equation (2.9).*

$$y_t = (1-\beta_n)\, y_{t-1} + \beta_n\,(\pi_{0,n} + \pi_{1,n}x_t) + \varepsilon_t, \quad t = 1, \ldots, n,$$

where $x_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, $\beta_n = b\,/n$, $\pi_{0,n} = n^{1/2}c_0$, $\pi_{1,n} = n^{1/2}c_1$, *and* $n = 100$. *Let* T_n / W_n *denote the* t / *Wald statistics respectively corresponding to* $H_0 : \beta = \beta_n$ *and* $H_0 : \pi_1 = \pi_{1,n}$, *where* β_n *and* $\pi_{1,n}$ *denote the true values of* β *and* π_1.

Using Theorem 4, Figures 7 – 10 provide the simulated finite-sample and asymptotic densities of T_n / W_n *given the true values of* $\{b, c_0, c_1\}$. *We consider* $\beta_n \in \{0.02, 0.05, 0.1\}$ *and* $\pi_{0,n} = \pi_{1,n} = 2$, *i.e.,* $b \in \{2, 5, 10\}$ *and* $c_0 = c_1 = 0.2$. *The asymptotic approximations based on Theorem 4 fit the finite-sample densities very well.*

3.2 Robust Confidence Sets

In this subsection we obtain the confidence sets (CS) of $\mathbf{R}\theta_n$ by inverting the t / Wald tests. In the following we focus on the two-sided confidence intervals based on the Wald tests. The one-sided / two-sided confidence intervals based on the t tests are analogous.

In Theorem 4 we have already show that when $\theta_n = \theta_0 \in \Theta_n^*$ or $\gamma_n \in \Gamma_n(h, b, \mathbf{c})$, the Wald statistics is pivotally asymptotically $\chi^2(d_r)$-distributed. And when $\gamma_n \in \Gamma_n(1, b, \mathbf{c})$, the Wald statistics has a nonstandard and non-pivotal asymptotic distribution depending on the values of the localization parameters $\{b, \mathbf{c}\}$. Without any prior information about which category the parameters γ_n belongs to, a conservative and robust confidence set (CS_n^R) is defined as a union of CS_n^L, the CS when $\gamma_n \in \Gamma_n(1, b, \mathbf{c})$, and CS_n^D, the CS when $\theta_n = \theta_0 \in \Theta_n^*$ or $\gamma_n \in \Gamma_n(h, b, \mathbf{c})$.

$$CS_n^R(\mathbf{R}\theta_n; 1-\alpha, \varphi_0) = CS_n^L(\mathbf{R}\theta_n; 1-\alpha, \varphi_0) \cup CS_n^D(\mathbf{R}\theta_n; 1-\alpha). \tag{3.5}$$

$1-\alpha$ denotes the confidence coefficient. L and D respectively represent 'local-to-zero β_n' and 'distant-from-zero β_n'.

CS_n^D is obtained by simply inverting the Wald test. Let $\chi^2_{d_r, 1-\alpha}$ be the $(1-\alpha)$-quantile of $\chi^2(d_r)$. Since the Wald statistics is pivotally asymptotically $\chi^2(d_r)$-distributed when

$\theta_n = \theta_0 \in \Theta_n^*$, or $\gamma_n \in \Gamma_n(h, b, \mathbf{c})$,

$$CS_n^D(\mathbf{R}\theta_n; 1-\alpha) = \left\{ \upsilon : W_n(\upsilon) \leq \chi^2_{d_r,1-\alpha} \right\}. \tag{3.6}$$

For CS_n^L, *i.e.*, the CS when $\gamma_n \in \Gamma_n(1, b, \mathbf{c})$, first we consider a simple case. Since $\mathbf{R}\theta_n = \mathbf{R}\left[\beta_n, \pi_n^\top\right]^\top = \mathbf{R}\left[n^{-1}b, n^{1/2}\mathbf{c}^\top\right]^\top$ when $\gamma_n \in \Gamma_n(1, b, \mathbf{c})$, let $\mathcal{H}(\mathbf{R}, \upsilon)$ be the 'null-imposition set' for the localization parameters $\{b, \mathbf{c}\}$, which contains all possible values of the localization parameters $\{b, \mathbf{c}\}$ under the null $\mathbf{R}\theta_n = \upsilon$.

$$\mathcal{H}(\mathbf{R}, \upsilon) = \left\{ b, \mathbf{c} : \mathbf{R}\left[n^{-1}b, n^{1/2}\mathbf{c}^\top\right]^\top = \upsilon, \quad \left\{n^{-1}b, n^{1/2}\mathbf{c}^\top\right\} \in \Theta_n \right\}. \tag{3.7}$$

The definition of the null-imposition set $\mathcal{H}(\mathbf{R}, \upsilon)$ is similar to equation (5.2) in Andrews and Cheng (2012). If $\mathcal{H}(\mathbf{R}, \upsilon)$ is a singleton for every υ, CS_n^L can also be obtained by simply inverting the Wald test. Let $\xi_{1-\alpha}(\mathcal{W}(b_\upsilon, \mathbf{c}_\upsilon; \varphi_0))$ be the $(1-\alpha)$-quantile of $\mathcal{W}(b_\upsilon, \mathbf{c}_\upsilon; \varphi_0)$, and $\mathbf{R}\left[n^{-1}b_\upsilon, n^{1/2}\mathbf{c}_\upsilon^\top\right]^\top = \upsilon$. When $n \to \infty$,

$$CS_n^L(\mathbf{R}\theta_n; 1-\alpha, \varphi_0) = \left\{ \upsilon : W_n(\upsilon) \leq \xi_{1-\alpha}(\mathcal{W}(b_\upsilon, \mathbf{c}_\upsilon; \varphi_0)) \right\}. \tag{3.8}$$

For example, suppose that $\mathbf{R} = \mathbb{I}_{d_\pi+1}$ and $\mathbf{R}\theta_n = \theta_n$, *i.e.*, we are interested in the confidence set of θ_n. Since $\theta_n = \{\beta_n, \pi_n\} = \left\{n^{-1}b, n^{1/2}\mathbf{c}\right\}$, for any given null hypothesis $H_0 : \theta_n = \upsilon_\theta$, the values of the localization parameters are available under the null hypothesis. Therefore the asymptotic distribution of the Wald statistics is also available. When $\mathcal{H}(\mathbf{R}, \upsilon)$ is a singleton, the robust CS is defined as:

$$\begin{aligned} CS_n^R(\mathbf{R}\theta_n; 1-\alpha, \varphi_0) &= \left\{ \upsilon : W_n(\upsilon) \leq c^R(\mathbf{R}\theta_n; 1-\alpha, \varphi_0) \right\} \\ &= \left\{ \upsilon : W_n(\upsilon) \leq \max\left\{ \chi^2_{d_r,1-\alpha}, \xi_{1-\alpha}(\mathcal{W}(b_\upsilon, \mathbf{c}_\upsilon; \varphi_0)) \right\} \right\}. \end{aligned} \tag{3.9}$$

However, the null-imposition set $\mathcal{H}(\mathbf{R}, \upsilon)$ may not be a unit set. For example, suppose $\mathbf{R}\theta_n = \pi_n$, *i.e.*, we are only interested in the confidence set of π_n. Then for any given null hypothesis $H_0 : \pi_n = \upsilon_\pi$, even though the value of $\mathbf{c} = n^{-1/2}\pi_n$ is available under the null, the value of $b = n\beta_n$ is still unknown. Since the asymptotic null distribution of the Wald statistics depends on the value of b, we are not able to determine the asymptotic null distribution, and the corresponding CS.

For the case when $\gamma_n \in \Gamma_n(1, b, \mathbf{c})$ and the null-imposition set $\mathcal{H}(\mathbf{R}, \upsilon)$ is not a unit set, we consider two different methods to obtain the CS, the null-imposed least-favorable method

(Andrews and Cheng, 2012) and the projection-based method (Dufour, 1997). The null-imposed least-favorable method establishes the confidence set $CS_n^{L,LF}$ by selecting $\{b_\upsilon, \mathbf{c}_\upsilon\}$ with the greatest critical value among $\mathcal{H}(\mathbf{R}, \upsilon)$. When $n \to \infty$,

$$CS_n^{L,LF}(\mathbf{R}\theta_n; 1-\alpha, \varphi_0) = \left\{ \upsilon : W_n(\upsilon) \leq \sup_{\{b,\mathbf{c}\} \in \mathcal{H}(\mathbf{R},\upsilon)} \xi_{1-\alpha}(\mathcal{W}(b, \mathbf{c}; \varphi_0)) \right\}. \tag{3.10}$$

For example, in the case when $\mathbf{R}\theta_n = \pi_n$, since the value of $b = n\beta_n$ is unknown under the null hypotheses, the null-imposed least-favorable method constructs the $CS_n^{L,LF}$ by selecting the value of b maximizing $\xi_{1-\alpha}(\mathcal{W}(b, \mathbf{c}; \varphi_0))$. $CS_n^{L,LF}$ is conservative since the greatest critical value is used. The robust CS based on the null-imposed least-favorable method is defined as:

$$\begin{aligned} CS_n^{R,LF}(\mathbf{R}\theta_n; 1-\alpha, \varphi_0) &= \left\{ \upsilon : W_n(\upsilon) \leq c^{R,LF}(\mathbf{R}\theta_n; 1-\alpha, \varphi_0) \right\} \\ &= \left\{ \upsilon : W_n(\upsilon) \leq \max\left\{ \chi^2_{d_r,1-\alpha}, \sup_{\{b,\mathbf{c}\} \in \mathcal{H}(\mathbf{R},\upsilon)} \xi_{1-\alpha}(\mathcal{W}(b_\upsilon, \mathbf{c}_\upsilon; \varphi_0)) \right\} \right\}. \end{aligned} \tag{3.11}$$

The projection-based method establishes the confidence set $CS_n^{L,P}$ by projecting an $(d_\pi + 1)$-sphere to the $\mathbb{R}^{d_r}$-space. For any given null hypothesis $H_0 : \mathbf{R}\theta_n = \upsilon$, let $\mathbf{R} = \mathbf{PQ}$, where $\mathbf{P} \in \mathbb{R}^{d_r \times (d_\pi+1)}$, $\mathbf{Q} \in \mathbb{R}^{(d_\pi+1)\times(d_\pi+1)}$, $\operatorname{rank}(\mathbf{P}) = d_r$ and $\operatorname{rank}(\mathbf{Q}) = d_\pi + 1$. The the null hypothesis $H_0 : \mathbf{R}\theta_n = \upsilon$ can be written as

$$H_0 : \mathbf{PQ}\theta_n = \mathbf{P}\varpi. \tag{3.12}$$

The matrices $\mathbf{P}$ and $\mathbf{Q}$ always exist since one can always select $\{\mathbf{P}, \mathbf{Q}\} = \{\mathbf{R}, \mathbb{I}_{d_\pi+1}\}$. Let

$$\mathcal{H}(\mathbf{Q}, \varpi) = \left\{ b, \mathbf{c} : \mathbf{Q}\left[n^{-1}b, n^{1/2}\mathbf{c}^\top\right]^\top = \varpi, \quad \left\{n^{-1}b, n^{1/2}\mathbf{c}^\top\right\} \in \Theta_n \right\}. \tag{3.13}$$

By $\operatorname{rank}(\mathbf{Q}) = d_\pi + 1$, $\mathcal{H}(\mathbf{Q}, \varpi)$ is a singleton. The CS for $\mathbf{Q}\theta_n$ can be obtained by equation (3.8).

$$CS_n^L(\mathbf{Q}\theta_n; 1-\alpha, \varphi_0) = \left\{ \varpi : W_n(\varpi) \leq \xi_{1-\alpha}(\mathcal{W}(b_\varpi, \mathbf{c}_\varpi; \varphi_0)) \right\}.$$

The confidence set $CS_n^{L,P}$ is established by projecting CS_n^L, an $(d_\pi + 1)$-sphere, to the $\mathbb{R}^{d_r}$-space.

$$\begin{aligned} CS_n^{L,P}(\mathbf{R}\theta_n; 1-\alpha, \varphi_0) &= \mathbf{P}CS_n^L(\mathbf{Q}\theta_n; 1-\alpha, \varphi_0) \\ &= \left\{ \upsilon : \upsilon = \mathbf{P}\varpi, \quad W_n(\varpi) \leq \xi_{1-\alpha}(\mathcal{W}(b_\varpi, \mathbf{c}_\varpi; \varphi_0)) \right\} \end{aligned} \tag{3.14}$$

In practice, a simple choice for the matrices $\{\mathbf{P}, \mathbf{Q}\}$ is $\{\mathbf{R}, \mathbb{I}_{d_\pi+1}\}$, the matrix $\mathbf{R}$ itself and the identity matrix $\mathbb{I}_{d_\pi+1}$. For example, in the case when $\mathbf{R}\theta_n = \pi_n$, the projection-based method constructs the $CS_n^{L,P}$ by projecting the CS of θ_n to the $\mathbb{R}^{d_\pi}$-space. $CS_\infty^{L,P}$ is also conservative since for any set $\mathcal{C} \subset \mathbb{R}^{(d_\pi+1)}$, the event $\{\mathbf{Q}\theta_n \in \mathcal{C}\}$ entails $\{\mathbf{PQ}\theta_n \in \mathbf{P}\mathcal{C}\}$. However, intuitively the projection-based method uses the information from the estimates for all parameters of interest, and it is possible to obtain a more informative but still conservative confidence set compared to the null-imposed least-favorable one under certain circumstances. The robust CS based on the projection-based method is defined as:

$$\begin{aligned} CS_n^{R,P}(\mathbf{R}\theta_n; 1-\alpha, \varphi_0) &= \mathbf{P}CS_n^R(\mathbf{Q}\theta_n; 1-\alpha, \varphi_0) \\ &= \left\{\upsilon : \upsilon = \mathbf{P}\varpi, \quad W_n(\varpi) \leq \max\left\{\chi^2_{d_\pi+1,1-\alpha}, \xi_{1-\alpha}(\mathcal{W}(b_\varpi, \mathbf{c}_\varpi; \varphi_0))\right\}\right\} \end{aligned} \tag{3.15}$$

For any finite-sample confidence set CS_n, the asymptotic size ($AsySz$) approximates the smallest finite-sample coverage probability.

$$AsySz(CS_n) = \liminf_{n\to\infty} \inf_{\gamma_n \in \Gamma_n} \mathbb{P}(\mathbf{R}\theta_n \in CS_n). \tag{3.16}$$

Notice that in the definition of the asymptotic size (equation (3.16)) $\liminf_{n\to\infty}$ is taken before $\inf_{\gamma_n\in\Gamma}$, i.e., the asymptotic size is defined as the probability limit (as $n \to \infty$) of the infimum of the exact finite-sample coverage probability. This definition reflects the fact that we are interested in the exact coverage probability, and asymptotic coverage probability is simply used to approximate the exact one. Since the exact finite-sample coverage probability are unavailable, in the following Theorem 5 we show that we can exchange $\liminf_{n\to\infty}$ and $\inf_{\gamma_n\in\Gamma}$. That is, we show that the asymptotic size can be obtain by taking the infimum of the asymptotic coverage probability. Similar arguments can be found in Andrews and Cheng (2012), Guggenberger (2012), Li (2013), Mikusheva (2007, 2012) and many others. Theorem 5 shows the correctness of the asymptotic sizes of CS_n^R and $CS_n^{R,LF}$. The projection-based $CS_n^{R,P}$, however, may be asymptotic oversized, *i.e.*, may have an asymptotic size higher than the required confidence coefficient $1-\alpha$.

Theorem 5 *Suppose that Assumptions 1, 2 and 3 hold and $y_0 = o_p(n^{1/2})$ when $\gamma_n \in \Gamma_n(1, b, \mathbf{c})$.*

1. *When the null-imposition set $\mathcal{H}(\mathbf{R}, \upsilon)$ is a singleton for every υ, $AsySz\left(CS_n^R(\mathbf{R}\theta_n; 1-\alpha, \varphi_0)\right) = 1-\alpha$.*
2. *$AsySz\left(CS_n^{R,LF}(\mathbf{R}\theta_n; 1-\alpha, \varphi_0)\right) = 1-\alpha$.*

3. $AsySz\left(CS_n^{R,P}\left(\mathbf{R}\theta_n; 1-\alpha, \varphi_0\right)\right) \geq 1-\alpha.$

Remark 4 1. *As in Andrews and Cheng (2012, 2013a, 2013b), we obtain the CS by inverting the tests. One may consider to obtain the CS directly from the asymptotic distributions of* $\widehat{\theta}_n$*, as in Mikusheva (2012). However, we have already shown that when* $\gamma_n \in \Gamma_n(h, b, \mathbf{c})$*, though the asymptotic distributions of* $\widehat{\theta}_n$ *will depend on the unknown values of* $\{h, b, \mathbf{c}\}$*, the t / Wald statistics will have standard and pivotal asymptotic distributions. Therefore, to consider the t / Wald statistics is much simpler then considering the estimates* $\widehat{\theta}_n$*.*

2. *By virtue of our linear drifting sequence approaches, as in Stock (1991), there is a surjective mapping from the values of localization parameters* $\{b, \mathbf{c}\}$ *to the null hypotheses corresponding to the tests to be inverted in obtaining the CS. Therefore, the null-imposed least-favorable method takes the supremum of the critical values of tests only with respect to the possible values of* $\{b, \mathbf{c}\}$ *in* $\mathcal{H}(\mathbf{R}, \upsilon)$*. Without the onto mapping, e.g., if we simply assume* $n^{-1/2}\pi_n \to \mathbf{c}$*, the simple least-favorable method would take the supremum w.r.t. all possible values of the* $\{b, \mathbf{c}\}$ *in the parameter space* Θ_n*. A wider and less informative confidence set may be obtained.*

3. *For the projection-based method, again, under our linear drifting sequence approaches, a confidence set for* $\mathbf{Q}\theta_n$ *is directly available since the null-imposition set* $\mathcal{H}(\mathbf{Q}, \varpi)$ *is a unit set. For example, when* $\mathbf{Q} = \mathbb{I}_{d_\pi+1}$*, we are able to construct a confidence set for* θ_n *since the values of* $\{b, \mathbf{c}\}$ *are known under the null hypotheses of the tests to be inverted. Without the onto mapping, the confidence set for* $\mathbf{Q}\theta_n$ *will not be directly available.*

4. *In the case when the CS of* π_n *is interested, the null-imposed least-favorable selects the value of* $b = n\beta_n$ *maximizing the critical values of the Wald tests to be inverted. Since the null-imposed least-favorable CS* ($CS_n^{L,LF}$) *may be very large, McCloskey (2011) proposed a Bonferroni-based size-correction method to obtain a more informative CS. When the parameters of interest are* $\{\beta_n, \pi_n\}$ *and* π_n *is not identified if and only if* $\beta_n = 0$*, McCloskey (2011) suggested to obtain a CS for* β_n *at first, and to select the value of the localization parameter corresponding to* β_n *within the obtained CS. However, for the problem considered in this paper, Theorem 2 shows that the asymptotic distribution of* $\widehat{\beta}_n$ *will depend on not only* b *but also* $\mathbf{c}$*, i.e., the localization parameters corresponding to the true values of* $\{\beta_n, \pi_n\}$*. Therefore it is not feasible to construct a*

CS with correct asymptotic coverage probability of β_n without any information of the true value of π_n.

Again, when $\gamma_n \in \Gamma_n(1, b, \mathbf{c})$, the CS of $\mathbf{R}\theta_n$ depends on unknown nuisance parameters $\varphi_0 = \{\mu_X, \mathbf{M}_X, \sigma_\varepsilon^2\}$. Since the nuisance parameters φ_0 can be consistently estimated by $\widehat{\varphi}_n = \left\{\widehat{\mu}_{X,n}, \widehat{\mathbf{M}}_{X,n}, \widehat{\sigma}_n^2\right\}$. Therefore, the CS can be obtained by replacing φ_0 with $\widehat{\varphi}_n$.

Example 4 (Example 1 continued) *Again, consider the following model as equation (2.9).*

$$y_t = (1 - \beta_n)\, y_{t-1} + \beta_n \left(\pi_{0,n} + \pi_{1,n} x_t\right) + \varepsilon_t, \quad t = 1, \ldots, n,$$

where $x_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, $\beta_n \in [0.02, 0.5]$, $\pi_{0,n} \in [0,2]$, $\pi_{1,n} \in [0,2]$, and $n = 100$. In this example we construct the CS for $\mathbf{R}\theta_n = \pi_{1,n}$ with $1 - \alpha = 0.8$ and 0.9 by the null-imposed least-favorable method ($CS_n^{L,LF}$) and the projection-based method ($CS_n^{L,P}$) and compare the CSs with the CS from the standard (Newey and McFadden, 1994) based on the $\chi^2(1)$ distribution.

Figures 11 and 12 provide the simulated coverage probabilities (CP) of the three CSs when $\pi_{1,n} = 2$. For both cases $1 - \alpha = 0.8$ and 0.9 and for every values of β_n and $\pi_{0,n}$, $CS_n^{L,LF}$ and $CS_n^{L,P}$ have CPs greater than the confidence coefficient $1 - \alpha$, while the CPs of the $\chi^2(1)$ CS are seriously downward biased, especially when β_n is close to zero. Under most circumstances $CS_n^{L,LF}$ has coverage probabilities closer to $1 - \alpha$, especially when β_n is close to zero. However, when β_n is distant from zero, $CS_n^{L,P}$ may have better coverage probabilities.

For all results $5,000$ simulation repetitions are used. For values of β_n and $\pi_{0,n}$ / $\pi_{1,n}$, $11,466$ grids are generated in the true parameter space $\Theta_n^* = [0, 0.5] \times [0, 4]^2$, where grids for β_n and $\pi_{0,n}$ / $\pi_{1,n}$ are respectively of size 0.02 and 0.2.

4 Empirical Application: U.S.'s Forecast-Based MPRF

According to our asymptotic theory, we construct the conservative confidence sets for the reaction coefficients $\{\pi_{\dot{p}}, \pi_x\}$ in U.S.'s forecast-based $MPRF$ and examine if $\{\pi_{\dot{p}}, \pi_x\}$ belong to the determinacy region $\mathcal{DR} = \{\pi_{\dot{p}} > 1, \pi_x > 0\}$. When $\pi_{\dot{p}} > 1$ and $\pi_x > 0$, regardless of the values of other unknown parameters, the $MPRF$ sufficiently satisfies the determinacy condition, *i.e.*, the monetary authority adjusts the nominal interest rates with 'sufficient strength' in response to inflations and output gaps (Woodford, 2003; Galí, 2008)

In the NLS estimation we use the Greenbook projections, *i.e.*, the real-time data of the *ex ante* forecasts for inflations and the expected output gaps of the Federal Reserve. The real-time data is available in the Federal Reserve Bank of Philadelphia for 1987:3–2007:4, *i.e.*, $n = 82$. As in Nikolsko-Rzhevskyy (2011), we consider the following model with $k = 0$ or 1.

$$i_t = (1-\beta)\, i_{t-1} + \beta\left(\pi_\alpha + \pi_{\dot{p}} \mathbb{E}_t \dot{p}_{t,k} + \pi_x \mathbb{E}_t x_{t,k}\right) + \varepsilon_t, \tag{4.1}$$

where $\{i_t\}$ denote the average of effective federal funds target rates at the last month of each quarter, and $\{\mathbb{E}_t \dot{p}_{t,k}, \mathbb{E}_t x_{t,k}\}$ are Greenbook projections for the annualized inflations and the average output gaps between periods t and $t+k$. $t = 1, \ldots, 82$. Figure 13 provides the plots of the data. Table 1 reports the NLS estimates, where in the parentheses we report the estimates of standard errors according to Equation (3.4).

Let $\beta = \beta_n = n^{-1}b$ and $\pi_\alpha = \pi_{\alpha,n} = n^{1/2}c_\alpha$. The null-imposed least-favorable CS ($CS_n^{L,LF}$) of $\{\pi_{\dot{p}}, \pi_x\}$ is obtained by selecting the values of b and c_α maximizing the critical values of the Wald tests corresponding to different values of $\{\pi_{\dot{p}}, \pi_x\}$. Figure 14 reports the $CS_n^{L,LF}$ and the standard CS based on $\chi^2(2)$ distribution. For both cases for $k = 0$ and 1, the $CS_n^{L,LF}$ with confidence coefficients $1 - \alpha = 0.8$, 0.9 and 0.95 contain many values not in the region $\mathcal{DR} = \{\pi_{\dot{p}} > 1, \pi_x > 0\}$.

As a robustness check, we also construct the projection-based CS ($CS_n^{L,P}$) of $\{\pi_{\dot{p}}, \pi_x\}$, which are obtained by projecting the CS of all parameters of interest from the $\mathbb{R}^4$ space to the $\mathbb{R}^2$ space for $\{\pi_{\dot{p}}, \pi_x\}$. Figure 15 reports the $CS_n^{L,P}$s. For all cases $CS_n^{L,P}$ are even larger than $CS_n^{L,LF}$ and contain even more values not in $\mathcal{DR}$. Our empirical results suggest that the NLS estimates for the reaction coefficients are not accurate sufficiently to rule out the possibility of indeterminacy.

For all results $5,000$ simulation repetitions are used. For values of parameters, $45,056$ grids are generated in the true parameter space $\Theta_n^* = [0, 0.2] \times [-1, 2]^3$, where grids for β and π_α / $\pi_{\dot{p}}$ / π_x are respectively of size 0.02 and 0.2.

5 Concluding Remarks

In this paper we modify the method of Andrews and Cheng (2012) on inference with weak / semi-strong identification and establish the asymptotic distributions of the NLS estimator / tests for the forecast-based monetary policy reaction function ($MPRF$) with a close-to-unity smoothing coefficient. Conservative confidence sets with correct / over asymptotic coverage probability for linear functions of parameters are obtained by the null-imposed least-favorable method ($NILF$) and the projection-based method. Our empirical result suggests that the NLS estimates for the reaction coefficients are not accurate sufficiently to rule out the possibility of indeterminacy for U.S.'s forecast-based $MPRF$ for 1987:3–2007:4.

6 Appendix A: Details of Lemma 2 and Theorem 4

In this section we provide details of Lemma 2 and Theorem 4. Proofs are collected in Appendix B.

Lemma 5 *(Lemma 2) Suppose that Assumptions 1, 2 and 3 hold, $\gamma_n \in \Gamma_n(1, b, \mathbf{c})$, and $y_0 = o_p\left(n^{1/2}\right)$. Let $\mathcal{Z}$ be a standard-normally distributed random variable, $\mathcal{W}_\varepsilon(\cdot)$ be a standard Wiener processes and $\mathcal{J}_{-b,\varepsilon}(\cdot)$ be an Ornstein–Uhlenbeck process such that for any $r \in [0,1]$, when $n \to \infty$,*

$$n^{-1/2}\sum_{t=1}^{n} X_t\varepsilon_t \Rightarrow \sigma_\varepsilon \mathbf{M}_X^{1/2}\mathcal{Z}, \quad n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor}\varepsilon_t \Rightarrow \sigma_\varepsilon \mathcal{W}_\varepsilon(r), \quad \textit{and}$$

$$n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor}\left(1-\frac{b}{n}\right)^{\lfloor nr\rfloor - t}\varepsilon_t \Rightarrow \mathcal{J}_{-b,\varepsilon}(r) = \int_0^r \exp\left(-b(r-s)\right) d\mathcal{W}_\varepsilon(s).$$

Then for any $\mathbb{R}^{d_\pi}$-valued π with $n^{-1/2}\pi \Rightarrow \kappa_\pi$ as $n \to \infty$,

1. *$\left(\partial Q_n(0,\pi;\gamma_n)\right)/\partial\beta \Rightarrow \mathcal{G}(\kappa_\pi, b, \mathbf{c}; \varphi_0)$, where*

$$\begin{aligned}
&\mathcal{G}(\kappa_\pi, b, \mathbf{c}; \varphi_0)\\
=\;& \sigma_\varepsilon^2\int_0^1 \mathcal{J}_{-b,\varepsilon}(r)\, d\mathcal{W}_\varepsilon(r) + \sigma_\varepsilon\left(\int_0^1 (1-\exp(-br))\, d\mathcal{W}_\varepsilon(r)\right)\mathbf{c}^\top\mu_X - \sigma_\varepsilon\kappa_\pi^\top \mathbf{M}_X^{1/2}\mathcal{Z}\\
&-b\sigma_\varepsilon^2\int_0^1 \mathcal{J}_{-b,\varepsilon}^2(r)\, dr - 2b\sigma_\varepsilon\left(\int_0^1 (1-\exp(-br))\,\mathcal{J}_{-b,\varepsilon}(r)\, dr\right)\mathbf{c}^\top\mu_X\\
&-b\left(\int_0^1 (1-\exp(-br))^2\, dr\right)\left(\mathbf{c}^\top\mu_X\right)^2 + b\sigma_\varepsilon\left(\int_0^1 \mathcal{J}_{-b,\varepsilon}(r)\, dr\right)(\mathbf{c}+\kappa_\pi)^\top\mu_X\\
&+b\left(\int_0^1 (1-\exp(-br))\, dr\right)(\mathbf{c}+\kappa_\pi)^\top\mu_X\mathbf{c}^\top\mu_X - b\kappa_\pi^\top\mathbf{M}_{\mathbf{X}}\mathbf{c}.
\end{aligned}$$

2. *$n^{-1}\left[\partial^2 Q_n(0,\pi;\gamma_n)/\partial\beta^2\right] \Rightarrow \mathcal{H}(\kappa_\pi, b, \mathbf{c}; \varphi_0)$, where*

$$\begin{aligned}
&\mathcal{H}(\kappa_\pi, b, \mathbf{c}; \varphi_0)\\
=\;& \sigma_\varepsilon^2\int_0^1 \mathcal{J}_{-b,\varepsilon}^2(r)\, dr + 2\sigma_\varepsilon\left(\int_0^1 (1-\exp(-br))\,\mathcal{J}_{-b,\varepsilon}(r)\, dr\right)\mathbf{c}^\top\mu_X\\
&+\left(\int_0^1 (1-\exp(-br))^2\, dr\right)\left(\mathbf{c}^\top\mu_X\right)^2 + \kappa_\pi^\top\mathbf{M}_X\kappa_\pi\\
&-2\sigma_\varepsilon\left(\int_0^1 \mathcal{J}_{-b,\varepsilon}(r)\, dr\right)\kappa_\pi^\top\mu_X - 2\left(\int_0^1 (1-\exp(-br))\, dr\right)\kappa_\pi^\top\mu_X\mathbf{c}^\top\mu_X.
\end{aligned}$$

Theorem 6 *(Theorem 4) Suppose that Assumptions 1, 2 and 3 hold.*

1. *When* $\theta_n = \theta_0 \in \Theta_n^*$*, i.e.,* $\beta_n = \beta_0$ *and* $\pi_n = \pi_0$ *for any* $n \in \mathbb{N}$*,* $T_n \overset{A}{\sim} \mathcal{N}(0,1)$*, and* $W_n \overset{A}{\sim} \chi^2(d_r)$*.*

2. *When* $\gamma_n \in \Gamma_n(1,b,\mathbf{c})$ *i.e.,* $\beta_n = b/n$ *with* $0 < b < \infty$ *and* $\pi_n = n^{1/2}\mathbf{c}$*, and* $y_0 = o_p\left(n^{1/2}\right)$*,*

$$\begin{aligned}
B^{-1}(1)\widehat{\mathbf{V}}_n B^{-1}(1) &= \begin{bmatrix} n^{-1/2} & \mathbf{0}_{1\times d_\pi} \\ \mathbf{0}_{d_\pi \times 1} & n\mathbb{I}_{d_\pi} \end{bmatrix} \widehat{\mathbf{V}}_n \begin{bmatrix} n^{-1/2} & \mathbf{0}_{1\times d_\pi} \\ \mathbf{0}_{d_\pi \times 1} & n\mathbb{I}_{d_\pi} \end{bmatrix} \\
&\Rightarrow \mathcal{V}_1(b,\mathbf{c};\varphi_0) = \begin{bmatrix} \mathcal{V}_1^{\beta\beta}(b,\mathbf{c};\varphi_0) & \mathcal{V}_1^{\beta\pi}(b,\mathbf{c};\varphi_0) \\ \mathcal{V}_1^{\pi\beta}(b,\mathbf{c};\varphi_0) & \mathcal{V}_1^{\pi\pi}(b,\mathbf{c};\varphi_0) \end{bmatrix},
\end{aligned}$$

where $\mathcal{V}_1^{\beta\pi}(b,\mathbf{c};\varphi_0) = \left(\mathcal{V}_1^{\pi\beta}(b,\mathbf{c};\varphi_0)\right)^\top$*,*

$$\begin{aligned}
\mathcal{V}_1^{\beta\beta}(b,\mathbf{c};\varphi_0) &= \sigma_\varepsilon^2 \int_0^1 \mathcal{J}_{-b,\varepsilon}^2(r)\,dr + 2\sigma_\varepsilon \left(\int_0^1 (1-\exp(-br))\,\mathcal{J}_{-b,\varepsilon}(r)\,dr\right)\mathbf{c}^\top \mu_X \\
&\quad + \left(\int_0^1 (1-\exp(-br))^2\,dr\right)\left(\mathbf{c}^\top\mu_X\right)^2 - 2\sigma_\varepsilon\left(\int_0^1 \mathcal{J}_{-b,\varepsilon}(r)\,dr\right)\widehat{\kappa}_\pi^\top \mu_X \\
&\quad -2\left(\int_0^1 (1-\exp(-br))\,dr\right)\widehat{\kappa}_\pi^\top \mu_X \mathbf{c}^\top \mu_X + \widehat{\kappa}_\pi^\top \mathbf{M}_X \widehat{\kappa}_\pi, \\
\mathcal{V}_1^{\pi\pi}(b,\mathbf{c};\varphi_0) &= \widehat{\lambda}_\beta^2(\widehat{\kappa}_\pi)\,\mathbf{M}_X, \\
\mathcal{V}_1^{\pi\beta}(b,\mathbf{c};\varphi_0) &= \widehat{\lambda}_\beta(\widehat{\kappa}_\pi) \times \left\{\mathbf{M}_X\widehat{\kappa}_\pi - \sigma_\varepsilon\left(\int_0^1 \mathcal{J}_{-b,\varepsilon}(r)\,dr\right)\mu_X\right. \\
&\quad \left. - \left(\int_0^1 (1-\exp(-br))\,dr\right)\mu_X \mathbf{c}^\top \mu_X\right\},
\end{aligned}$$

and

$$\begin{aligned}
T_n &\Rightarrow \mathcal{T}(b,\mathbf{c};\varphi_0) = \frac{\mathbf{R}\widehat{\tau}(b,\mathbf{c};\varphi_0)}{\left[\sigma_\varepsilon^2 \mathbf{R}\mathcal{V}_1^{-1}(b,\mathbf{c};\varphi_0)\,\mathbf{R}^\top\right]^{1/2}}, \\
W_n &\Rightarrow \mathcal{W}(b,\mathbf{c};\varphi_0) = \left[\mathbf{R}\widehat{\tau}(b,\mathbf{c};\varphi_0)\right]^\top \left[\sigma_\varepsilon^2 \mathbf{R}\mathcal{V}_1^{-1}(b,\mathbf{c};\varphi_0)\,\mathbf{R}^\top\right]^{-1} \mathbf{R}\widehat{\tau}(b,\mathbf{c};\varphi_0),
\end{aligned}$$

where $\widehat{\lambda}_\beta(\widehat{\kappa}_\pi) = \widehat{\lambda}_\beta(\widehat{\kappa}_\pi(b,\mathbf{c};\varphi_0), b, \mathbf{c};\varphi_0)$*,* $\widehat{\kappa}_\pi = \widehat{\kappa}_\pi(b,\mathbf{c};\varphi_0)$ *and* $\widehat{\tau}(b,\mathbf{c};\varphi_0)$ *are defined in Theorem 2.*

3. *When* $\gamma_n \in \Gamma_n(h,b,\mathbf{c})$*, i.e.,* $\beta_n = b/n^h$ *with* $0 < b < \infty$ *and* $\pi_n = n^{-1/2+h}\mathbf{c}$*, where* $0 < h < 1$*,* $T_n \overset{A}{\sim} \mathcal{N}(0,1)$*, and* $W_n \overset{A}{\sim} \chi^2(d_r)$*.*

7 Appendix B: Proofs of Theorems and Lemmas

Proof. (Lemma 1) For simplicity, we only illustrate the case when $d_\pi = 1$. When $\beta_n = 0$, $y_t = y_{t-1} + \varepsilon_t$ for $t = 1, \ldots, n$. By the law of number for stationary ergodic sequences (White, 2001, Theorem 3.34, p. 44), the central limit theorem for stationary ergodic adapted mixingales (White, 2001, Theorem 5.16, p. 125), and Lemma 6 with $b \to 0$,

$$
\begin{aligned}
n^{-1}\sum_{t=1}^{n} X_t^2 &\to_{a.s.} \mathbf{M}_X, \quad n^{-1/2}\sum_{t=1}^{n} X_t\varepsilon_t \Rightarrow \sigma_\varepsilon \mathbf{M}_X^{1/2}\mathcal{Z} \sim \mathcal{N}\left(0, \sigma_\varepsilon^2 \mathbf{M}_{\mathbf{X}}\right), \\
n^{-2}\sum_{t=1}^{n} y_{t-1}^2 &\Rightarrow \sigma_\varepsilon^2 \int_0^1 \mathcal{W}_\varepsilon^2(r)\, dr, \quad n^{-1}\sum_{t=1}^{n} y_{t-1}\varepsilon_t \Rightarrow \sigma_\varepsilon^2 \int_0^1 \mathcal{W}_\varepsilon(r)\, d\mathcal{W}_\varepsilon(r), \quad \text{and} \\
n^{-3/2}\sum_{t=1}^{n} X_t y_{t-1} &\Rightarrow \mu_X \sigma_\varepsilon \int_0^1 \mathcal{W}_\varepsilon(r)\, dr.
\end{aligned}
$$

Then by the first order condition of equation (2.1),

$$
\begin{aligned}
n^{-1/2}\widehat{\pi}_n &= \frac{\left(n^{-2}\sum_{t=1}^{n} y_{t-1}^2\right)\left(n^{-1/2}\sum_{t=1}^{n} X_t\varepsilon_t\right) - \left(n^{-1}\sum_{t=1}^{n} y_{t-1}\varepsilon_t\right)\left(n^{-3/2}\sum_{t=1}^{n} X_t y_{t-1}\right)}{\left(n^{-3/2}\sum_{t=1}^{n} X_t y_{t-1}\right)\left(n^{-1/2}\sum_{t=1}^{n} X_t\varepsilon_t\right) - \left(n^{-1}\sum_{t=1}^{n} y_{t-1}\varepsilon_t\right)\left(n^{-1}\sum_{t=1}^{n} X_t^2\right)} \\
&\Rightarrow \frac{\sigma_\varepsilon^2 \int_0^1 \mathcal{W}_\varepsilon^2(r)\, dr \cdot \sigma_\varepsilon \mathbf{M}_X^{1/2}\mathcal{Z} - \sigma_\varepsilon^2 \int_0^1 \mathcal{W}_\varepsilon(r)\, d\mathcal{W}_\varepsilon(r) \cdot \mu_X \sigma_\varepsilon \int_0^1 \mathcal{W}_\varepsilon(r)\, dr}{\mu_X \sigma_\varepsilon \int_0^1 \mathcal{W}_\varepsilon(r)\, dr \cdot \sigma_\varepsilon \mathbf{M}_X^{1/2}\mathcal{Z} - \sigma_\varepsilon^2 \int_0^1 \mathcal{W}_\varepsilon(r)\, d\mathcal{W}_\varepsilon(r) \cdot \mathbf{M}_X} = \widehat{\kappa}_\pi = O_p(1),
\end{aligned}
$$

$$
\begin{aligned}
n^{-1}\widehat{\beta}_n &= \frac{\left(n^{-1/2}\sum_{t=1}^{n} X_t\varepsilon_t\right) n^{-1/2}\widehat{\pi}_n - \left(n^{-1}\sum_{t=1}^{n} y_{t-1}\varepsilon_t\right)}{\left(n^{-1}\sum_{t=1}^{n} X_t^2\right)\left(n^{-1/2}\widehat{\pi}_n\right)^2 - 2\left(n^{-3/2}\sum_{t=1}^{n} X_t y_{t-1}\right) n^{-1/2}\widehat{\pi}_n + \left(n^{-2}\sum_{t=1}^{n} y_{t-1}^2\right)} \\
&\Rightarrow \frac{\sigma_\varepsilon \mathbf{M}_X^{1/2}\mathcal{Z} \cdot \widehat{\kappa}_\pi - \sigma_\varepsilon^2 \int_0^1 \mathcal{W}_\varepsilon(r)\, d\mathcal{W}_\varepsilon(r)}{\mathbf{M}_X \cdot \widehat{\kappa}_\pi^2 - 2\mu_X \sigma_\varepsilon \int_0^1 \mathcal{W}_\varepsilon(r)\, dr \cdot \widehat{\kappa}_\pi + \sigma_\varepsilon^2 \int_0^1 \mathcal{W}_\varepsilon^2(r)\, dr} = O_p(1).
\end{aligned}
$$

■

Proof. (Lemma 2 / Lemma 5)

1. $\left(\left(\partial Q_n\left(0, \pi; \gamma_n\right)\right)/\partial \beta\right)$ By Lemma 6, the law of number for stationary ergodic sequences (White, 2001, Theorem 3.34, p. 44), and the central limit theorem for stationary

ergodic adapted mixingales (White, 2001, Theorem 5.16, p. 125),

$$
\begin{aligned}
&\frac{\partial}{\partial \beta} Q_n(0,\pi;\gamma_n) = n^{-1}\sum_{t=1}^{n}(y_t - y_{t-1})\left(y_{t-1} - X_t^\top \pi\right) \\
= \;& n^{-1}\sum_{t=1}^{n}\left(\varepsilon_t - \beta_n y_{t-1} + \beta_n X_t^\top \pi_n\right)\left(y_{t-1} - X_t^\top \pi\right) \\
= \;& n^{-1}\sum_{t=1}^{n} y_{t-1}\varepsilon_t - n^{-1}\pi^\top \sum_{t=1}^{n} X_t\varepsilon_t - n^{-1}\beta_n\sum_{t=1}^{n} y_{t-1}^2 + n^{-1}\beta_n \pi^\top \sum_{t=1}^{n} X_t y_{t-1} \\
& + n^{-1}\beta_n \pi_n^\top \sum_{t=1}^{n} X_t y_{t-1} - n^{-1}\beta_n \pi^\top\left(\sum_{t=1}^{n} X_t X_t^\top\right)\pi_n \\
\Rightarrow \;& \sigma_\varepsilon^2 \int_0^1 \mathcal{J}_{-b,\varepsilon}(r)\, d\mathcal{W}_\varepsilon(r) + \sigma_\varepsilon\left(\int_0^1 (1-\exp(-br))\, d\mathcal{W}_\varepsilon(r)\right)\mathbf{c}^\top \mu_X - \sigma_\varepsilon \kappa_\pi^\top \mathbf{M}_X^{1/2}\mathcal{Z} \\
& - b\sigma_\varepsilon^2 \int_0^1 \mathcal{J}_{-b,\varepsilon}^2(r)\, dr - 2b\sigma_\varepsilon\left(\int_0^1 (1-\exp(-br))\,\mathcal{J}_{-b,\varepsilon}(r)\, dr\right)\mathbf{c}^\top \mu_X \\
& - b\left(\int_0^1 (1-\exp(-br))^2\, dr\right)\left(\mathbf{c}^\top \mu_X\right)^2 + b\sigma_\varepsilon\left(\int_0^1 \mathcal{J}_{-b,\varepsilon}(r)\, dr\right)(\mathbf{c}+\kappa_\pi)^\top \mu_X \\
& + b\left(\int_0^1 (1-\exp(-br))\, dr\right)(\mathbf{c}+\kappa_\pi)^\top \mu_X \mathbf{c}^\top \mu_X - b\kappa_\pi^\top \mathbf{M}_X \mathbf{c}.
\end{aligned}
$$

2. ($n^{-1}\left[\partial^2 Q_n(0,\pi;\gamma_n)/\partial\beta^2\right]$) By Lemma 6 and the law of large number for stationary ergodic sequences (White, 2001, Theorem 3.34, p. 44),

$$
\begin{aligned}
&n^{-1}\frac{\partial^2}{\partial \beta^2} Q_n(0,\pi;\gamma_n) = n^{-2}\sum_{t=1}^{n}\left(y_{t-1} - X_t^\top \pi\right)^2 \\
= \;& n^{-2}\sum_{t=1}^{n} y_{t-1}^2 + n^{-2}\pi^\top \sum_{t=1}^{n} X_t X_t^\top \pi - 2n^{-2}\pi^\top \sum_{t=1}^{n} X_t y_{t-1} \\
\Rightarrow \;& \sigma_\varepsilon^2 \int_0^1 \mathcal{J}_{-b,\varepsilon}^2(r)\, dr + 2\sigma_\varepsilon\left(\int_0^1 (1-\exp(-br))\,\mathcal{J}_{-b,\varepsilon}(r)\, dr\right)\mathbf{c}^\top \mu_X \\
& + \left(\int_0^1 (1-\exp(-br))^2\, dr\right)\left(\mathbf{c}^\top \mu_X\right)^2 + \kappa_\pi^\top \mathbf{M}_X \kappa_\pi \\
& - 2\sigma_\varepsilon\left(\int_0^1 \mathcal{J}_{-b,\varepsilon}(r)\, dr\right)\kappa_\pi^\top \mu_X - 2\left(\int_0^1 (1-\exp(-br))\, dr\right)\kappa_\pi^\top \mu_X \mathbf{c}^\top \mu_X.
\end{aligned}
$$

■

Proof. (Theorem 2) For notational simplicity, let $\widehat{\lambda}_\beta$ denote $\widehat{\lambda}_\beta\left(\widehat{\kappa}_\pi(b,\mathbf{c};\varphi_0),b,\mathbf{c};\varphi_0\right)$, $\widehat{\kappa}_\pi$ denote $\widehat{\kappa}_\pi(b,\mathbf{c};\varphi_0)$, and $q(\lambda_\beta,\kappa_\pi)$ denote $q(\lambda_\beta,\kappa_\pi,b,\mathbf{c};\varphi_0)$. Also, let $\widehat{\lambda}_{\beta,n} = n\widehat{\beta}_n$ and $\widehat{\kappa}_{\pi,n} =$

$n^{-1/2}\widehat{\pi}_n$. Then it suffices to show $\left\{\widehat{\lambda}_{\beta,n}, \widehat{\kappa}_{\pi,n}\right\} \Rightarrow \left\{\widehat{\lambda}_{\beta}, \widehat{\kappa}_{\pi}\right\}$. Let

$$\begin{aligned} q_n(\lambda_\beta, \kappa_\pi) &= q_n(\lambda_\beta, \kappa_\pi, b, \mathbf{c}; \varphi_0) \\ &= \frac{\partial}{\partial \beta} Q_n\left(0, n^{1/2}\kappa_\pi; \gamma_n\right) \cdot \lambda_\beta + \frac{1}{2} n^{-1} \frac{\partial^2}{\partial \beta^2} Q_n\left(0, n^{1/2}\kappa_\pi; \gamma_n\right) \cdot \lambda_\beta^2. \end{aligned}$$

Then by equations (2.1) and (2.3) / equations (2.6) and (2.7), $\left\{\widehat{\lambda}_{\beta,n}, \widehat{\kappa}_{\pi,n}\right\}$ / $\left\{\widehat{\lambda}_{\beta}, \widehat{\kappa}_{\pi}\right\}$ are respectively the unique minimizers of $q_n(\lambda_\beta, \kappa_\pi)$ and $q(\lambda_\beta, \kappa_\pi)$ in $\mathbb{R}^{d_\pi+1}$, *i.e.*,

$$q_n\left(\widehat{\lambda}_{\beta,n}, \widehat{\kappa}_{\pi,n}\right) = \min_{\lambda_\beta, \kappa_\pi} q_n(\lambda_\beta, \kappa_\pi), \quad \text{and} \quad q\left(\widehat{\lambda}_{\beta}, \widehat{\kappa}_{\pi}\right) = \min_{\lambda_\beta, \kappa_\pi} q(\lambda_\beta, \kappa_\pi).$$

By Lemma 2 and equation (2.5), for any given $\{\lambda_\beta, \kappa_\pi\} \in \mathcal{C}$, $q_n(\lambda_\beta, \kappa_\pi) \Rightarrow q(\lambda_\beta, \kappa_\pi)$ when $n \to \infty$. Since $q_n(\lambda_\beta, \kappa_\pi)$ and $q(\lambda_\beta, \kappa_\pi)$ are concave functions with respect to $\{\lambda_\beta, \kappa_\pi\}$, by the fact that pointwise convergence of concave functions on a dense subset of an open set implies uniform convergence on any compact subset of the open set (Newey and McFadden, 1994, proof of Theorem 2.7, pp. 2133, 2134), $q_n(\lambda_\beta, \kappa_\pi) \Rightarrow q(\lambda_\beta, \kappa_\pi)$ uniformly on any compact set of $\mathbb{R}$ when $n \to \infty$.

Consider a compact set $\mathbb{C} \subset \mathbb{R}$. Let $\mathbb{Z}_n$ / $\mathbb{Z}$ be the inverse images of $q_n(\lambda_\beta, \kappa_\pi)$ / $q(\lambda_\beta, \kappa_\pi)$ in $\mathbb{R}^{d_\pi+1}$ respectively, *i.e.*, $\mathbb{Z}_n = \left\{\{\lambda_\beta, \kappa_\pi\} \in \mathbb{R}^{d_\pi+1} : q_n(\lambda_\beta, \kappa_\pi) \in \mathbb{C}\right\}$, and $\mathbb{Z} = \left\{\{\lambda_\beta, \kappa_\pi\} \in \mathbb{R}^{d_\pi+1} : q(\lambda_\beta, \kappa_\pi) \in \mathbb{C}\right\}$. By the compactness of $\mathbb{C}$ and the continuity of $q_n(\lambda_\beta, \kappa_\pi)$ / $q(\lambda_\beta, \kappa_\pi)$ with respect to $\{\lambda_\beta, \kappa_\pi\}$, $\mathbb{Z}_n$ / $\mathbb{Z}$ are also compact. And since $q_n(\lambda_\beta, \kappa_\pi) \Rightarrow q(\lambda_\beta, \kappa_\pi)$ uniformly on $\mathbb{C}$ when $n \to \infty$, $\mathbb{Z}_n \to \mathbb{Z}$ when $n \to \infty$. Let $\left\{\widehat{\lambda}^*_{\beta,n}, \widehat{\kappa}^*_{\pi,n}\right\}$ / $\left\{\widehat{\lambda}^*_{\beta}, \widehat{\kappa}^*_{\pi}\right\}$ be the minimizers of of $q_n(\lambda_\beta, \kappa_\pi)$ and $q(\lambda_\beta, \kappa_\pi)$ in $\mathbb{Z}_n$ / $\mathbb{Z}$, *i.e.*,

$$q_n\left(\widehat{\lambda}^*_{\beta,n}, \widehat{\kappa}^*_{\pi,n}\right) = \min_{\left\{\lambda_\beta, \kappa_\pi\right\} \in \mathbb{Z}_n} q_n(\lambda_\beta, \kappa_\pi), \quad \text{and} \quad q\left(\widehat{\lambda}^*_{\beta}, \widehat{\kappa}^*_{\pi}\right) = \min_{\left\{\lambda_\beta, \kappa_\pi\right\} \in \mathbb{Z}} q(\lambda_\beta, \kappa_\pi).$$

By the concavity of $q_n(\lambda_\beta, \kappa_\pi)$ and $q(\lambda_\beta, \kappa_\pi)$, $\left\{\widehat{\lambda}^*_{\beta,n}, \widehat{\kappa}^*_{\pi,n}\right\}$ / $\left\{\widehat{\lambda}^*_{\beta}, \widehat{\kappa}^*_{\pi}\right\}$ are unique. If $\left\{\widehat{\lambda}^*_{\beta,n}, \widehat{\kappa}^*_{\pi,n}\right\}$ are tight for every $n \in \mathbb{N}$, by the compactness of $\mathbb{Z}_n$, $\left\{\widehat{\lambda}^*_{\beta,n}, \widehat{\kappa}^*_{\pi,n}\right\}$ will be uniformly tight with respect to n. Then by the Argmax continuous mapping theorem (van der Vaart and Wellner, 1996, p.286), when $n \to \infty$,

$$\left\{\widehat{\lambda}^*_{\beta,n}, \widehat{\kappa}^*_{\pi,n}\right\} = \underset{\left\{\lambda_\beta, \kappa_\pi\right\} \in \mathbb{Z}_n}{\arg\min}\, q_n(\lambda_\beta, \kappa_\pi) \Rightarrow \underset{\left\{\lambda_\beta, \kappa_\pi\right\} \in \mathbb{Z}}{\arg\min}\, q(\lambda_\beta, \kappa_\pi) = \left\{\widehat{\lambda}^*_{\beta}, \widehat{\kappa}^*_{\pi}\right\}.$$

Since $\mathbb{C}$ is arbitrary, the desired results directly follow. That is, for any compact subset

$\mathbb{C} \subset \mathbb{R}$ to which $\min_{\lambda_\beta,\kappa_\pi} q_n(\lambda_\beta,\kappa_\pi)$ and $\min_{\lambda_\beta,\kappa_\pi} q(\lambda_\beta,\kappa_\pi)$ belong, when $n \to \infty$,

$$\begin{aligned}\left\{\widehat{\lambda}_{\beta,n},\widehat{\kappa}_{\pi,n}\right\} &= \arg\min_{\lambda_\beta,\kappa_\pi} q_n(\lambda_\beta,\kappa_\pi) = \arg\min_{\left\{\lambda_\beta,\kappa_\pi\right\}\in\mathbb{Z}_n} q_n(\lambda_\beta,\kappa_\pi) \\ &\Rightarrow \arg\min_{\left\{\lambda_\beta,\kappa_\pi\right\}\in\mathbb{Z}} q(\lambda_\beta,\kappa_\pi) = \arg\min_{\lambda_\beta,\kappa_\pi} q(\lambda_\beta,\kappa_\pi) = \left\{\widehat{\lambda}_\beta,\widehat{\kappa}_\pi\right\}.\end{aligned}$$

It only remains to show the tightness of $\left\{\widehat{\lambda}_{\beta,n},\widehat{\kappa}_{\pi,n}\right\} = \left\{n^{-1}\widehat{\beta}_n, n^{-1/2}\widehat{\pi}_n\right\}$. For simplicity, we only illustrate the case when $d_\pi = 1$. By the first order condition of equation (2.1), the law of number for stationary ergodic sequences (White, 2001, Theorem 3.34, p. 44), the central limit theorem for stationary ergodic adapted mixingales (White, 2001, Theorem 5.16, p. 125), and Lemma 6,

$$\widehat{\kappa}_{\pi,n} = \frac{\left\{\begin{array}{c}\left(n^{-2}\sum_{t=1}^n y_{t-1}^2\right)\left(n^{-1/2}\sum_{t=1}^n X_t\varepsilon_t\right) - \left(n^{-1}\sum_{t=1}^n y_{t-1}\varepsilon_t\right)\left(n^{-3/2}\sum_{t=1}^n X_t y_{t-1}\right) \\ +b\mathbf{c}\left(n^{-2}\sum_{t=1}^n y_{t-1}^2\right)\left(n^{-1}\sum_{t=1}^n X_t^2\right) - b\mathbf{c}\left(n^{-3/2}\sum_{t=1}^n X_t y_{t-1}\right)^2\end{array}\right\}}{\left\{\begin{array}{c}\left(n^{-3/2}\sum_{t=1}^n X_t y_{t-1}\right)\left(n^{-1/2}\sum_{t=1}^n X_t\varepsilon_t\right) - \left(n^{-1}\sum_{t=1}^n y_{t-1}\varepsilon_t\right)\left(n^{-1}\sum_{t=1}^n X_t^2\right) \\ +b\left(n^{-2}\sum_{t=1}^n y_{t-1}^2\right)\left(n^{-1}\sum_{t=1}^n X_t^2\right) - b\left(n^{-3/2}\sum_{t=1}^n X_t y_{t-1}\right)^2\end{array}\right\}} = O_p(1),$$

$$\widehat{\lambda}_{\beta,n} = \frac{\left\{\begin{array}{c}\left(bn^{-3/2}\sum_{t=1}^n X_t y_{t-1} + b\mathbf{c}n^{-1}\sum_{t=1}^n X_t^2 + n^{-1/2}\sum_{t=1}^n X_t\varepsilon_t\right)n^{-1/2}\widehat{\pi}_n \\ -\left(bn^{-2}\sum_{t=1}^n y_{t-1}^2 + b\mathbf{c}n^{-3/2}\sum_{t=1}^n X_t y_{t-1} + n^{-1}\sum_{t=1}^n y_{t-1}\varepsilon_t\right)\end{array}\right\}}{\left(n^{-1}\sum_{t=1}^n X_t^2\right)\left(n^{-1/2}\widehat{\pi}_n\right)^2 - 2\left(n^{-3/2}\sum_{t=1}^n X_t y_{t-1}\right)n^{-1/2}\widehat{\pi}_n + \left(n^{-2}\sum_{t=1}^n y_{t-1}^2\right)} = O_p(1).$$

■

Proof. (Lemma 3) The consistency of $\widehat{\mu}_{X,n}$ and $\widehat{\mathbf{M}}_{X,n}$ directly follows the law of number for stationary ergodic sequences (White, 2001, Theorem 3.34, p. 44). For $\widehat{\sigma}_n^2$, by Lemma 6, Theorem 2, and the Kolmogorov law of large number (White, 2001, Theorem 3.1, p. 32),

$$\begin{aligned}\widehat{\sigma}_n^2 &= n^{-1}\sum_{t=1}^n\left[y_t - \left(1-\widehat{\beta}_n\right)y_{t-1} - \widehat{\beta}_n X_t^\top\widehat{\pi}_n\right]^2 \\ &= n^{-1}\sum_{t=1}^n\left[\varepsilon_t + \left(\widehat{\beta}_n - \beta_n\right)y_{t-1} - \beta_n X_t^\top\left(\widehat{\pi}_n - \pi_n\right) - \left(\widehat{\beta}_n - \beta_n\right)X_t^\top\widehat{\pi}_n\right]^2 \\ &= n^{-1}\sum_{t=1}^n\varepsilon_t^2 + O_p\left(n^{-1}\right) \xrightarrow{p} \sigma_\varepsilon^2.\end{aligned}$$

■

Proof. (Lemma 4)

1. $\left(n^{1/2}B^{-1}(h)\,D_{\theta}Q_n(\theta_n;\gamma_n)\right)$

$$n^{1/2}B^{-1}(h)\,D_{\theta}Q_n(\theta_n;\gamma_n)=\left[\begin{array}{c} n^{-1/2-h/2}\sum_{t=1}^{n}\left(y_{t-1}-X_t^{\top}\pi_n\right)\varepsilon_t \\ -\beta_n n^{-1/2+h}\sum_{t=1}^{n}X_t\varepsilon_t \end{array}\right].$$

By Lemma 7,

$$n^{-1/2-h/2}\sum_{t=1}^{n}\left(y_{t-1}-X_t^{\top}\pi_n\right)\varepsilon_t=n^{-1/2-h/2}\sum_{t=1}^{n}y_{t-1}\varepsilon_t+o_p(1),$$

therefore, by Lemma 7 and the central limit theorem for stationary ergodic adapted mixingales (White, 2001, Theorem 5.16, p. 125),

$$\begin{aligned}&n^{1/2}B^{-1}(h)\,D_{\theta}Q_n(\theta_n;\gamma_n)\\ =&\left[\begin{array}{c} n^{-1/2-h/2}\sum_{t=1}^{n}y_{t-1}\varepsilon_t+O_p\left(n^{-h/2}\right) \\ -bn^{-1/2}\sum_{t=1}^{n}X_t\varepsilon_t \end{array}\right]\overset{A}{\sim}\mathcal{N}\left(\mathbf{0}_{(d_\pi+1)\times 1},\left[\begin{array}{cc} (2b)^{-1}\sigma_\varepsilon^4 & \mathbf{0}_{1\times d_\pi} \\ \mathbf{0}_{d_\pi\times 1} & \sigma_\varepsilon^2 b^2\mathbf{M}_X \end{array}\right]\right).\end{aligned}$$

2. $\left(B^{-1}(h)\,D_{\theta\theta^{\top}}Q_n(\theta_n;\gamma_n)\,B^{-1}(h)\right)$

$$\begin{aligned}&B^{-1}(h)\,D_{\theta\theta^{\top}}Q_n(\theta_n;\gamma_n)\,B^{-1}(h)\\ =&\left[\begin{array}{cc} n^{-1-h}\sum_{t=1}^{n}\left(y_{t-1}-X_t^{\top}\pi_n\right)^2 & \\ -n^{-1+h/2}\sum_{t=1}^{n}X_t\left[\beta_n\left(y_{t-1}-X_t^{\top}\pi_n\right)+\varepsilon_t\right] & \beta_n^2 n^{-1+2h}\sum_{t=1}^{n}X_tX_t^{\top} \end{array}\right].\end{aligned}$$

By Lemma 7 and the law of large number for stationary ergodic sequences (White, 2001, Theorem 3.34, p. 44),

$$\begin{aligned} n^{-1-h}\sum_{t=1}^{n}\left(y_{t-1}-X_t^{\top}\pi_n\right)^2 &= n^{-1-h}\sum_{t=1}^{n}y_{t-1}^2+o_p(1)\xrightarrow{p}\frac{\sigma_\varepsilon^2}{2b},\\ \beta_n^2 n^{-1+2h}\sum_{t=1}^{n}X_tX_t^{\top} &= b^2 n^{-1}\sum_{t=1}^{n}X_tX_t^{\top}\xrightarrow{p}b^2\mathbf{M}_X,\end{aligned}$$

and

$$\begin{aligned}&-n^{-1+h/2}\sum_{t=1}^{n}X_t\left[\beta_n\left(y_{t-1}-X_t^{\top}\pi_n\right)+\varepsilon_t\right]\\ =&-n^{-2+h/2}b\sum_{t=1}^{n}X_ty_{t-1}+n^{-2+h/2}b\sum_{t=1}^{n}X_tX_t^{\top}\pi_n-n^{-1+h/2}\sum_{t=1}^{n}X_t\varepsilon_t\\ =&\;O_p\left(n^{-3/2+3h/2}\right)+O_p\left(n^{-1/2+h/2}\right)+O_p\left(n^{-1/2+h/2}\right)\xrightarrow{p}\mathbf{0}_{d_\pi\times 1}.\end{aligned}$$

Therefore,

$$B^{-1}(h)\, D_{\theta\theta^\top} Q_n(\theta_n;\gamma_n)\, B^{-1}(h) \xrightarrow{p} \left[\begin{array}{cc} (2b)^{-1}\sigma_\varepsilon^2 & \\ \mathbf{0}_{d_\pi\times 1} & b^2\mathbf{M}_X \end{array}\right].$$

■

Proof. (Theorem 3) First we show that $\widehat{\beta}_n - \beta_n = O\left(n^{-1/2-h/2}\right)$ and $\widehat{\pi}_n - \pi_n = O\left(n^{-1/2+h}\right)$. Again, for simplicity, we only illustrate the case when $d_\pi = 1$. By the first order condition of equation (2.1), the law of number for stationary ergodic sequences (White, 2001, Theorem 3.34, p. 44), the central limit theorem for stationary ergodic adapted mixingales (White, 2001, Theorem 5.16, p. 125), and Lemma 7,

$$\begin{aligned} n^{1/2-h}\left(\widehat{\pi}_n - \pi_n\right) &= \frac{\left\{\begin{array}{c}\left(n^{-1-h}\sum_{t=1}^n y_{t-1}^2\right)\left(n^{-1/2}\sum_{t=1}^n X_t\varepsilon_t\right) \\ +b\mathbf{c}\left(n^{-1-h}\sum_{t=1}^n y_{t-1}^2\right)\left(n^{-1}\sum_{t=1}^n X_t^2\right) + o_p(1)\end{array}\right\}}{b\left(n^{-1-h}\sum_{t=1}^n y_{t-1}^2\right)\left(n^{-1}\sum_{t=1}^n X_t^2\right) + o_p(1)} = O_p(1), \\ n^{1/2+h/2}\left(\widehat{\beta}_n - \beta_n\right) &= \frac{-n^{-1/2-h/2}\sum_{t=1}^n y_{t-1}\varepsilon_t + o_p(1)}{n^{-1-h}\sum_{t=1}^n y_{t-1}^2 + o_p(1)} = O_p(1). \end{aligned}$$

Then we show that $R(\theta^*;\gamma_n) = o_p(n^{-1})$. By equation (2.10),

$$\begin{aligned} & Q_n\left(\widehat{\theta}_n;\gamma_n\right) - Q_n(\theta_n;\gamma_n) \\ =\; & D_\theta^\top Q_n(\theta_n;\gamma_n)\left(\widehat{\theta}_n - \theta_n\right) + \frac{1}{2}\left(\widehat{\theta}_n - \theta_n\right)^\top D_{\theta\theta^\top} Q_n(\theta_n;\gamma_n)\left(\widehat{\theta}_n - \theta_n\right) + R(\theta^*;\gamma_n), \end{aligned}$$

where θ^* is in between $\widehat{\theta}_n$ and θ_n, $\beta^* = (\beta^* - \beta_n) + \beta_n = O\left(n^{-1/2-h/2}\right) + O\left(n^{-h}\right)$, and $\pi^* = O\left(n^{-1/2+h}\right)$. By the cubic approximation for $Q_n(\theta;\gamma_n)$ around θ_n:

$$\begin{aligned} R(\theta^*;\gamma_n) &= \frac{1}{2}\left(\widehat{\beta}_n - \beta_n\right)^2 \frac{\partial^3 Q_n(\theta^*;\gamma_n)}{\partial\beta^2\partial\pi}\left(\widehat{\pi}_n - \pi_n\right) \\ &\quad + \frac{1}{2}\left(\widehat{\beta}_n - \beta_n\right)\left(\widehat{\pi}_n - \pi_n\right)^\top \frac{\partial^3 Q_n(\theta^*;\gamma_n)}{\partial\beta\partial\pi\partial\pi^\top}\left(\widehat{\pi}_n - \pi_n\right), \end{aligned}$$

in which

$$\begin{aligned} \frac{\partial^3 Q_n(\theta^*;\gamma_n)}{\partial\beta^2\partial\pi} &= -2n^{-1}\sum_{t=1}^n\left(y_{t-1} - X_t^\top\pi^*\right)X_t^\top, \\ \frac{\partial^3 Q_n(\theta^*;\gamma_n)}{\partial\beta\partial\pi\partial\pi^\top} &= 2\beta^* n^{-1}\sum_{t=1}^n X_t X_t^\top. \end{aligned}$$

By Lemma 7, the law of large number for stationary ergodic sequences (White, 2001, Theorem 3.34, p. 44), $\beta^* = O\left(n^{-1/2-h/2}\right) + O\left(n^{-h}\right)$ and $\pi^* = O\left(n^{-1/2+h}\right)$,

$$\begin{aligned}
\frac{\partial^3 Q_n(\theta^*;\gamma_n)}{\partial\beta^2\partial\pi} &= -2n^{-1}\sum_{t=1}^{n} y_{t-1}X_t^\top + 2\pi^{*\top}n^{-1}\sum_{t=1}^{n} X_tX_t^\top = O_p\left(n^{-1/2+h}\right), \\
\frac{\partial^3 Q_n(\theta^*;\gamma_n)}{\partial\beta\partial\pi\partial\pi^\top} &= 2\beta^* n^{-1}\sum_{t=1}^{n} X_tX_t^\top = O\left(n^{-1/2-h/2}\right) + O\left(n^{-h}\right).
\end{aligned}$$

Therefore, by $\widehat{\beta}_n - \beta_n = O\left(n^{-1/2-h/2}\right)$, and $\widehat{\pi}_n - \pi_n = O\left(n^{-1/2+h}\right)$,

$$\begin{aligned}
R(\theta^*;\gamma_n) &= \frac{1}{2}\left(\widehat{\beta}_n - \beta_n\right)^2 \frac{\partial^3 Q_n(\theta^*;\gamma_n)}{\partial\beta^2\partial\pi}\left(\widehat{\pi}_n - \pi_n\right) \\
&\quad + \frac{1}{2}\left(\widehat{\beta}_n - \beta_n\right)\left(\widehat{\pi}_n - \pi_n\right)^\top \frac{\partial^3 Q_n(\theta^*;\gamma_n)}{\partial\beta\partial\pi\partial\pi^\top}\left(\widehat{\pi}_n - \pi_n\right) \\
&= \left[O\left(n^{-1/2-h/2}\right)\right]^2 \cdot O_p\left(n^{-1/2+h}\right) \cdot O\left(n^{-1/2+h}\right) \\
&\quad + O\left(n^{-1/2-h/2}\right) \cdot O\left(n^{-1/2+h}\right) \cdot \left[O\left(n^{-1/2-h/2}\right) + O\left(n^{-h}\right)\right] \cdot O\left(n^{-1/2+h}\right) \\
&= O_p\left(n^{-2+h}\right) + O_p\left(n^{-3/2+h/2}\right) = o_p\left(n^{-1}\right).
\end{aligned}$$

Let

$$\begin{aligned}
J_n &= B^{-1}(h)\, D_{\theta\theta^\top} Q_n(\theta_n;\gamma_n)\, B^{-1}(h), \quad Z_n^* = -n^{1/2} J_n^{-1} B^{-1}(h)\, D_\theta Q_n(\theta_n;\gamma_n), \\
\Delta_n^*(\theta) &= n^{1/2} B(h)(\theta - \theta_n). \quad \text{and} \quad q_n(\Delta_n^*(\theta);\gamma_n) = n\left(Q_n(\theta;\gamma_n) - Q_n(\theta_n;\gamma_n)\right).
\end{aligned}$$

Then by equation (2.10), Lemma 4, and the fact that $R(\theta^*;\gamma_n) = o_p(n^{-1})$,

$$\begin{aligned}
q_n(\Delta_n^*(\theta);\gamma_n) &= -Z_n^{*\top} J_n \Delta_n^*(\theta) + \frac{1}{2}\left[\Delta_n^*(\theta)\right]^\top J_n \Delta_n^*(\theta) + o_p(1) \\
&= \frac{1}{2}\left[\Delta_n^*(\theta) - Z_n^*\right]^\top J_n \left[\Delta_n^*(\theta) - Z_n^*\right] - \frac{1}{2} Z_n^{*\top} J_n Z_n^* + o_p(1).
\end{aligned}$$

By definition (equation (2.1)), $\widehat{\theta}_n$ is the minimizer of $Q_n(\theta;\gamma_n) - Q_n(\theta_n;\gamma_n)$, and therefore $\Delta_n^*(\theta)$ is the minimizer of $q_n(\Delta_n^*(\theta);\gamma_n)$, *i.e.*,

$$q_n\left(\Delta_n^*\left(\widehat{\theta}_n\right);\gamma_n\right) = \min_\theta q_n(\Delta_n^*(\theta);\gamma_n).$$

Therefore $\Delta_n^*\left(\widehat{\theta}_n\right) \overset{A}{=} Z_n^*$. *By Lemma 4,*

$$
\begin{aligned}
n^{1/2}B\left(h\right)\left(\widehat{\theta}_n-\theta_n\right) &\overset{A}{=} -n^{1/2}J_n^{-1}B^{-1}\left(h\right)D_\theta Q_n\left(\theta_n;\gamma_n\right)\\
\Rightarrow\ & \mathcal{V}^{*-1}\left(b;\varphi_0\right)\mathcal{G}^*\left(b;\varphi_0\right)\sim\mathcal{N}\left(\mathbf{0}_{(d_\pi+1)\times 1},\sigma_\varepsilon^2\mathcal{V}^{*-1}\left(b;\varphi_0\right)\right).
\end{aligned}
$$

■

Proof. (Theorem 4 / Theorem 6) 1. directly follows by Theorem 4.1 of Newey and McFadden (1994, p. 2156). For 2. by equations (3.4) and (2.11),

$$
B^{-1}\left(1\right)\widehat{\mathbf{V}}_nB^{-1}\left(1\right)=\left[\begin{array}{cc} n^{-2}\sum_{t=1}^n\left(y_{t-1}-X_t^\top\widehat{\pi}_n\right)^2 & \\ -n^{-1/2}\widehat{\beta}_n\sum_{t=1}^nX_t\left(y_{t-1}-X_t^\top\widehat{\pi}_n\right) & n\widehat{\beta}_n^2\sum_{t=1}^nX_tX_t^\top\end{array}\right].
$$

By Lemma 6, the law of large number and Theorem 2:

$$
\begin{aligned}
& n^{-2}\sum_{t=1}^n\left(y_{t-1}-X_t^\top\widehat{\pi}_n\right)^2=n^{-2}\sum_{t=1}^ny_{t-1}^2-2n^{-2}\widehat{\pi}_n^\top\sum_{t=1}^nX_ty_{t-1}+n^{-2}\widehat{\pi}_n^\top\sum_{t=1}^nX_tX_t^\top\widehat{\pi}_n\\
\Rightarrow\ & \sigma_\varepsilon^2\int_0^1\mathcal{J}_{-b,\varepsilon}^2\left(r\right)dr+2\sigma_\varepsilon\left(\int_0^1\left(1-\exp\left(-br\right)\right)\mathcal{J}_{-b,\varepsilon}\left(r\right)dr\right)\mathbf{c}^\top\mu_X\\
& +\left(\int_0^1\left(1-\exp\left(-br\right)\right)^2dr\right)\left(\mathbf{c}^\top\mu_X\right)^2-2\sigma_\varepsilon\left(\int_0^1\mathcal{J}_{-b,\varepsilon}\left(r\right)dr\right)\widehat{\kappa}_\pi^\top\mu_X\\
& -2\left(\int_0^1\left(1-\exp\left(-br\right)\right)dr\right)\widehat{\kappa}_\pi^\top\mu_X\mathbf{c}^\top\mu_X+\widehat{\kappa}_\pi^\top\mathbf{M}_X\widehat{\kappa}_\pi,
\end{aligned}
$$

$$
n\widehat{\beta}_n^2\sum_{t=1}^nX_tX_t^\top=\left[n\left(\widehat{\beta}_n-\beta_n\right)\right]^2\cdot n^{-1}\sum_{t=1}^nX_tX_t^\top\Rightarrow\widehat{\lambda}_\beta^2\left(\widehat{\kappa}_\pi\right)\mathbf{M}_X,
$$

and

$$
\begin{aligned}
& -n^{-1/2}\widehat{\beta}_n\sum_{t=1}^nX_t\left(y_{t-1}-X_t^\top\widehat{\pi}_n\right)=n\widehat{\beta}_n\left(n^{-3/2}\sum_{t=1}^nX_tX_t^\top\widehat{\pi}_n-n^{-3/2}\sum_{t=1}^nX_ty_{t-1}\right)\\
\Rightarrow\ & \widehat{\lambda}_\beta\left(\widehat{\kappa}_\pi\right)\left\{\mathbf{M}_X\widehat{\kappa}_\pi-\sigma_\varepsilon\left(\int_0^1\mathcal{J}_{-b,\varepsilon}\left(r\right)dr\right)\mu_X-\left(\int_0^1\left(1-\exp\left(-br\right)\right)dr\right)\mu_X\mathbf{c}^\top\mu_X\right\},
\end{aligned}
$$

where $\widehat{\lambda}_\beta\left(\widehat{\kappa}_\pi\right)=\widehat{\lambda}_\beta\left(\widehat{\kappa}_\pi\left(b,\mathbf{c};\varphi_0\right),b,\mathbf{c};\varphi_0\right)$ and $\widehat{\kappa}_\pi=\widehat{\kappa}_\pi\left(b,\mathbf{c};\varphi_0\right)$. And the results follow by Theorem 2 and Lemma 3.

For 3., it suffices to show $\widehat{\sigma}_n^2\xrightarrow{p}\sigma_\varepsilon^2$ and $B^{-1}\left(h\right)\widehat{\mathbf{V}}_nB^{-1}\left(h\right)\xrightarrow{p}\mathcal{V}_h\left(b;\varphi_0\right)$. For $\widehat{\sigma}_n^2$, by Lemma 7, Theorem 3, and the Kolmogorov law of large number (White, 2001, Theorem 3.1,

p. 32),

$$\begin{aligned}
\widehat{\sigma}_n^2 &= n^{-1}\sum_{t=1}^{n}\left[y_t-\left(1-\widehat{\beta}_n\right)y_{t-1}-\widehat{\beta}_n X_t^\top\widehat{\pi}_n\right]^2 \\
&= n^{-1}\sum_{t=1}^{n}\left[\varepsilon_t+\left(\widehat{\beta}_n-\beta_n\right)y_{t-1}-\beta_n X_t^\top\left(\widehat{\pi}_n-\pi_0\right)-\left(\widehat{\beta}_n-\beta_n\right)X_t^\top\widehat{\pi}_n\right]^2 \\
&= n^{-1}\sum_{t=1}^{n}\varepsilon_t^2+O_p\left(n^{-1}\right)\xrightarrow{p}\sigma_\varepsilon^2.
\end{aligned}$$

For $B^{-1}(h)\,\widehat{\mathbf{V}}_n B^{-1}(h)$, by equations (3.4) and (2.11),

$$B^{-1}(h)\,\widehat{\mathbf{V}}_n B^{-1}(h)=\begin{bmatrix} n^{-1-h}\sum_{t=1}^{n}\left(y_{t-1}-X_t^\top\widehat{\pi}_n\right)^2 & \\ -n^{-1-h/2}\widehat{\beta}_n\sum_{t=1}^{n}X_t\left(y_{t-1}-X_t^\top\widehat{\pi}_n\right) & n^{-1+2h}\widehat{\beta}_n^2\sum_{t=1}^{n}X_tX_t^\top \end{bmatrix}.$$

The results follow by Lemma 7, Theorem 3, and the law of large number for stationary ergodic sequences (White, 2001, Theorem 3.34, p. 44),

$$\begin{aligned}
n^{-1-h}\sum_{t=1}^{n}\left(y_{t-1}-X_t^\top\widehat{\pi}_n\right)^2 &= n^{-1-h}\sum_{t=1}^{n}y_{t-1}^2+O_p\left(n^{-1+h}\right)\xrightarrow{p}\frac{\sigma_\varepsilon^2}{2b}, \\
n^{-1+2h}\widehat{\beta}_n^2\sum_{t=1}^{n}X_tX_t^\top &= \left(n^h\left(\widehat{\beta}_n-\beta_n\right)+b\right)^2 n^{-1}\sum_{t=1}^{n}X_tX_t^\top\xrightarrow{p}b^2\mathbf{M}_X, \\
-n^{-1-h/2}\widehat{\beta}_n\sum_{t=1}^{n}X_t\left(y_{t-1}-X_t^\top\widehat{\pi}_n\right) &= O_p\left(n^{-1/2-h/2}\right)\xrightarrow{p}\mathbf{0}_{d_\pi\times 1}.
\end{aligned}$$

And the remains directly follow by Theorem 3. ■

Proof. (Theorem 5) We first prove 2., *i.e.*,

$$\liminf_{n\to\infty}\inf_{\gamma_n\in\Gamma}CP_n^{R,LF}(\gamma_n)=\min\left\{\inf_{\{b,\mathbf{c}\}\in\mathcal{H}(\mathbf{R},\upsilon)}CP_\infty^{L,LF}(b,\mathbf{c}),CP_\infty^{D}\right\}=1-\alpha.$$

where, by definition,

$$AsySz\left(CS_n^{R,LF}\left(\mathbf{R}\theta_n;1-\alpha,\varphi_0\right)\right)=\liminf_{n\to\infty}\inf_{\gamma_n\in\Gamma}CP_n^{R,LF},$$

and

$$
\begin{aligned}
CP_n^{R,LF}(\gamma_n) &= \mathbb{P}\left(W_n(\upsilon) \le \max\left\{\sup_{\{b,\mathbf{c}\}\in\mathcal{H}(\mathbf{R},\upsilon)} \xi_{1-\alpha}\left(\mathcal{W}\left(b_\upsilon,\mathbf{c}_\upsilon;\varphi_0\right)\right), \chi^2_{d_r,1-\alpha}\right\}\right), \\
CP_\infty^{L,LF}(b,\mathbf{c}) &= \liminf_{n\to\infty}\mathbb{P}\left(W_n(\upsilon) \le \sup_{\{b,\mathbf{c}\}\in\mathcal{H}(\mathbf{R},\upsilon)} \xi_{1-\alpha}\left(\mathcal{W}\left(b_\upsilon,\mathbf{c}_\upsilon;\varphi_0\right)\right)\middle|\, \gamma_n\in\Gamma_n(1,b,\mathbf{c})\right) = 1-\alpha, \\
CP_\infty^{D} &= \liminf_{n\to\infty}\mathbb{P}\left(W_n(\upsilon)\le\chi^2_{d_r,1-\alpha}\middle|\,\theta_n=\theta_0\in\Theta_n^* \quad\text{or}\quad \gamma_n\in\Gamma_n(h,b,\mathbf{c})\right) = 1-\alpha.
\end{aligned}
$$

Our proof is similar to the proof of Lemma 2.1 in Andrews and Cheng (2012). Because in the problem considered in this paper, the parameter causing the potential weak identification, β, is only one-dimensional, our proof is much simpler.

Since for any function $f_n(x)$,

$$
\inf_x \liminf_{n\to\infty} f_n(x) \le \liminf_{n\to\infty}\inf_x f_n(x),
$$

therefore

$$
\liminf_{n\to\infty}\inf_{\gamma_n\in\Gamma} CP_n^{R,LF} \le \min\left\{\inf_{\{b,\mathbf{c}\}\in\mathcal{H}(\mathbf{R},\upsilon)} CP_\infty^{L,LF}(b,\mathbf{c}), CP_\infty^D\right\} = 1-\alpha.
$$

Let $\{\gamma_n^*\in\Gamma_n : n\in\mathbb{N}\}$ be a sequence such that

$$
\liminf_{n\to\infty} CP_n^{R,LF}(\gamma_n^*) = \liminf_{n\to\infty}\inf_{\gamma_n\in\Gamma} CP_n^{R,LF}(\gamma_n).
$$

Such a sequence always exists since according to the axiom of choice, we can always select each element in this sequence as the infimizer of CP_n^L for every $n\in\mathbb{N}$. Let $\gamma_n^* = \{\theta_n^*,\zeta_0\} = \{\beta_n^*,\pi_n^*,\zeta_0\}$. Then by Definition 1, either $\gamma_n^*\in\Gamma_n(1,b,\mathbf{c})$, or $\theta_n^*=\theta_0\in\Theta_n^*$ / $\gamma_n^*\in\Gamma_n(h,b,\mathbf{c})$. In the former case $\liminf_{n\to\infty} CP_n^{R,LF}(\gamma_n^*) = CP_\infty^L(b^*,\mathbf{c}^*)$, where $b^* = n^{-1}\beta_n^*$ and $\mathbf{c}^* = n^{1/2}\pi_n^*$, and $CP_\infty^L(b^*,\mathbf{c}^*) = 1-\alpha$ for all $\{b^*,\mathbf{c}^*\}\in\mathcal{H}(\mathbf{R},\upsilon)$. In the later case $\liminf_{n\to\infty} CP_n^{R,LF}(\gamma_n^*) = CP_\infty^D = 1-\alpha$. Therefore

$$
\liminf_{n\to\infty}\inf_{\gamma_n\in\Gamma} CP_n^{R,LF} \ge \min\left\{\inf_{\{b,\mathbf{c}\}\in\mathcal{H}(\mathbf{R},\upsilon)} CP_\infty^{L,LF}(b,\mathbf{c}), CP_\infty^D\right\} = 1-\alpha.
$$

1. directly follows since 1. is a special case of 2. For 3., since

$$
AsySz\left(CS_n^R(\mathbf{Q}\theta_n;1-\alpha,\varphi_0)\right) = 1-\alpha,
$$

and for any set $\mathcal{C}$, $\{\theta_n \in \mathcal{C}\}$ entails $\{\mathbf{P}\theta_n \in \mathbf{P}\mathcal{C}\}$, therefore,

$$\begin{aligned} & \liminf_{n\to\infty} \inf_{\gamma_n\in\Gamma} \mathbb{P}\left(\mathbf{R}\theta_n \in CS_n^{R,P}\left(\mathbf{R}\theta_n; 1-\alpha, \varphi_0\right)\right) \\ \geq & \liminf_{n\to\infty} \inf_{\gamma_n\in\Gamma} \mathbb{P}\left(\mathbf{Q}\theta_n \in CS_n^{R}\left(\mathbf{Q}\theta_n; 1-\alpha, \varphi_0\right)\right) = 1-\alpha. \end{aligned}$$

■

8 Appendix C: Supplementary Results and Proofs

This section states and proves some results used in the proofs of the theorems.

Lemma 6 *Suppose that Assumptions 1, 2 and 3 hold, $y_0 = o_p\left(n^{1/2}\right)$, and $\gamma_n \in \Gamma_n\left(1, b, \mathbf{c}\right)$. Let $\mathcal{W}_\varepsilon(\cdot)$ and $\mathcal{W}_X(\cdot)$ be two standard Wiener processes (one-dimensional and d_π-dimensional, respectively), and $\mathcal{J}_{-b,\varepsilon}(\cdot)$ and $\mathcal{J}_{-b,X}(\cdot)$ be an Ornstein–Uhlenbeck process. For any $r \in [0,1]$, when $n \to \infty$,*

$$\begin{aligned} n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor}\varepsilon_t &\Rightarrow \sigma_\varepsilon \mathcal{W}_\varepsilon(r), \quad n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor}\left(X_{t-i}-\mu_X\right)^\top \Rightarrow \mathbf{\Sigma}_X^{1/2}\mathcal{W}_X(r), \\ \mathcal{J}_{-b,\varepsilon}(r) &= \int_0^r \exp\left(-b\left(r-s\right)\right) d\mathcal{W}_\varepsilon(s) \quad \text{and} \quad \mathcal{J}_{-b,X}(r) = \int_0^r \exp\left(-b\left(r-s\right)\right) d\mathcal{W}_X(s) \end{aligned}$$

Then as $n \to \infty$, we have the following results.

1. $n^{-1/2}y_{\lfloor nr\rfloor} \Rightarrow \sigma_\varepsilon \mathcal{J}_{-b,\varepsilon}(r) + \mathbf{c}^\top \mu_X\left(1-\exp\left(-br\right)\right)$.
2. $n^{-3/2}\sum_{t=1}^n y_{t-1} \Rightarrow \sigma_\varepsilon \int_0^1 \mathcal{J}_{-b,\varepsilon}(r)\,dr + \mathbf{c}^\top\mu_X\left(\int_0^1\left(1-\exp\left(-br\right)\right)dr\right)$.
3. $n^{-2}\sum_{t=1}^n y_{t-1}^2 \Rightarrow \sigma_\varepsilon^2 \int_0^1 \mathcal{J}_{-b,\varepsilon}^2(r)\,dr + 2\sigma_\varepsilon \mathbf{c}^\top\mu_X\left(\int_0^1\left(1-\exp\left(-br\right)\right)\mathcal{J}_{-b,\varepsilon}(r)\,dr\right)$
$+\left(\mathbf{c}^\top\mu_X\right)^2\left(\int_0^1\left(1-\exp\left(-br\right)\right)^2 dr\right)$.
4. $n^{-1}\sum_{t=1}^n y_{t-1}\varepsilon_t \Rightarrow \sigma_\varepsilon^2 \int_0^1 \mathcal{J}_{-b,\varepsilon}(r)\,d\mathcal{W}_\varepsilon(r) + \sigma_\varepsilon\mathbf{c}^\top\mu_X\left(\int_0^1\left(1-\exp\left(-br\right)\right)d\mathcal{W}_\varepsilon(r)\right)$.
5. $n^{-1}\sum_{t=1}^n \left(X_t-\mu_X\right)y_{t-1} \Rightarrow \sigma_\varepsilon \mathbf{\Sigma}_X^{1/2}\int_0^1 \mathcal{J}_{-b,\varepsilon}(r)\,d\mathcal{W}_X(r)$
$+\mathbf{c}^\top\mu_X\mathbf{\Sigma}_X^{1/2}\left(\int_0^1\left(1-\exp\left(-br\right)\right)d\mathcal{W}_X(r)\right)$.
6. $n^{-3/2}\sum_{t=1}^n X_t y_{t-1} \Rightarrow \sigma_\varepsilon\left(\int_0^1 \mathcal{J}_{-b,\varepsilon}(r)\,dr\right)\mu_X + \left(\int_0^1\left(1-\exp\left(-br\right)\right)dr\right)\mu_X\mathbf{c}^\top\mu_X$.

Proof. For 1., under Assumption 1, equation (1.2) can be written as:

$$\begin{aligned} y_{\lfloor nr \rfloor} &= \sum_{i=0}^{\infty} (1-\beta_n)^i \varepsilon_{\lfloor nr \rfloor - i} + \beta_n \sum_{i=0}^{\infty} (1-\beta_n)^i X_{\lfloor nr \rfloor - i}^{\top} \pi_n \\ &= (1-\beta_n)^{\lfloor nr \rfloor} y_0 + \sum_{i=0}^{\lfloor nr \rfloor - 1} (1-\beta_n)^i \varepsilon_{t-i} + \beta_n \sum_{i=0}^{\lfloor nr \rfloor - 1} (1-\beta_n)^i \mu_X^{\top} \pi_n \\ &\quad + \beta_n \sum_{i=0}^{\lfloor nr \rfloor - 1} (1-\beta_n)^i \left(X_{\lfloor nr \rfloor - i} - \mu_X \right)^{\top} \pi_n. \end{aligned}$$

Where $(1-\beta_n)^{\lfloor nr \rfloor} \to \exp(-br)$, $\beta_n \sum_{i=0}^{\lfloor nr \rfloor - 1} (1-\beta_n)^i = 1 - (1-\beta_n)^{\lfloor nr \rfloor} \to 1 - \exp(-br)$, and for any $r \in [0,1]$, by Lemma 1 of Phillips (1987), as $n \to \infty$,

$$n^{-1/2} \sum_{i=0}^{\lfloor nr \rfloor - 1} (1-\beta_n)^i \varepsilon_{\lfloor nr \rfloor - i} \Rightarrow \sigma_{\varepsilon} \mathcal{J}_{-b,\varepsilon}(r), \quad n^{-1/2} \sum_{i=0}^{\lfloor nr \rfloor - 1} (1-\beta_n)^i \left(X_{\lfloor nr \rfloor - i} - \mu_X \right) \Rightarrow \mathbf{\Sigma}_X^{1/2} \mathcal{J}_{-b,X}(r).$$

Therefore for any $r \in [0,1]$, as $n \to \infty$, 1. follows by

$$\begin{aligned} n^{-1/2} y_{\lfloor nr \rfloor} &= n^{-1/2} \sum_{i=0}^{\lfloor nr \rfloor - 1} (1-\beta_n)^i \varepsilon_{\lfloor nr \rfloor - i} + n^{-1/2} \beta_n \sum_{i=0}^{\lfloor nr \rfloor - 1} (1-\beta_n)^i \mu_X^{\top} \pi_n + o_p(1) \\ &\Rightarrow \sigma_{\varepsilon} \mathcal{J}_{-b,\varepsilon}(r) + \mathbf{c}^{\top} \mu_X (1 - \exp(-br)). \end{aligned}$$

2. – 5. follow by

$$
\begin{aligned}
n^{-3/2}\sum_{t=1}^{n} y_{t-1} &\Rightarrow \int_0^1 \left[\sigma_\varepsilon \mathcal{J}_{-b,\varepsilon}(r) + \mathbf{c}^\top \mu_X \left(1-\exp(-br)\right)\right] dr \\
&= \sigma_\varepsilon \int_0^1 \mathcal{J}_{-b,\varepsilon}(r)\, dr + \mathbf{c}^\top \mu_X \left(\int_0^1 \left(1-\exp(-br)\right) dr\right), \\
n^{-2}\sum_{t=1}^{n} y_{t-1}^2 &\Rightarrow \int_0^1 \left[\sigma_\varepsilon \mathcal{J}_{-b,\varepsilon}(r) + \mathbf{c}^\top \mu_X \left(1-\exp(-br)\right)\right]^2 dr \\
&= \sigma_\varepsilon^2 \int_0^1 \mathcal{J}_{-b,\varepsilon}^2(r)\, dr + 2\sigma_\varepsilon \mathbf{c}^\top \mu_X \left(\int_0^1 \left(1-\exp(-br)\right) \mathcal{J}_{-b,\varepsilon}(r)\, dr\right) \\
&\quad + \left(\mathbf{c}^\top \mu_X\right)^2 \left(\int_0^1 \left(1-\exp(-br)\right)^2 dr\right), \\
n^{-1}\sum_{t=1}^{n} y_{t-1}\varepsilon_t &\Rightarrow \int_0^1 \left[\sigma_\varepsilon \mathcal{J}_{-b,\varepsilon}(r) + \mathbf{c}^\top \mu_X \left(1-\exp(-br)\right)\right] d\sigma_\varepsilon \mathcal{W}_\varepsilon(r) \\
&= \sigma_\varepsilon^2 \int_0^1 \mathcal{J}_{-b,\varepsilon}(r)\, d\mathcal{W}_\varepsilon(r) + \sigma_\varepsilon \mathbf{c}^\top \mu_X \left(\int_0^1 \left(1-\exp(-br)\right) d\mathcal{W}_\varepsilon(1)\right), \\
n^{-1}\sum_{t=1}^{n} \left(X_t - \mu_X\right) y_{t-1} &\Rightarrow \int_0^1 \left[\sigma_\varepsilon \mathcal{J}_{-b,\varepsilon}(r) + \mathbf{c}^\top \mu_X \left(1-\exp(-br)\right)\right] d\mathbf{\Sigma}_X^{1/2} \mathcal{W}_X(r) \\
&= \sigma_\varepsilon \mathbf{\Sigma}_X^{1/2} \int_0^1 \mathcal{J}_{-b,\varepsilon}(r)\, d\mathcal{W}_X(r) + \mathbf{c}^\top \mu_X \mathbf{\Sigma}_X^{1/2} \left(\int_0^1 \left(1-\exp(-br)\right) d\mathcal{W}_X(r)\right).
\end{aligned}
$$

And for 6., by 2. and 5.,

$$
\begin{aligned}
n^{-3/2}\sum_{t=1}^{n} X_t y_{t-1} &= n^{-3/2}\mu_X \sum_{t=1}^{n} y_{t-1} + O_p\left(n^{-1/2}\right) \\
&\Rightarrow \sigma_\varepsilon \left(\int_0^1 \mathcal{J}_{-b,\varepsilon}(r)\, dr\right) \mu_X + \left(\int_0^1 \left(1-\exp(-br)\right) dr\right) \mu_X \mathbf{c}^\top \mu_X.
\end{aligned}
$$

■

Lemma 7 *Suppose that Assumption 1 holds and* $\gamma_n \in \Gamma_n(h, b, \mathbf{c})$. *Then as* $n \to \infty$*:*

1. $n^{-1/2-h}\sum_{t=1}^{n} y_t \overset{A}{\sim} \mathcal{N}\left(\mathbf{c}^\top \mu_X, b^{-2}\sigma_\varepsilon^2\right)$
2. $n^{-1/2-h/2}\sum_{t=1}^{n} y_{t-1}\varepsilon_t \overset{A}{\sim} \mathcal{N}\left(0, (2b)^{-1}\sigma_\varepsilon^4\right)$.
3. $n^{-1-h}\sum_{t=1}^{n} y_{t-1}^2 \xrightarrow{p} (2b)^{-1}\sigma_\varepsilon^2$.
4. $n^{-1/2-h/2}\sum_{t=1}^{n} \left(X_t - \mu_X\right) y_{t-1} \overset{A}{\sim} \mathcal{N}\left(0, (2b)^{-1}\sigma_\varepsilon^2 \mathbf{\Sigma}_X\right)$.

5. $n^{-1/2-h}\sum_{t=1}^{n} X_t y_{t-1} \overset{A}{\sim} \mathcal{N}\left(\mu_X \mathbf{c}^\top \mu_X, b^{-2}\sigma_\varepsilon^2 \mu_X \mu_X^\top\right).$

Proof. Let

$$\eta_t = \sum_{i=0}^{\infty} (1-\beta_n)^i \varepsilon_{t-i}, \quad \text{and} \quad \xi_t = \sum_{i=0}^{\infty} (1-\beta_n)^i (X_{t-i} - \mu_X).$$

Then by Assumption 1, equation (1.2) can be written as:

$$\begin{aligned} y_t &= \sum_{i=0}^{\infty} (1-\beta_n)^i \varepsilon_{t-i} + \beta_n \sum_{i=0}^{\infty} (1-\beta_n)^i X_{t-i}^\top \pi_n \\ &= \sum_{i=0}^{\infty} (1-\beta_n)^i \varepsilon_{t-i} + \beta_n \sum_{i=0}^{\infty} (1-\beta_n)^i \mu_X^\top \pi_n + \beta_n \sum_{i=0}^{\infty} (1-\beta_n)^i (X_{t-i} - \mu_X)^\top \pi_n \\ &= \mu_X^\top \pi_n + \eta_t + \beta_n \xi_t \pi_n. \end{aligned}$$

By Theorem 2, Lemma 1 and Lemma 2 of Giraitis and Phillips (2006), as $n \to \infty$,

$$\begin{aligned} n^{-1/2-h}\sum_{t=1}^{n} \eta_t &= b^{-1} n^{-1/2} (1-\rho_n) \sum_{t=1}^{n} \eta_t \overset{A}{\sim} \mathcal{N}\left(0, \frac{\sigma_\varepsilon^2}{b^2}\right), \\ n^{-1/2-h}\sum_{t=1}^{n} \xi_t &= b^{-1} n^{-1/2} (1-\rho_n) \sum_{t=1}^{n} \xi_t \overset{A}{\sim} \mathcal{N}\left(0, \frac{1}{b^2}\mathbf{\Sigma}_X\right), \\ n^{-1/2-h/2}\sum_{t=1}^{n} \eta_{t-1}\varepsilon_t &= \left(2b - n^{-h}b^2\right)^{-1/2} n^{-1/2} \left(1-\rho_n^2\right)^{1/2} \sum_{t=1}^{n} \eta_{t-1}\varepsilon_t \overset{A}{\sim} \mathcal{N}\left(0, \frac{\sigma_\varepsilon^4}{2b}\right), \\ n^{-1/2-h/2}\sum_{t=1}^{n} \xi_{t-1}\varepsilon_t &= \left(2b - n^{-h}b^2\right)^{-1/2} n^{-1/2} \left(1-\rho_n^2\right)^{1/2} \sum_{t=1}^{n} \eta_{t-1}\varepsilon_t \overset{A}{\sim} \mathcal{N}\left(0, \frac{\sigma_\varepsilon^2}{2b}\mathbf{\Sigma}_X\right), \\ n^{-1-h}\sum_{t=1}^{n} \eta_{t-1}^2 &= \left(2b - n^{-h}b^2\right)^{-1} n^{-1} \left(1-\rho_n^2\right) \sum_{t=1}^{n} \eta_{t-1}^2 \xrightarrow{p} \frac{\sigma_\varepsilon^2}{2b}, \quad \text{and} \\ n^{-1-h}\sum_{t=1}^{n} \xi_{t-1}\xi_{t-1}^\top &= \left(2b - n^{-h}b^2\right)^{-1} n^{-1} \left(1-\rho_n^2\right) \sum_{t=1}^{n} \xi_{t-1}\xi_{t-1}^\top \xrightarrow{p} \frac{1}{2b}\mathbf{\Sigma}_X. \end{aligned}$$

Therefore,

$$\begin{aligned}
n^{-1/2-h}\sum_{t=1}^{n} y_t &= \mathbf{c}^{\top}\mu_X + n^{-1/2-h}\sum_{t=1}^{n}\eta_t + o_p(1) \overset{A}{\sim} \mathcal{N}\left(\mathbf{c}^{\top}\mu_X, \frac{\sigma_\varepsilon^2}{b^2}\right), \\
n^{-1/2-h/2}\sum_{t=1}^{n} y_{t-1}\varepsilon_t &= n^{-1/2-h/2}\sum_{t=1}^{n}\eta_{t-1}\varepsilon_t + o_p(1) \overset{A}{\sim} \mathcal{N}\left(0, \frac{\sigma_\varepsilon^4}{2b}\right), \\
n^{-1-h}\sum_{t=1}^{n} y_{t-1}^2 &= n^{-1-h}\sum_{t=1}^{n}\eta_{t-1}^2 + o_p(1) \xrightarrow{p} \frac{\sigma_\varepsilon^2}{2b}, \\
n^{-1/2-h/2}\sum_{t=1}^{n}(X_t-\mu_X)\,y_{t-1} &= n^{-1/2-h/2}\sum_{t=1}^{n}\eta_{t-1}(X_t-\mu_X) + o_p(1) \overset{A}{\sim} \mathcal{N}\left(0, \frac{\sigma_\varepsilon^2}{2b}\mathbf{\Sigma}_X\right).
\end{aligned}$$

And for 5., by 1. and 4.,

$$n^{-1/2-h}\sum_{t=1}^{n} X_t y_{t-1} = n^{-1/2-h}\mu_X\sum_{t=1}^{n} y_{t-1} + o_p(1) \overset{A}{\sim} \mathcal{N}\left(\mu_X\mathbf{c}^{\top}\mu_X, \frac{\sigma_\varepsilon^2}{b^2}\mu_X\mu_X^{\top}\right).$$

■

References

[1] Anatolyev, S., Gospodinov, N., 2011. Methods for Estimation and Inference in Modern Econometrics. Chapman and Hall/CRC Press. Boca Raton, Florida, U.S.

[2] Andrews, D.W.K., 1994. Empirical process methods in econometrics. In: Engle, R.F., McFadden, D.L., (Eds.), Handbook of Econometrics, vol. IV. Elsevier, Amsterdam, The Netherlands.

[3] Andrews, D.W.K., Cheng, X., 2012. Estimation and inference with weak, semi-strong, and strong identification. Econometrica 80, 2153–2211.

[4] Andrews, D.W.K., Cheng, X., 2013a. GMM estimation and uniform subvector inference with possible identification failure. Econometric Theory, forthcoming.

[5] Andrews, D.W.K., Cheng, X., 2013b. Maximum likelihood estimation and uniform inference with sporadic identification failure. Journal of Econometrics 173, 36–56.

[6] Andrews, D.W.K., Cheng, X., Guggenberger, P., 2011. Generic results for establishing the asymptotic size of confidence sets and tests. Cowles Foundation Discussion Paper No. 1813.

[7] Bunzel, H., Enders, W., 2010. The Taylor rule and "opportunistic" monetary policy. Journal of Money, Credit, and Banking 42, 931–949.

[8] Clarida, R., Galí, J., Gertler, M., 2000. Monetary policy rules and macroeconomic stability: Theory and some evidence. Quarterly Journal of Economics 115, 147–180.

[9] Cochrane, J., 2011. Determinacy and identification with Taylor rules. Journal of Political Economy 119, 565–615.

[10] Croushore, D., Stark, T., 2001. A real-time data set for macroeconomists. Journal of Econometrics 105, 111–130.

[11] Dufour, J.M., 1997. Some impossibility theorems in econometrics with applications to structural and dynamic models. Econometrica 65, 1365–1387.

[12] Elliott, G., Rothenberg, T.J., Stock, J.H., 1996. Efficient tests of an autoregressive unit root. Econometrica 64, 813–836.

[13] Galí, J., 2008. Monetary Policy, Inflation, and the Business Cycle: An Introduction to the ew Keynesian Framework. Princeton University Press, Princeton, New Jersey, U.S.

[14] Giraitis, L., Phillips, P.C.B., 2006. Uniform limit theory for stationary autoregression. Journal of Time Series Analysis 27, 51–60.

[15] Guerron-Quintana, P., Inoue, A., Kilian, L., 2009. Inference in weakly identified DSGE models. Working paper.

[16] Guggenberger, P., 2012. On the asymptotic size distortion of tests when instruments locally violate the exogeneity assumption. Econometric Theory 28, 387–421.

[17] Hansen, B.E., 1999. The grid bootstrap and the autoregressive model. The Review of Economics and Statistics 81, 594–607.

[18] Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

[19] Hodrick, R., Prescott, E.C., 1997. Postwar U.S. business cycles: An empirical investigation. Journal of Money, Credit, and Banking 29, 1–16.

[20] Inoue, A., Kilian, L., 2002. Bootstrapping autoregressive processes with possible unit roots. Econometrica 70, 377–391.

[21] Inoue, A., Rossi, B., 2011. Testing for weak identification in possibly nonlinear models. Journal of Econometrics 161, 246–261.

[22] Jeganathan, P. 1991. On the asymptotic behavior of least-squares estimators in AR time series with roots near the unit circle. Econometric Theory 7, 269–306.

[23] Krinsky, I., Robb, A.L., 1986. On approximating the statistical properties of elasticities. Review of Economics and Statistics 68, 715–719.

[24] Li, J., 2013. Robust estimation and inference for jumps in noisy high frequency data: a local-to-continuity theory for the pre-averaging method. Econometrica 81, 1673–1693.

[25] Mavroeidis, S., 2004. Weak identification of forward-looking models in monetary economics. Oxford Bulletin of Economics and Statistics 66, 609–635.

[26] Mavroeidis, S., 2010. Monetary policy rules and macroeconomic stability: Some new evidence. American Economic Review 100, 491–503.

[27] McCloskey, A., 2011. Bonferroni-based size-correction for nonstandard testing problems. Working paper.

[28] Mikusheva, A., 2007. Uniform inference in autoregressive models. Econometrica 75, 1411–1452.

[29] Mikusheva, A., 2012. One-dimensional inference in autoregressive models with the potential presence of a unit root. Econometrica 80, 173–212.

[30] Nelson, C.R., Startz, R., 2007. The zero-information-limit condition and spurious inference in weakly identified models. Journal of Econometrics 138, 47–62.

[31] Newey, W. K., McFadden, D., 1994. Large sample estimation and hypothesis testing. In: Engle, R.F., McFadden, D.L., (Eds.), Handbook of Econometrics, vol. IV. Elsevier, Amsterdam, The Netherlands.

[32] Nikolsko-Rzhevskyy, A., 2011. Monetary policy estimation in real time: forward-looking Taylor rules without forward-looking data. Journal of Money, Credit and Banking 43, 871–897.

[33] Nikolsko-Rzhevskyy, A., Papell, D.H., 2012. Taylor rules and the great inflation. Journal of Macroeconomics 34, 903–918

[34] Orphanides, A., 2001. Monetary policy rules based on real-time-data. American Economic Review 91, 964–985.

[35] Orphanides, A., 2004. Monetary policy rules, macroeconomic stability, and inflation: A view from the trenches. Journal of Money, Credit, and Banking 36, 151–175.

[36] Phillips, P.C.B., 1987. Towards a unified asymptotic theory for autoregression. Biometrika 74, 535–547.

[37] Phillips, P.C.B., Magdalinos, T., 2007. Limit theory for moderate deviations from a unit root. Journal of Econometrics 136, 115–130.

[38] Stock, J.H., 1991. Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series. Journal of monetary economics 28, 435–459.

[39] Taylor, J.B., 1993. Discretion versus policy rules in practice. Carnegie-Rochester Conference Series on Public Policy 39, 195–214.

[40] van der Meer, T., Pap, G., van Zuijlen, M.C.A., 1999. Asymptotic inference for nearly unstable AR (p) processes. Econometric Theory 15, 184–217.

[41] van der Vaart, A.W., Wellner, L.A., 1996. Weak convergence and empirical processes. New York: Springer.

[42] White, H.L., 2001. Asymptotic theory for econometricians, revised edition. Orlando, FL: Academic Press.

[43] Woodford, M., 2003. Interest and Prices: Foundations of a Theory of Monetary Policy. Princeton, NJ: Princeton University Press.

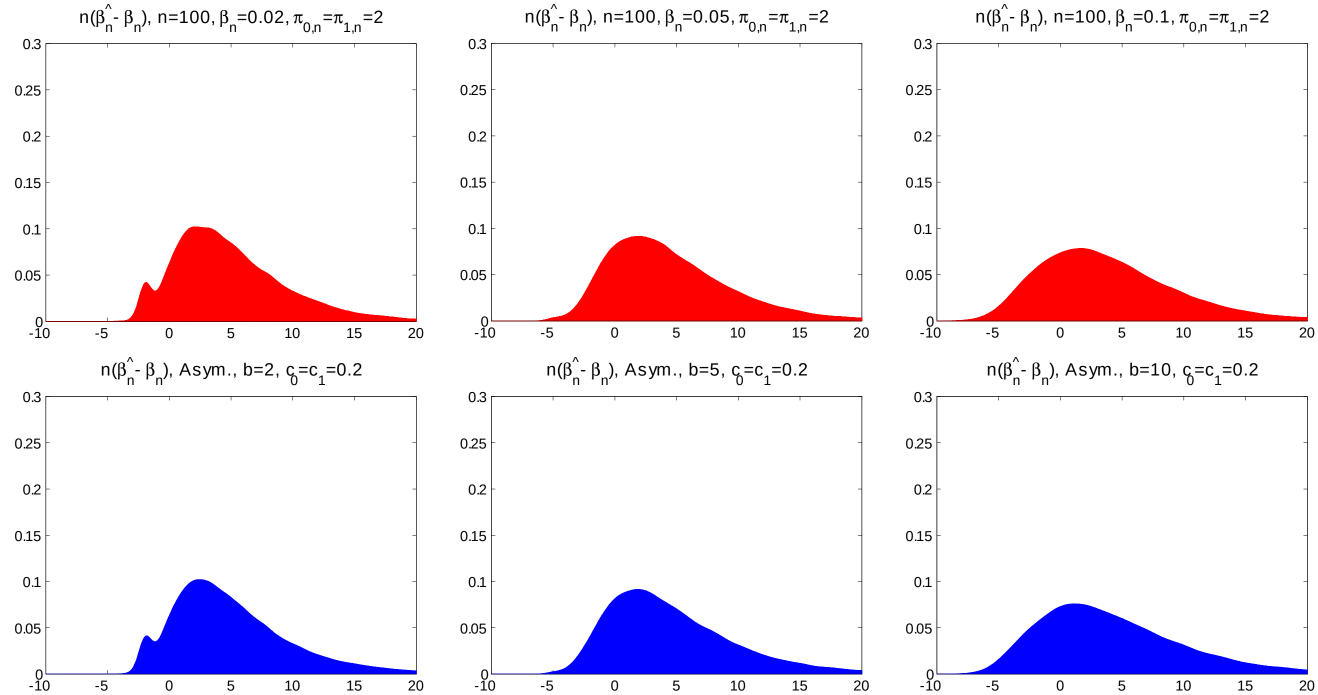
[44] Woutersen, T., Ham, J.C., 2013. Calculating confidence intervals for continuous and discontinuous functions of parameters. Working paper.

Table 1: NLS estimates for the forecast-based monetary policy reaction function

	$\pi_{\dot{p}}$	π_x	π_α	β	σ_ε^2	R^2
$k=0$	0.895	1.171	2.359	0.109	0.198	0.957
	(0.325)	(0.306)	(1.073)	(0.030)		
$k=1$	1.491	0.985	0.765	0.194	0.160	0.965
	(0.211)	(0.148)	(0.654)	(0.034)		

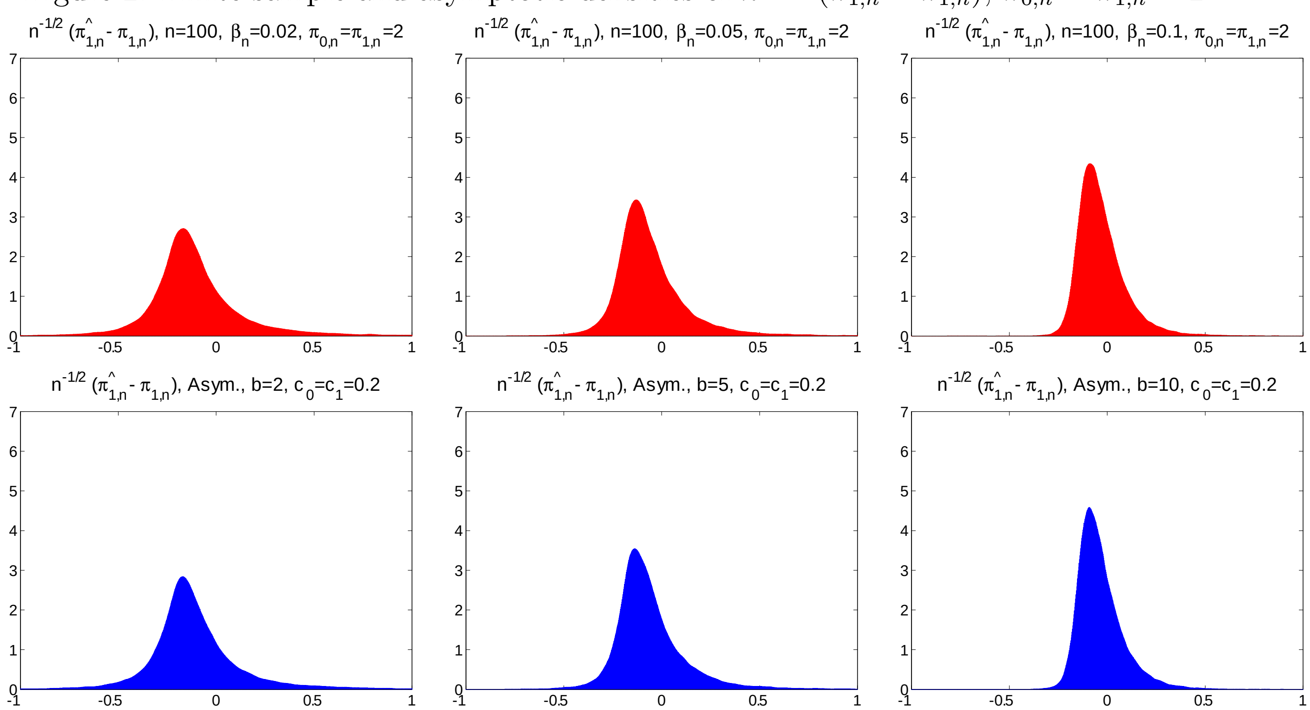
In parentheses are estimates of standard errors according to the standard asymptotic theory as in Theorem 1.

Figure 1: Finite-sample and asymptotic densities of $n\left(\widehat{\beta}_n - \beta_n\right)$, $\pi_{0,n} = \pi_{1,n} = 2$



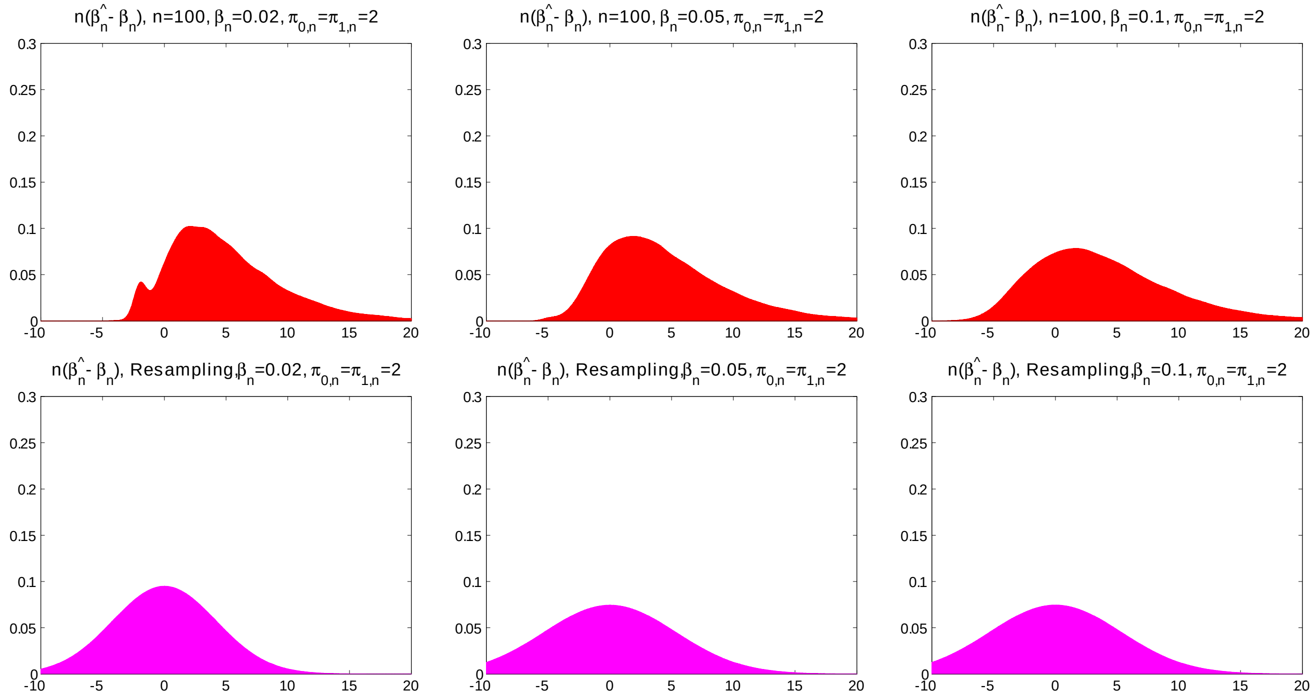
The first and rows are respectively the simulated finite-sample densities and the asymptotic densities in Example 1.

Figure 2: Finite-sample and asymptotic densities of $n^{-1/2}\left(\widehat{\pi}_{1,n} - \pi_{1,n}\right)$, $\pi_{0,n} = \pi_{1,n} = 2$



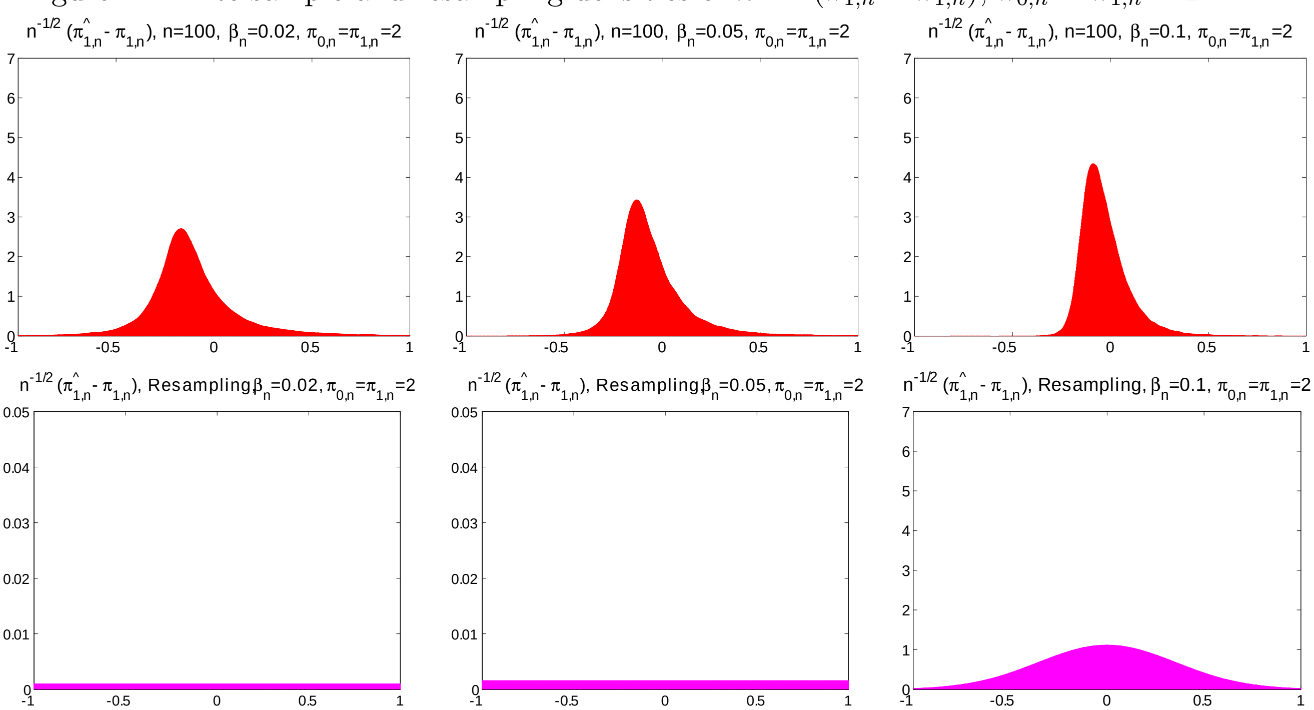
The first and second rows are respectively the simulated finite-sample densities and the asymptotic densities in Example 1.

Figure 3: Finite-sample and resampling densities of $n\left(\widehat{\beta}_n - \beta_n\right)$, $\pi_{0,n} = \pi_{1,n} = 2$



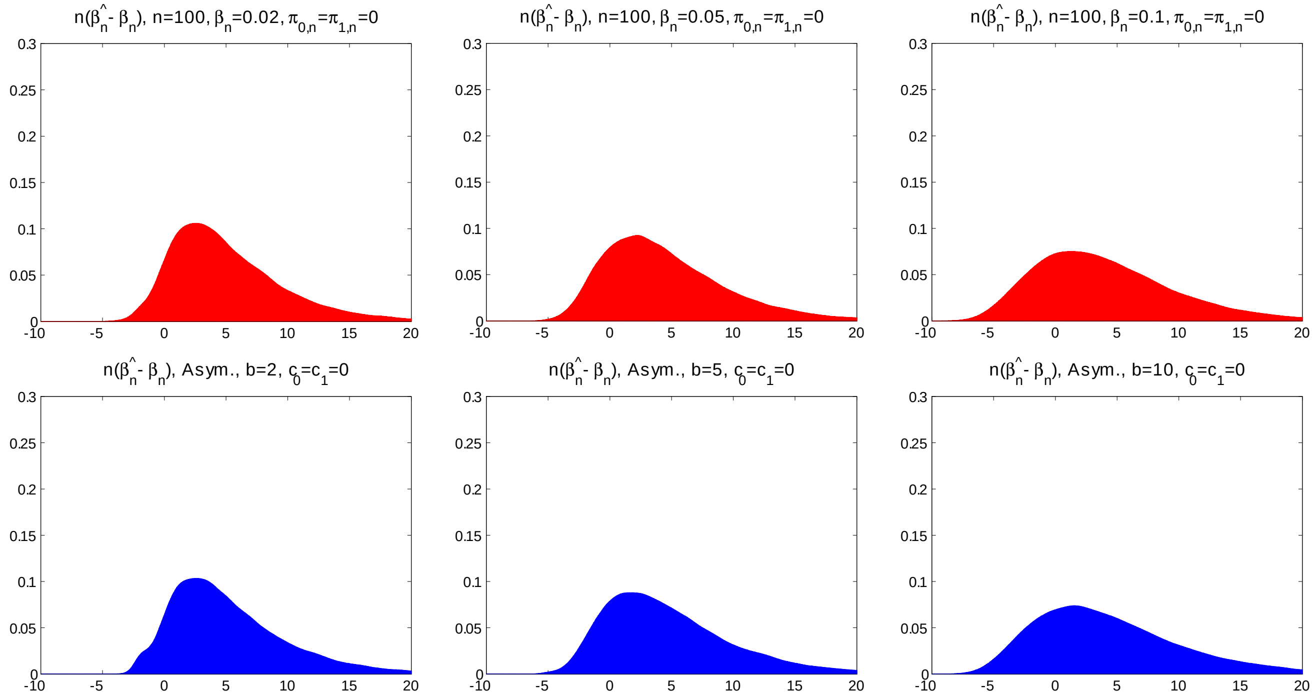
The first and second rows are respectively the simulated finite-sample densities and the resampling densities in Example 1.

Figure 4: Finite-sample and resampling densities of $n^{-1/2}\left(\widehat{\pi}_{1,n} - \pi_{1,n}\right)$, $\pi_{0,n} = \pi_{1,n} = 2$



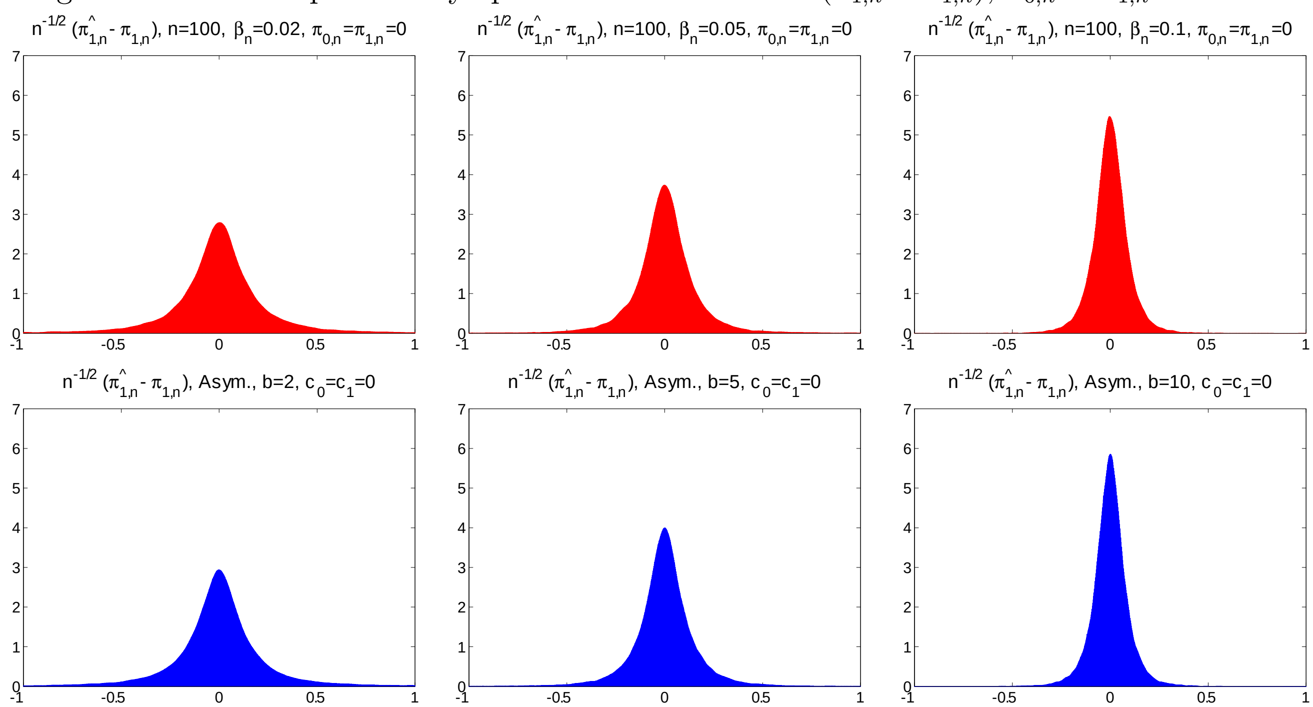
The first and second rows are respectively the simulated finite-sample densities and the resampling densities in Example 1.

Figure 5: Finite-sample and asymptotic densities of $n\left(\widehat{\beta}_n - \beta_n\right)$, $\pi_{0,n} = \pi_{1,n} = 0$



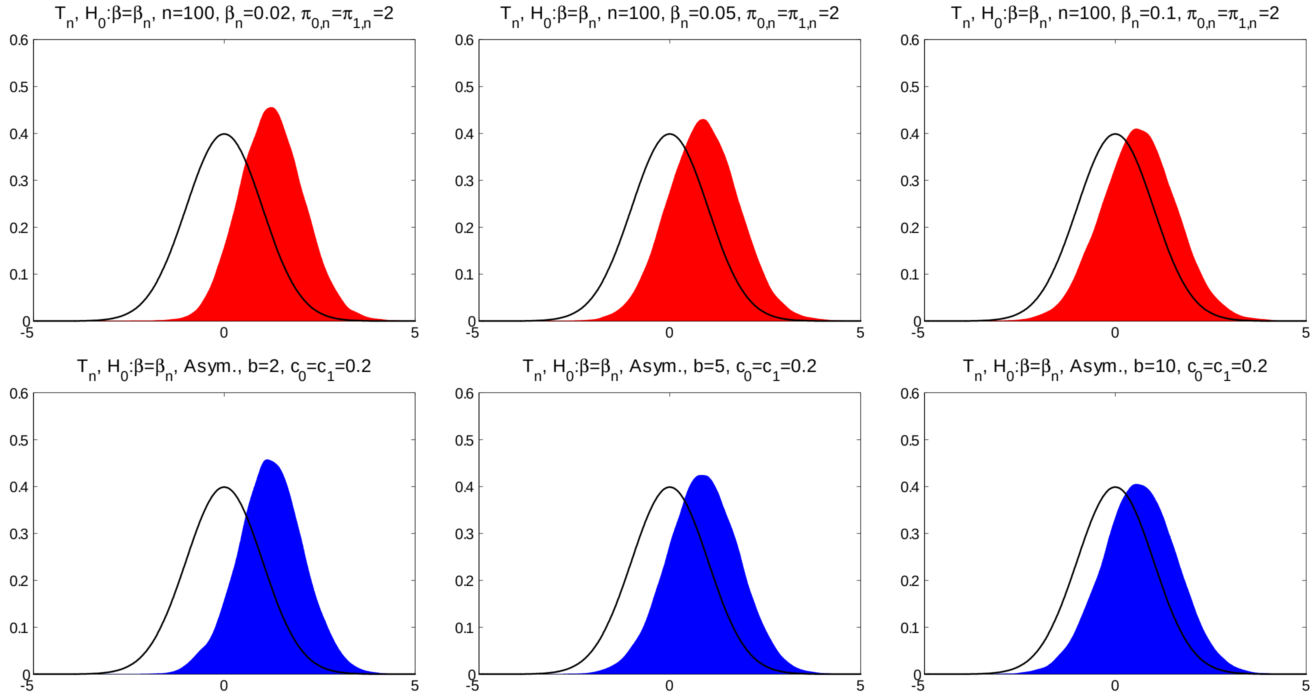
The first and rows are respectively the simulated finite-sample densities and the asymptotic densities in Example 2.

Figure 6: Finite-sample and asymptotic densities of $n^{-1/2}\left(\widehat{\pi}_{1,n} - \pi_{1,n}\right)$, $\pi_{0,n} = \pi_{1,n} = 0$



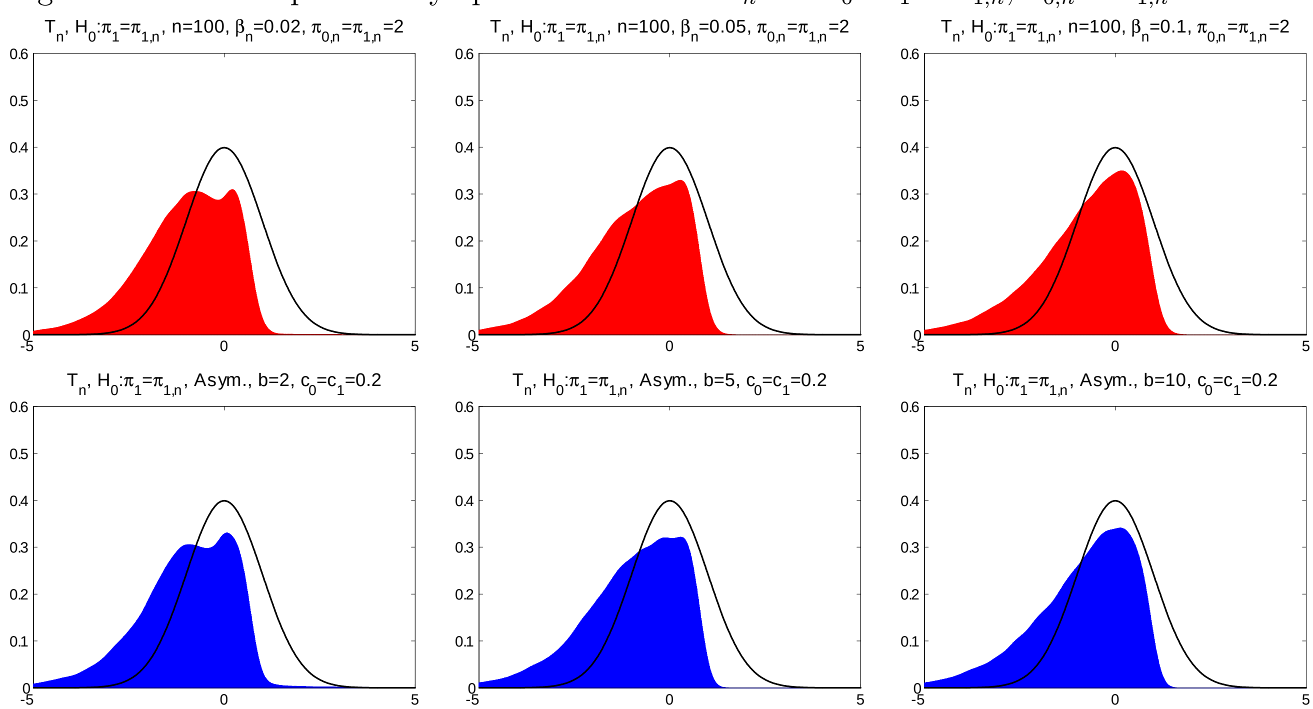
The first and second rows are respectively the simulated finite-sample densities and the asymptotic densities in Example 2.

Figure 7: Finite-sample and asymptotic densities of T_n for $H_0 : \beta = \beta_n$, $\pi_{0,n} = \pi_{1,n} = 2$



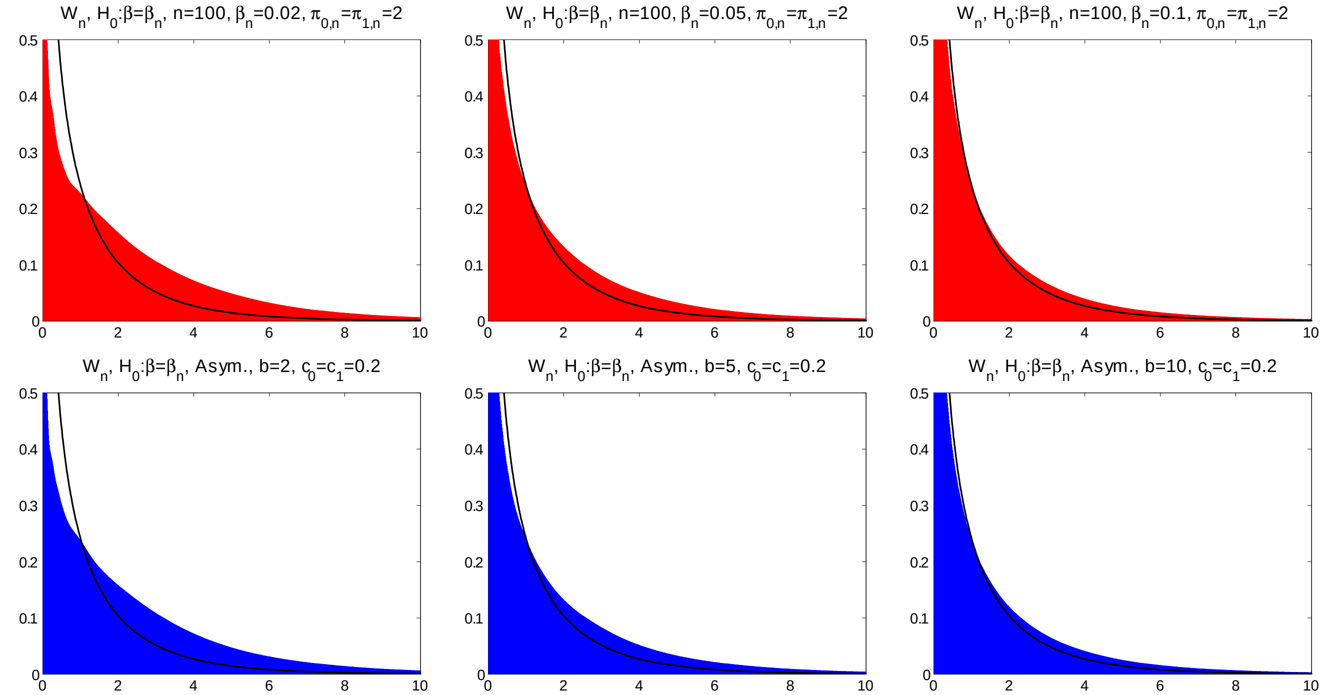
The first and second rows are respectively the simulated finite-sample densities and the asymptotic densities in Example 3.

Figure 8: Finite-sample and asymptotic densities of T_n for $H_0 : \pi_1 = \pi_{1,n}$, $\pi_{0,n} = \pi_{1,n} = 2$



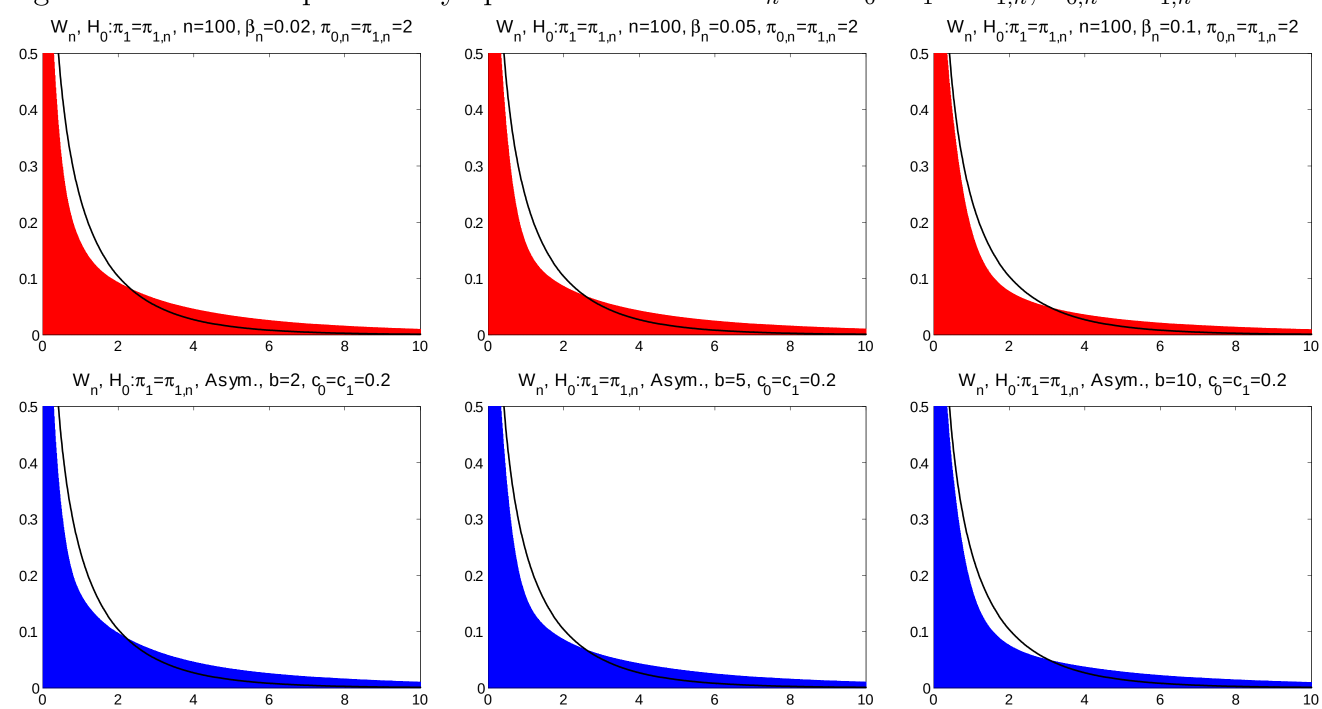
The first and second rows are respectively the simulated finite-sample densities and the asymptotic densities in Example 3.

Figure 9: Finite-sample and asymptotic densities of W_n for $H_0 : \beta = \beta_n$, $\pi_{0,n} = \pi_{1,n} = 2$



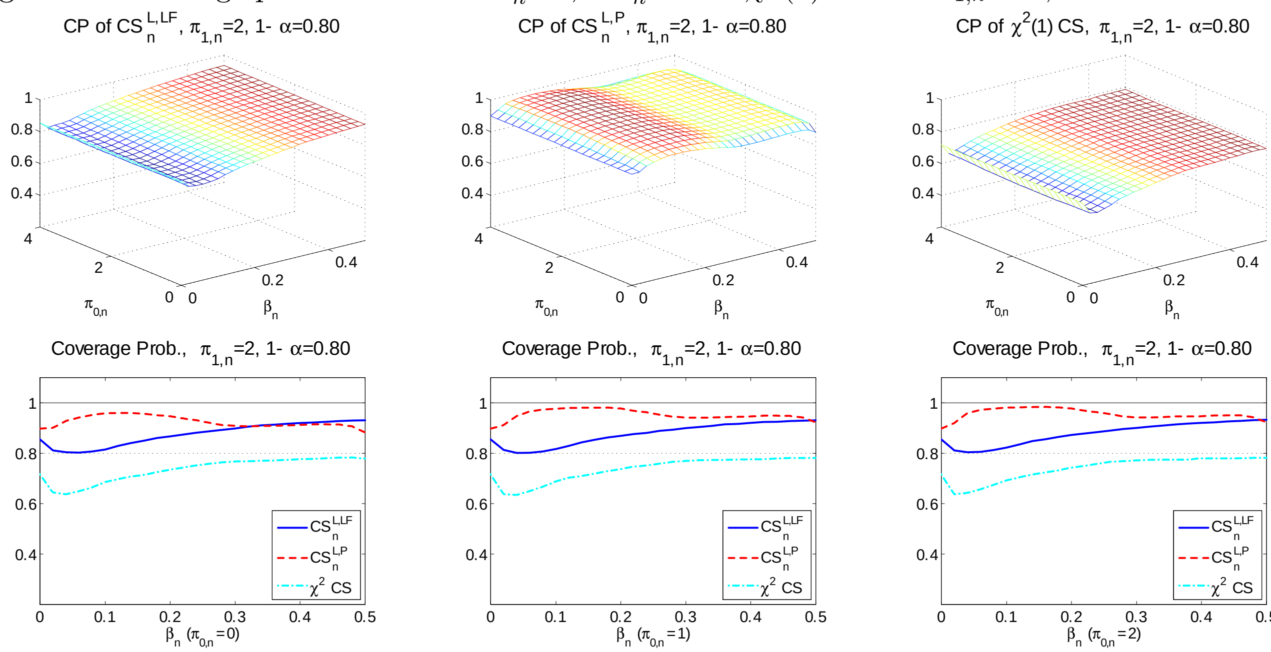
The first and second rows are respectively the simulated finite-sample densities and the asymptotic densities in Example 3.

Figure 10: Finite-sample and asymptotic densities of W_n for $H_0 : \pi_1 = \pi_{1,n}$, $\pi_{0,n} = \pi_{1,n} = 2$



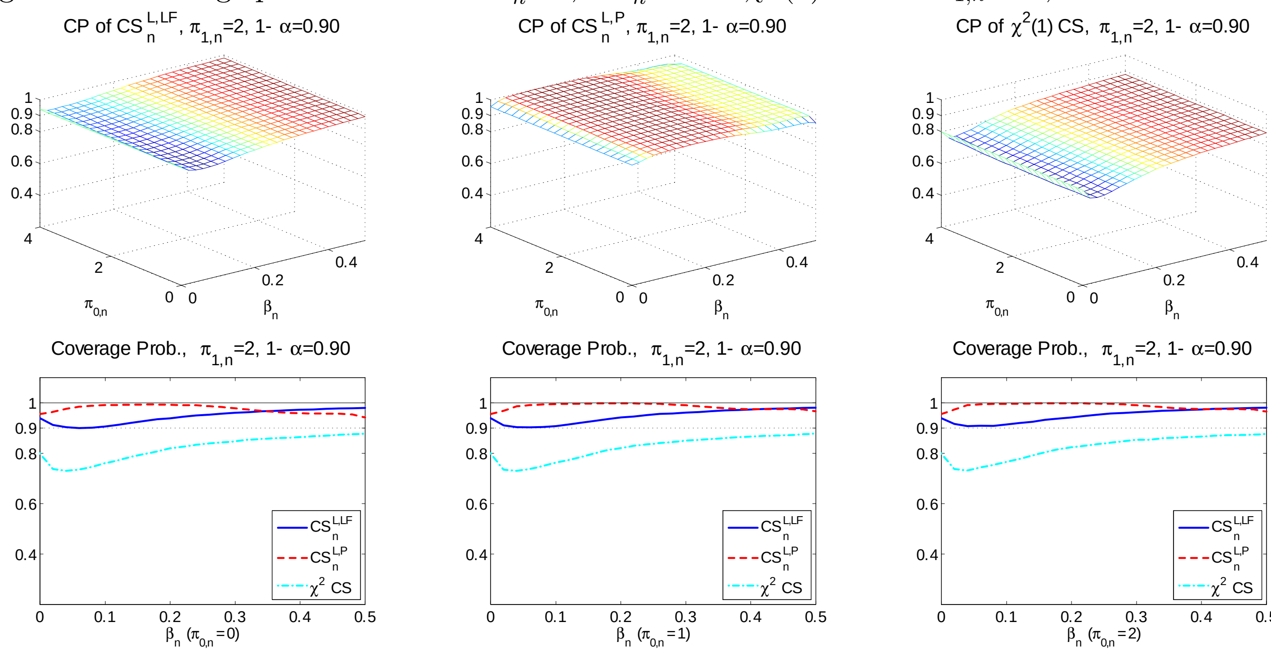
The first and second rows are respectively the simulated finite-sample densities and the asymptotic densities in Example 3.

Figure 11: Coverage probabilities of $CS_n^{L,LF}$, $CS_n^{L,P}$ and $\chi^2(1)$ CS for $\pi_{1,n} = 2$, $1 - \alpha = 0.8$



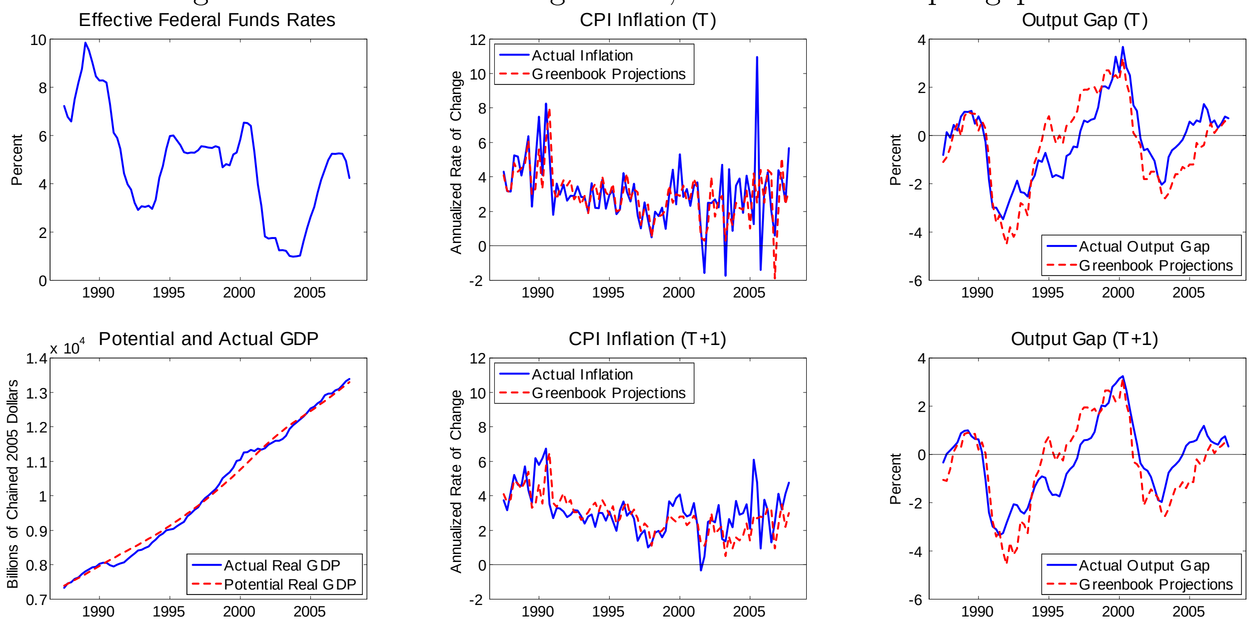
The first to third panels in the first row are respectively simulated coverage probabilities of the least-favorable confidence sets $CS_n^{L,LF}$, the projection-based confidence sets $CS_n^{L,P}$ and the standard confidence sets based on the $\chi^2(1)$ distribution of Example 4. The first to third panels in the second row are coverage probabilities of $CS_n^{L,LF}$, $CS_n^{L,P}$ and $\chi^2(1)$ CS with $\pi_{0,n} = 0$, 1, and 2.

Figure 12: Coverage probabilities of $CS_n^{L,LF}$, $CS_n^{L,P}$ and $\chi^2(1)$ CS for $\pi_{1,n} = 2$, $1 - \alpha = 0.9$



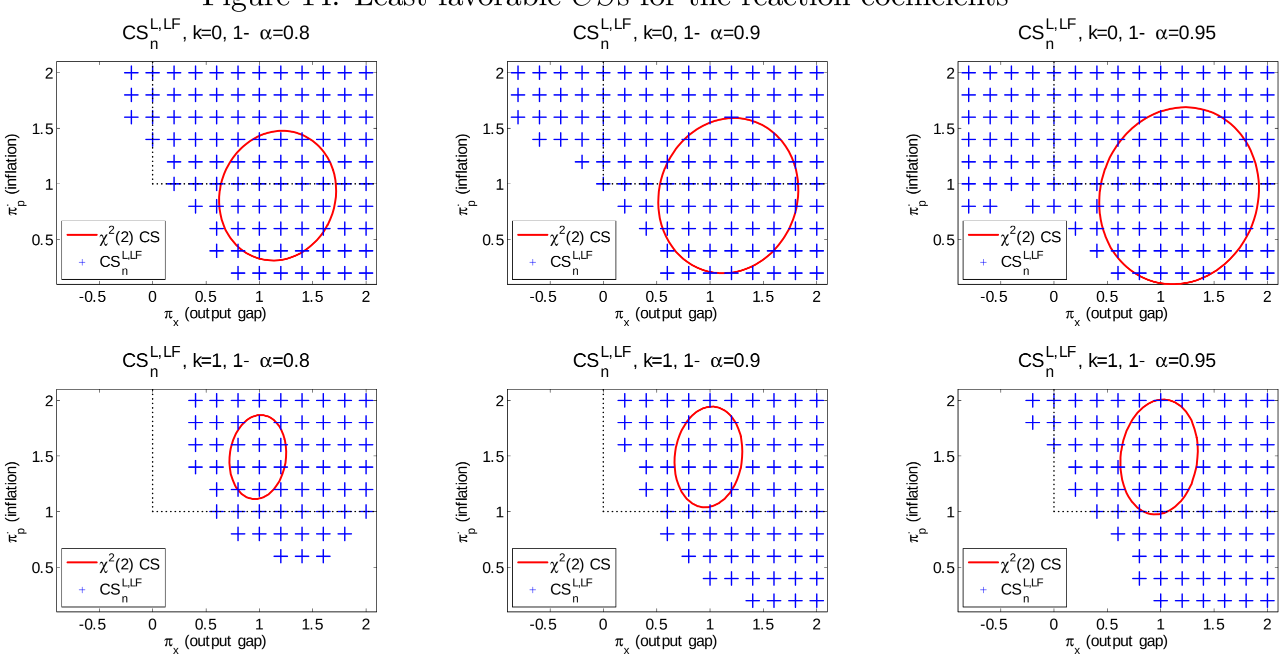
The first to third panels in the first row are respectively simulated coverage probabilities of the least-favorable confidence sets $CS_n^{L,LF}$, the projection-based confidence sets $CS_n^{L,P}$ and the standard confidence sets based on the $\chi^2(1)$ distribution of Example 4. The first to third panels in the second row are coverage probabilities of $CS_n^{L,LF}$, $CS_n^{L,P}$ and $\chi^2(1)$ CS with $\pi_{0,n} = 0$, 1, and 2.

Figure 13: Federal funds target rates, inflations and output gaps



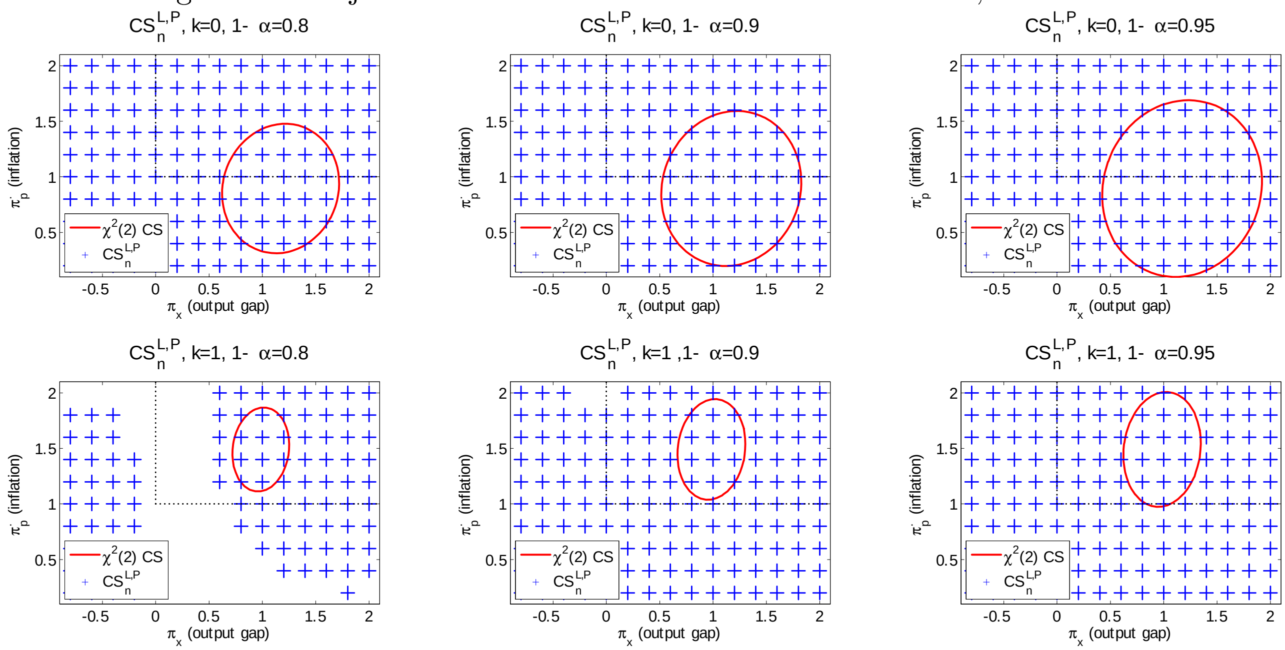
The effective federal funds target rates are the monthly averages of the last month in each quarter. The inflation rates, potential GDP and actual GDP are from the Federal Reserve Economic Data (FRED®) in Federal Reserve Bank of St. Louis. The Greenbook projections are from the Real-Time Data Research Center in Federal Reserve Bank of Philadelphia. The dates correspond to the publication dates of Greenbooks.

Figure 14: Least-favorable CSs for the reaction coefficients



The first and second rows are respectively for $k = 0$ and 1. The first to third panels are respectively for $1 - \alpha = 0.8$, 0.9 and 0.95. The dot line denotes the determinacy region $\mathcal{DR} = \{\pi_{\dot{p}} > 1, \pi_x > 0\}$.

Figure 15: Projection-based CSs for the reaction coefficients, $k = 1$



The first and second rows are respectively for $k = 0$ and 1. The first to third panels are respectively for $1 - \alpha = 0.8$, 0.9 and 0.95. The dot line denotes the determinacy region $\mathcal{DR} = \{\pi_{\dot{p}} > 1, \pi_x > 0\}$.