ESTIMATION AND INFERENCE UNDER WEAK IDENTIFICATION AND PERSISTENCE: AN APPLICATION TO FORECAST-BASED MONETARY POLICY REACTION FUNCTION

A Dissertation

by

JUI-CHUNG YANG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Ke-Li Xu
Co-Chair of Committee,	Qi Li
Committee Members,	Dennis W. Jansen
	Faming Liang
Head of Department,	Timothy Gronberg

August 2014

Major Subject: Economics

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, *i.e.*, the nominal interest rates are highly persistent. Using asymptotic theories for near unit root processes and novel drifting sequence approaches, we modify the method of Andrews and Cheng (2012, *Econometrica*) on inference under weak identification to accommodate the persistence issue. 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 and Wald statistics of a general class of models.

Despite the not-consistent-estimability when the smoothing coefficient is close to unity, the conservative confidence sets of weakly-identified parameters of interest can be obtained by inverting the t or the Wald tests. We show that the null-imposed least-favorable confidence sets will have correct asymptotic sizes while the projection-based and Bonferroni-based methods may lead to asymptotic over-coverage. An identification-category-selection procedure is proposed to select between the standard confidence set and the conservative one under weak identification. Our empirical application suggests that for the model in which the expected inflations and output gaps have a forecast horizon zero, 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. However, for the model with forecast horizon one, the possibility of indeterminacy may be ruled out.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

In a seminal paper, Clarida, Galí and Gertler (2000) proposed the monetary policy reaction function ($MPRF$) in the study of the implications of monetary policy for macroeconomic fluctuations. In $MPRF$, the nominal interest rate i_t is modeled as a weighted average of the interest rate in the previous period i_{t-1}, and the monetary authority's target rate i_t^*. The target rate i_t^* is assumed to follow a forward-looking Taylor monetary policy rule (Taylor, 1993; Clarida *et al.*, 2000), *i.e.*, i_t^* is a function of the expected annualized inflation $\mathrm{E}_t\dot{p}_{t,k}$ and the expected average output gap $\mathrm{E}_t x_{t,k}$ between periods t and $t+k$:

$$\begin{aligned} i_t &= \rho i_{t-1} + (1-\rho)\, i_t^* + \varepsilon_t \qquad (1.1) \\ &= \rho i_{t-1} + (1-\rho)\left(\pi_\alpha + \pi_{\dot{p}} \mathrm{E}_t \dot{p}_{t,k} + \pi_x \mathrm{E}_t x_{t,k}\right) + \varepsilon_t. \end{aligned}$$

$\mathrm{E}_t(\cdot)$ denotes the expectation of the monetary authority at time t, and k denotes the forecast horizon. $\rho \in [0,1)$ is known as the smoothing coefficient, and $\{\pi_{\dot{p}}, \pi_x\}$ are known as the reaction coefficients. In this paper we use the real-time data, *i.e.*, the historical *ex ante* forecasts ($\{\mathrm{E}_t\dot{p}_{t,k}, \mathrm{E}_t x_{t,k}\}$) for the inflations and output gaps, and the model is consequently called the forecast-based $MPRF$. We are especially interested in the problem if the reaction coefficient for inflation $\pi_{\dot{p}}$ is greater than one, and the coefficient for output gap π_x is greater than zero. 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]. Throughout this paper, the region $\mathcal{DR} = \{\pi_{\dot{p}} > 1, \pi_x > 0\}$ is called the determinacy region.

The objective of the present study is to revisit the empirical findings of Clarida *et al.* (2000) about the determinacy of $MPRF$ of U.S. with more recent real-time data, in light of recent concerns over the issue of the weak identification of parameters (Andrews and Cheng, 2012. 2013a, 2013b). Specifically, in this paper we are interested in the inference of the forecast-based $MPRF$ of U.S. when the smoothing coefficient ρ is close to unity, based on the nonlinear least squares (NLS) estimation. Lately close-to-one estimates for ρ had been found by Bunzel and Enders (2010), Nikolsko-Rzhevskyy (2011) and Nikolsko-Rzhevskyy and Papell (2012). 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 able to 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 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 with a near unit root, and the NLS estimator will have a nonstandard asymptotic distribution. 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

[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).

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 and the test statistics for a class of models in which weak identification occurs in part of the parameter space when there is a unit root or near unit root. The current study modifies the method of Andrews and Cheng (2012) on inference under weak and 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 are appropriately chosen according to the nonstandard convergence or 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 and Wald statistics.

Second, despite the not-consistent-estimability when $\rho \approx 1$, the conservative confidence sets (CS) of weakly-identified parameters of interest can be obtained by inverting the t or 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 (Dufour, 1997) and Bonferroni-based method may lead to asymptotic over-coverage. All three methods will give conservative CS which will be robust to the identification failure of the $MPRF$ when $\rho \approx 1$. As in Andrews and Cheng (2012), we also propose an identification-category-selection (ICS) procedure to select the appropriate confidence set between the standard and usually more informative CS, and the conservative CS under weak identification.

Third, we obtain the conservative confidence sets of the reaction coefficients $\{\pi_{\dot{p}}, \pi_x\}$ in U.S.'s forecast-based $MPRF$ with forecast horizons $k = 0$ and 1 for

1987:3–2007:4 with confidence coefficients $1-\alpha = 0.8$, 0.9 and 0.95. In the case $k = 0$, our ICS procedure selects the conservative CSs, which contain many values of $\{\pi_{\dot{p}}, \pi_x\}$ not in the determinacy region $\mathcal{DR} = \{\pi_{\dot{p}} > 1, \pi_x > 0\}$. For the case $k = 1$, however, our ICS procedure selects the standard CSs, which are contained in the determinacy region $\mathcal{DR} = \{\pi_{\dot{p}} > 1, \pi_x > 0\}$. Our empirical application suggests that for the case $k = 0$, the NLS estimates for $\{\pi_{\dot{p}}, \pi_x\}$ are not accurate sufficiently to rule out the possibility of indeterminacy. But in the case $k = 1$, the possibility of indeterminacy may be ruled out.

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 their seminal paper, Clarida, Galí and Gertler (2000) estimated the monetary policy reaction function of U.S. for the pre-Volcker (1960:1 – 1979:2) and Volcker-Greenspan periods (1979:3 – 1996:4)[2]. Since the expectations of the inflation and the output gap of the Federal Reserve ($\{\mathrm{E}_t\dot{p}_{t,k}, \mathrm{E}_t x_{t,k}\}$) were unobservable to the public, Clarida *et al.* (2000) replaced the *ex ante* expectations by the observable *ex post* realizations ($\{\dot{p}_{t,k}, x_{t,k}\}$).

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

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

[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.

reaction coefficients $\{\pi_{\dot{p}}, \pi_x\}$ for the pre-Volcker / Volcker-Greenspan periods were respectively not in and 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 $\{i_t, \dot{p}_{t,k}, x_{t,k}\}$ are merely weakly correlated to $\{\dot{p}_{t,k}, x_{t,k}\}$. Recently Inoue and Rossi (2011) and Mavroeidis (2010) reexamined 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 are 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 ($\{\mathrm{E}_t\dot{p}_{t,k}, \mathrm{E}_t x_{t,k}\}$) 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 real-time data of many macroeconomic variables have been open to the public (after a five-year declassification period) in the Federal Reserve Bank of Philadelphia[5]. For

[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 $0.14 - 0.18$.

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

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 [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 and 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 and 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, but π is identified if and only if $\beta \neq 0$ and

[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 $p = 1$). Among many, they used the Hodrick-Prescott (1997) filter in computing output gaps.

the estimator for π may weakly converge to a nondegenerate random variable when $\beta \approx 0$. The problem considered in this paper is plausibly similar to Andrews and Cheng (2012) if we reparameterize $\rho = 1 - \beta$ in the $MPRF$. Consider the following data generating process (DGP):

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

where y_t denotes the interest rate i_t, and X_t denotes a constant one, the expected inflation and the expected output gap $(1, \mathrm{E}_t \dot{p}_{t,k}, \mathrm{E}_t x_{t,k})^\top$. As in Andrews and Cheng (2012), π can be identified if and only if $\beta = 1 - \rho \neq 0$. However, when equation (1.2) contains a close-to-zero β (close-to-one ρ), $\{y_t\}$ will be highly persistent. In this case, 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 are made to the 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, Andrews 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, to accommodate the persistence of $\{y_t\}$ when $\beta \approx 0$, drifting sequences different from Andrews and Cheng (2012) are appropriately chosen according to the nonstandard convergence or divergence rates of NLS estimators when $\beta = 0$.

Specifically, three different asymptotic approaches are considered. In the first asymptotic approach, the 'distant-from-zero β_n' class, $\beta = \beta_n$ drifts to zero with a standardization factor n^{-h} with $h \in (0, 1/2]$, while $\pi = \pi_n$ is treated as a fixed parameter. In the second asymptotic approach, the 'close-to-zero β_n' class, $\beta = \beta_n$ drifts to zero with a standardization factor n^{-h}, and $\pi = \pi_n$ drifts to $\pm\infty$ with a standardization factor $n^{-1/2+h}$ with $h \in [1/2, 1)$. And in the third asymptotic approach, the 'local-to-zero β_n' class, $\beta = \beta_n$ drifts to zero with a standardization factor n^{-1}, and $\pi = \pi_n$ drifts to $\pm\infty$ with a standardization factor $n^{1/2}$. The local-to-zero β_n class is chosen according to the convergence and divergence rates of the NLS estimator when $\beta = 0$. The distant-from-zero β_n class and the close-to-zero β_n class bridge the local-to-zero β_n class and the ordinary case in which both β_n and π_n are fixed parameters. 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 not seem intuitive. However, rather than any arbitrary artificial choice, the drifting-to-infinity sequences are logical outcomes of 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 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 and Wald test statistics. Specifically, when β is merely close to zero or distant from zero, the t and the Wald statistics will be asymptotically Gaussian and χ^2 distributed. However, when β is local to zero, both the t and the Wald statistics will have nonstandard and non-pivotal asymptotic distributions. Our Monte Carlo simulation shows that 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.

The confidence sets (CS) for any linear functions of parameters are obtained by inverting the t or the Wald tests. When β is not local to zero, since the t and the Wald statistics will have standard Gaussian and χ^2 asymptotic distributions, the CS will also be standard. When β is local to zero, however, 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), the projection-based method (Dufour, 1997) and the Bonferroni-based method. 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 a subspace in the parameter space. The Bonferroni-based method relies on the Bonferroni inequality and obtains the CSs for parameters of interest and parameters not of interest simultaneously. Though all three methods are conservative, we show that the $NILF$ CS will have correct asymptotic sizes. The projection-based and Bonferroni-based methods may lead to asymptotically over-coverage. However, since the information from the estimates for all parameters of interests are used, under certain circumstances, it is possible to obtain a more informative CS than the $NILF$ one by the projection-based and Bonferroni-based methods. All three methods require the computation of the test statistics for as many values of parameters as possible. In practice, we propose the use of the grid method. As in Andrews and Cheng (2012), we also propose an identification-category-selection (ICS) procedure to select the appropriate CS between the standard CS and the conservative one under weak identification.

According to our asymptotic theory, we construct the CS with confidence coefficients $1-\alpha = 0.8$, 0.9 and 0.95 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 real-time data for expected inflations and the expected output gaps ($\{\mathrm{E}_t \dot{p}_{t,k}, \mathrm{E}_t x_{t,k}\}$) from the Federal Reserve Bank of Philadelphia. As in Nikolsko-Rzhevskyy (2011), we consider the case with forecasting horizons $k = 0$ or 1. In the case $k = 0$, our ICS procedure selects the conservative CSs, which contain many values of $\{\pi_{\dot{p}}, \pi_x\}$ not in the determinacy region $\mathcal{DR} = \{\pi_{\dot{p}} > 1, \pi_x > 0\}$. For the case $k = 1$, however, our ICS procedure selects the conventional CSs, which are contained in the determinacy region $\mathcal{DR} = \{\pi_{\dot{p}} > 1, \pi_x > 0\}$. Our empirical application suggests that for the case $k = 0$, the NLS estimates for $\{\pi_{\dot{p}}, \pi_x\}$ are not accurate sufficiently to rule out the possibility of indeterminacy. But in the case $k = 1$, the possibility of indeterminacy may be ruled out.

The remainder of the paper is organized as follows. Section 2 provides the asymptotic theory for the NLS estimator for models as equation (1.2) when $\beta \approx 0$. Section 3 establishes the limiting properties of the t and the Wald test statistics and discusses 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 are collected in Appendix.

2. ASSUMPTIONS AND ASYMPTOTIC THEORY

Consider the following data generating process (DGP) as equation (2.1):

$$y_t = (1-\beta_n)\, y_{t-1} + \beta_n X_t^{\top} \pi_n + \varepsilon_t, \quad t = 1, \ldots, n. \tag{2.1}$$

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 ($\mathrm{E}_t \dot{p}_{t,k}$), the expected output gap ($\mathrm{E}_t x_{t,k}$) and a constant one as in equation (1.1).

In the section, we consider the nonlinear least squares (NLS) estimator for $\theta_n = \{\beta_n, \pi_n\}$.

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 $n \in N$. θ_n is an element of the interior of a convex parameter space Θ^*, which is contained in $(0, 1] \times R^{d_\pi}$.*

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

Assumption 3 *$\{\varepsilon_t\}$ and $\{X_t \varepsilon_t\}$ are martingale difference sequences (MDS). ε_t is independent to (y_{t-1}, X_t) with $E(\varepsilon_t) = 0$, $E|\varepsilon_t|^2 < \infty$ and $\mathrm{var}\,(\varepsilon_t) = \sigma_\varepsilon^2 > 0$ for all $t = 1, \ldots, n$.*

For notational simplicity, let $\varphi_0 = \{\mu_X, \mathbf{M}_X, \sigma_\varepsilon^2\}$ denote the nuisance parameters, where $\varphi_0 \in \Phi \subset \mathrm{R}^{d_\pi} \times \mathrm{R}^{d_\pi \times d_\pi} \times (0, \infty)$. Also let $\gamma_n = \{\theta_n, \varphi_0\} \in \Gamma = \Theta^* \times \Phi$ denote all the parameters in the model, including the parameters of interest $\theta_n = \{\beta_n, \pi_n\}$ and the nuisance parameters $\varphi_0 = \{\mu_X, \mathbf{M}_X, \sigma_\varepsilon^2\}$. $\{\varepsilon_t\}$ is assumed to be serially uncorrelated. If $\{\varepsilon_t\}$ is serially correlated, then by $\operatorname{Cov}(y_t, \varepsilon_t) \neq 0$ and $\operatorname{Cov}(y_t, y_{t-1}) \neq 0$, $\operatorname{Cov}(y_{t-1}, \varepsilon_t)$ will not be zero, *i.e.*, y_{t-1} will be endogenous, and the NLS estimator for θ_n will be biased.

$\theta_n = \{\beta_n, \pi_n\}$ belongs to the 'true parameter space' Θ^*. For any 'optimization parameter space' $\Theta \subset \mathrm{R}^{d_\pi+1}$ containing Θ^* (*i.e.*, $\Theta^* \subset \Theta$), the NLS estimator $\widehat{\theta}_n = \left\{\widehat{\beta}_n, \widehat{\pi}_n\right\}$ is defined as the minimizer of the objective function $Q_n(\theta)$.

$$Q_n\left(\widehat{\theta}_n\right) = \min_{\theta\in\Theta} Q_n(\theta) = \min_{\theta\in\Theta} \frac{1}{2n} \sum_{t=1}^{n} \left[y_t - (1-\beta)\, y_{t-1} - \beta X_t^\top \pi\right]^2. \tag{2.2}$$

In practice, the optimization parameter space Θ can be selected as a large set to prevent the misspecification of the parameter space. When the optimization parameter space Θ is large enough to rule out the possible boundary issues, the nonlinear least squares (NLS) estimator $\widehat{\theta}_n = \left\{\widehat{\beta}_n, \widehat{\pi}_n\right\}$ can also be defined by the first order condition, *i.e.*,

$$\frac{1}{n}\sum_{t=1}^{n} \left(y_{t-1} - X_t^\top \widehat{\pi}_n\right)\left[y_t - \left(1-\widehat{\beta}_n\right) y_{t-1} - \widehat{\beta}_n X_t^\top \widehat{\pi}_n\right] = 0,$$
$$\frac{1}{n}\sum_{t=1}^{n} X_t \left[y_t - \left(1-\widehat{\beta}_n\right) y_{t-1} - \widehat{\beta}_n X_t^\top \widehat{\pi}_n\right] = 0.$$

In the following we discuss the estimation of θ_n when β_n is close to zero and not close to zero separately. When $\beta_n = \beta_0 > 0$ and $\pi_n = \pi_0$, *i.e.,* when θ_n is fixed at the constant vector $\theta_0 = \{\beta_0, \pi_0\} \in \Theta^*$, 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* $\theta_n = \theta_0 \in \Theta^*$, *i.e.,* $\beta_n = \beta_0$ *and* $\pi_n = \pi_0$ *for any* $n \in N$. *Then* $\widehat{\theta}_n \xrightarrow{p} \theta_n = \theta_0$, *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,*

$$\mathcal{V}_0(\gamma_n) = E\begin{bmatrix} \left(y_{t-1} - X_t^\top \pi_0\right)^2 & -\beta_0\left(y_{t-1} - X_t^\top \pi_0\right) X_t^\top \\ -\beta_0 X_t\left(y_{t-1} - X_t^\top \pi_0\right) & \beta_0^2 X_t X_t^\top \end{bmatrix}.$$

However, when $\beta = 0$, the NLS objective function $Q_n(\theta)$ 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 $\beta \approx 0$, 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* $n \in N$*. 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 three different asymptotic approaches, $\Gamma(1, b, \mathbf{c})$, $\Gamma(h, b, \mathbf{c})$ and $\Gamma(h, b)$, to mimic the finite sample behaviors of $\widehat{\theta}_n = \left\{\widehat{\beta}_n, \widehat{\pi}_n\right\}$. Through out this paper, the three asymptotic approaches $\Gamma(1, b, \mathbf{c})$, $\Gamma(h, b, \mathbf{c})$ and $\Gamma(h, b)$ are respectively known as the 'local-to-zero β_n', 'close-to-zero β_n' and 'distant-from-zero β_n' classes.

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

$$
\begin{aligned}
&\textit{Local-to-zero } \beta_n : \Gamma(1, b, \mathbf{c}) = \left\{\{\gamma_n\} \in \Gamma : \beta_n = \frac{b}{n}, \pi_n = n^{1/2}\mathbf{c}\right\}, \\
&\textit{Close-to-zero } \beta_n : \Gamma(h, b, \mathbf{c}) = \left\{\{\gamma_n\} \in \Gamma : \beta_n = \frac{b}{n^h}, \pi_n = n^{-1/2+h}\mathbf{c}, h \in [1/2, 1)\right\}, \\
&\textit{Distant-from-zero } \beta_n : \Gamma(h, b) = \left\{\{\gamma_n\} \in \Gamma : \beta_n = \frac{b}{n^h}, h \in (0, 1/2]\right\}.
\end{aligned}
$$

For the local-to-zero β_n class $\Gamma(1, b, \mathbf{c})$, β_n and π_n are assumed to be sequences respectively drifting to zero and $\pm\infty$ when $n \to \infty$. The standardization factors n^{-1} and $n^{1/2}$ are appropriately chosen to match the convergence or divergence rates of the NLS estimator when $\beta_n = 0$ (Lemma 1). The distant-from-zero β_n class $\Gamma(h, b)$ and the close-to-zero β_n class $\Gamma(h, b, \mathbf{c})$ bridge the local-to-zero β_n class $\Gamma(1, b, \mathbf{c})$ and the ordinary case, in which both β_n and π_n are fixed parameters ($\theta_n = \theta_0 \in \Theta^* \subset (0, 1] \times \mathrm{R}^{d_\pi}$).

Notice that in the local-to-zero β_n class $\Gamma(1, b, \mathbf{c})$, the drifting sequence for $\rho_n = 1 - \beta_n$ is exactly the frequently used local-to-unity asymptotic approach (Phillips, 1987; Stock, 1991) in the near unit root literature. And in the close-to-zero β_n class $\Gamma(h, b, \mathbf{c})$ and the distant-from-zero β_n class $\Gamma(h, b)$, the drifting sequence for $\rho_n = 1 - \beta_n$ is the neighborhood-of-unity approach (Giraitis and Phillips, 2006; Phillips and Magdalinos, 2007). For the drifting sequence for π_n, in the distant-from-zero β_n class $\Gamma(h, b)$ we do not make any special assumption about π_n and simply treat π_n as a fixed parameter. In the local-to-zero β_n class $\Gamma(1, b, \mathbf{c})$ and the close-to-zero β_n class $\Gamma(h, b, \mathbf{c})$, however, π_n follows divergent sequences drifting to infinity. To the best of our knowledge, divergent drifting sequences have never appeared in the literature and may seem not intuitive. Rather than any arbitrary artificial choice, the drifting-to-infinity sequences are logical outcomes of the convergence or divergence rates of the NLS estimators when $\beta_n = 0$ (Lemma 1). We will discuss the divergent drifting sequences in more details in subsection 2.3.

In the following two sections we establish the asymptotic results under these three different asymptotic approaches. Our method is a modification of Andrews and Cheng (2012) on weak and 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 $O_p(n^{-1})$ and $O_p\left(n^{1/2}\right)$. Due to the different convergence or 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 and Wald test statistics with quadratic approximations for the objective function similar to Andrews and Cheng (2012).

In contrast to Andrews 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 ($\beta_n = n^{-1}b$ or $n^{-h}b$, and $\pi_n = n^{1/2}\mathbf{c}$ or $n^{-1/2+h}\mathbf{c}$). The linear drifting sequences and the property of the exponential function $\lim_{n\to\infty}\left(1-n^{-1}b\right)^n = \exp\left(-b\right)$ 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 $\gamma_n \in \Gamma\left(1, b, \mathbf{c}\right)$, *i.e.*, $\beta_n = n^{-1}b$ and $\pi_n = n^{1/2}\mathbf{c}$. When $\gamma_n \in \Gamma\left(1, b, \mathbf{c}\right)$, as in Andrews and Cheng (2012), we consider a quadratic approximation for $Q_n\left(\beta, \pi\right)$ in β around $\beta = 0$.

$$Q_n\left(\beta, \pi\right) - Q_n\left(0, \pi\right) = \frac{\partial}{\partial\beta}Q_n\left(0, \pi\right)\cdot\beta + \frac{1}{2}\frac{\partial^2}{\partial\beta^2}Q_n\left(\beta^*, \pi\right)\cdot\beta^2, \tag{2.3}$$

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

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

Since $\partial^2 Q_n(\beta,\pi)/\partial\beta^2$ does not depend on β, $\partial^2 Q_n(\beta^*,\pi)/\partial\beta^2 = \partial^2 Q_n(0,\pi)/\partial\beta^2$. Therefore, equation (2.3) can be written as:

$$Q_n(\beta,\pi) - Q_n(0,\pi) = \frac{\partial}{\partial\beta} Q_n(0,\pi) \cdot \beta + \frac{1}{2}\frac{\partial^2}{\partial\beta^2} Q_n(0,\pi) \cdot \beta^2. \tag{2.4}$$

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

$$n^{-1/2}\pi \Rightarrow \kappa_\pi. \tag{2.5}$$

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

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

$\mathcal{G}(\kappa_\pi, b, \mathbf{c}; \varphi_0)$ and $\mathcal{H}(\kappa_\pi, b, \mathbf{c}; \varphi_0)$ are functionals of a Wiener process $\mathcal{W}_\varepsilon(r)$ and an Ornstein–Uhlenbeck process $\mathcal{J}_{b,\varepsilon}(r)$. The exact functional forms of $\mathcal{G}(\kappa_\pi, b, \mathbf{c}; \varphi_0)$ and $\mathcal{H}(\kappa_\pi, b, \mathbf{c}; \varphi_0)$ are in Appendix A.

According to equation (2.4) and Lemma 2, let $q(\lambda_\beta, \kappa_\pi, b, \mathbf{c}; \varphi_0)$ be the asymptotic approximation of $Q_n(\beta,\pi) - Q_n(0,\pi)$,

$$q(\lambda_\beta, \kappa_\pi, b, \mathbf{c}; \varphi_0) = \mathcal{G}(\kappa_\pi, b, \mathbf{c}; \varphi_0) \cdot \lambda_\beta + \frac{1}{2}\mathcal{H}(\kappa_\pi, b, \mathbf{c}; \varphi_0) \cdot \lambda_\beta^2. \tag{2.6}$$

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

$$q\left(\widehat{\lambda}_\beta(\kappa_\pi, b, \mathbf{c}; \varphi_0), \kappa_\pi, b, \mathbf{c}; \varphi_0\right) = \inf_{\lambda_\beta} q(\lambda_\beta, \kappa_\pi, b, \mathbf{c}; \varphi_0), \tag{2.7}$$

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

$$\begin{aligned} & q\left(\widehat{\lambda}_\beta(\widehat{\kappa}_\pi(b, \mathbf{c}; \varphi_0), b, \mathbf{c}; \varphi_0), \widehat{\kappa}_\pi(b, \mathbf{c}; \varphi_0), b, \mathbf{c}; \varphi_0\right) \\ & = \inf_{\kappa_\pi} q\left(\widehat{\lambda}_\beta(\kappa_\pi, b, \mathbf{c}; \varphi_0), \kappa_\pi, b, \mathbf{c}; \varphi_0\right). \end{aligned} \tag{2.8}$$

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

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

$\widehat{\kappa}_\pi(b, \mathbf{c}; \varphi_0)$ and $\widehat{\lambda}_\beta(\widehat{\kappa}_\pi(b, \mathbf{c}; \varphi_0), b, \mathbf{c}; \varphi_0)$ are also defined in details in Appendix A.

Remark 1 *1. In Theorem 2 we show that when $\gamma_n \in \Gamma(1, b, \mathbf{c})$, $\widehat{\beta}_n$ is super-consistent with a convergence rate n, and $\widehat{\pi}_n$ does not possess a limiting distribution but actually diverge as $n \to \infty$ with a divergence rate $\sqrt{n}$. The asymptotic distributions of $n\left(\widehat{\beta}_n - \beta_n\right)$ and $n^{-1/2}(\widehat{\pi}_n - \pi_n)$ 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* $\gamma_n \in \Gamma(1, b, \mathbf{c})$*, the* t *and 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* $Q_n(\beta, \pi)$ *when* $\beta = 0$*,* $Q_n(0, \pi)$ *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* $Q_n(\beta, \pi) - Q_n(0, \pi)$*, which has the quadratic expansion as in equation (2.4). 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, $n\left(\widehat{\beta}_n - \beta_n\right)$ and $n^{-1/2}(\widehat{\pi}_n - \pi_n)$ will have asymptotic distributions depending 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:

$$\widehat{\mu}_{X,n} = n^{-1} \sum_{t=1}^{n} X_t, \quad \widehat{\mathbf{M}}_{X,n} = \frac{1}{n} \sum_{t=1}^{n} X_t X_t^\top, \quad \widehat{\sigma}_n^2 = n^{-1} \sum_{t=1}^{n} \widehat{\varepsilon}_t^2, \quad \text{where} \tag{2.9}$$
$$\widehat{\varepsilon}_t = y_t - \left(1 - \widehat{\beta}_n\right) y_{t-1} - \widehat{\beta}_n X_t^\top \widehat{\pi}_n, \quad t = 1 \ldots, n.$$

Lemma 3 *Suppose that all conditions of Theorem 2 are satisfied. Then* $\widehat{\varphi}_n \xrightarrow{p} \varphi_0$*, i.e.,* $\widehat{\mu}_{X,n} \xrightarrow{p} \mu_X$*,* $\widehat{\mathbf{M}}_{X,n} \xrightarrow{p} \mathbf{M}_X$*, and* $\widehat{\sigma}_n^2 \xrightarrow{p} \sigma_\varepsilon^2$*.*

Proposition 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 $n\left(\widehat{\beta}_n - \beta_n\right)$ and $n^{-1/2}\left(\widehat{\pi}_n - \pi_n\right)$ by Monte Carlo simulation. We omit the formal proof for the asymptotic theory of $n\left(\widehat{\beta}_n - \beta_n\right)$ and $n^{-1/2}\left(\widehat{\pi}_n - \pi_n\right)$ when φ_0 is replaced by its estimate $\widehat{\varphi}_n$ 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.

Example 1 *Consider the following model as equation (2.10):*

$$y_t = \left(1 - \beta_n\right) y_{t-1} + \beta_n \left(\pi_{0,n} + \pi_{1,n} x_t\right) + \varepsilon_t, \quad t = 1, \ldots, n, \tag{2.10}$$

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 2.1 and 2.2 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*

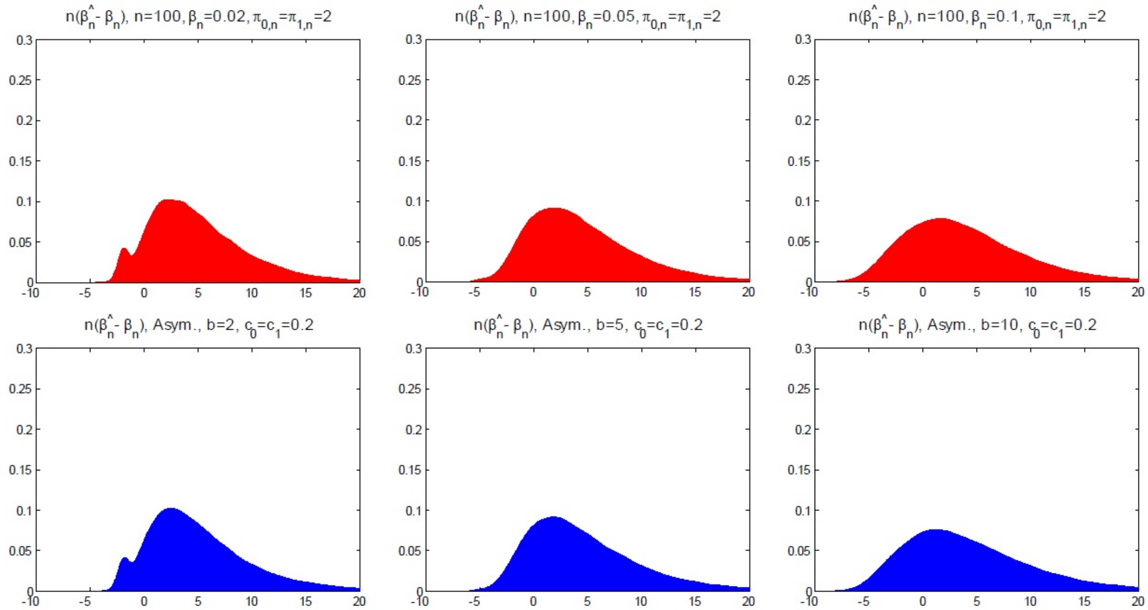

Figure 2.1: Finite-sample and asymptotic densities of $n(\widehat{\beta}_n - \beta_n)$, $\pi_{0,n} = \pi_{1,n} = 2$, Example 1
The first and rows are respectively the simulated finite-sample densities and the asymptotic densities of $n(\widehat{\beta}_n - \beta_n)$ with $\pi_{0,n} = \pi_{1,n} = 2$ in Example 1.

$\{b, c_0, c_1\}$. *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.*

For all results $50,000$ *simulation repetitions are used. 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* $\eta_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$*.*

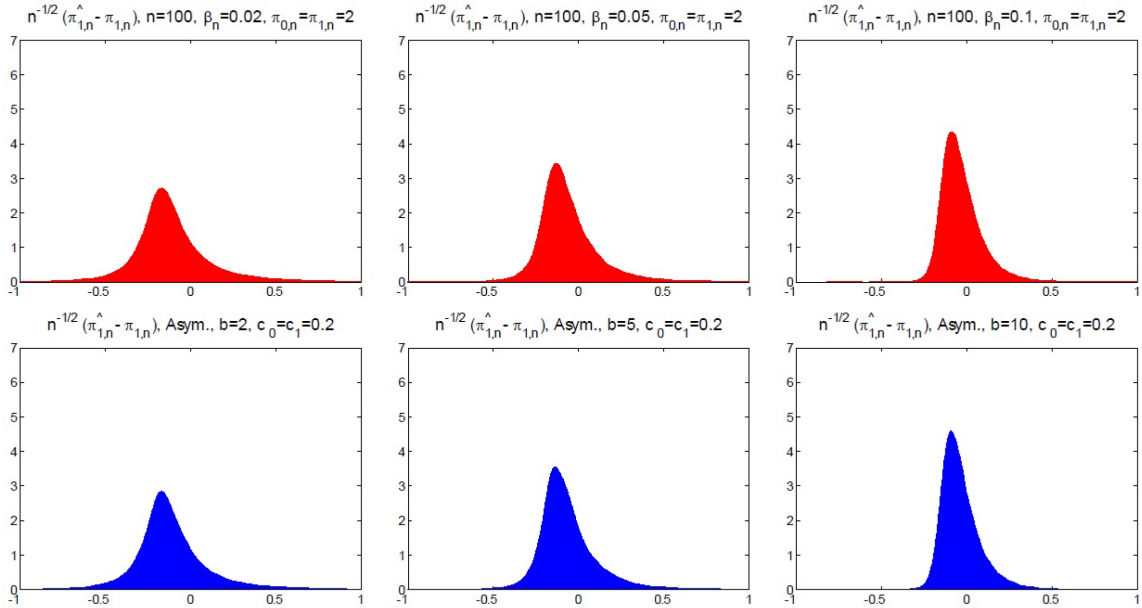

Figure 2.2: Finite-sample and asymptotic densities of $n^{-1/2}(\widehat{\pi}_{1,n}-\pi_{1,n})$, $\pi_{0,n} = \pi_{1,n} = 2$, Example 1
The first and rows are respectively the simulated finite-sample densities and the asymptotic densities of $n^{-1/2}(\widehat{\pi}_{1,n} - \pi_{1,n})$ with $\pi_{0,n} = \pi_{1,n} = 2$ in Example 1.

2.2 Estimation Results for Close-to-Zero β_n and 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 $\gamma_n \in \Gamma(h, b, \mathbf{c})$, the close-to-zero β_n class, and $\gamma_n \in \Gamma(h, b)$, the distant-from-zero β_n class. When $\gamma_n \in \Gamma(h, b, \mathbf{c})$, $\beta_n = n^{-h}b$ and $\pi_n = n^{-1/2+h}\mathbf{c}$ wit $h \in [1/2, 1)$. And when $\gamma_n \in \Gamma(h, b)$, π_n is fixed, and $\beta_n = n^{-h}b$ with $h \in (0, 1/2]$. When $\gamma_n \in \Gamma(h, b, \mathbf{c})$ or $\Gamma(h, b)$, we consider a quadratic approximation for $Q_n(\theta)$ around θ_n as Newey and McFadden (1994) and Andrews and Cheng (2012):

$$
\begin{aligned}
& Q_n(\theta) - Q_n(\theta_n) && (2.11) \\
&= D_\theta^\top Q_n(\theta_n)(\theta - \theta_n) + \frac{1}{2}(\theta - \theta_n)^\top D_{\theta\theta^\top} Q_n(\theta_n)(\theta - \theta_n) + R(\theta^*),
\end{aligned}
$$

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

$$D_\theta Q_n(\theta_n) = \begin{bmatrix} n^{-1}\sum_{t=1}^n \left(y_{t-1} - X_t^\top \pi_n\right)\varepsilon_t \\ -\beta_n n^{-1}\sum_{t=1}^n X_t \varepsilon_t \end{bmatrix},$$

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

Let

$$B(h) = \begin{bmatrix} n^{h/2} & \mathbf{0}_{1\times d_\pi} \\ \mathbf{0}_{d_\pi\times 1} & n^{-h}\mathrm{I}_{d_\pi} \end{bmatrix}. \tag{2.12}$$

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

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

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

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

Theorem 3 *Suppose that Assumptions 1, 2 and 3 hold and $\gamma_n \in \Gamma(h, b, \mathbf{c})$. 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* $\gamma_n \in \Gamma(h, b, \mathbf{c})$, $\widehat{\beta}_n - \beta_n = O_p\left(n^{-1/2-h/2}\right)$, *and* $\widehat{\pi}_n - \pi_n = O_p\left(n^{-1/2+h}\right)$. *Despite the non-standard convergence or divergence rates, the asymptotic distributions of* $n^{1/2+h/2}\left(\widehat{\beta}_n - \beta_n\right)$ *and* $n^{1/2-h}\left(\widehat{\pi}_n - \pi_n\right)$ *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* $\gamma_n \in \Gamma(h, b, \mathbf{c})$, *the asymptotic distributions of the* t *and Wald test statistics corresponding to* H_0 *will also be standard (Gaussian and* 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.*

3. Usually, the asymptotic distributions of the estimators will depend on true values of all parameters. One may expect $n^{1/2+h/2}\left(\widehat{\beta}_n - \beta_n\right)$ *and* $n^{1/2-h}\left(\widehat{\pi}_n - \pi_n\right)$ *to have limiting distributions depending on the values of both* β_n *and* π_n, *i.e., the values of both* b *and* $\mathbf{c}$. *However, in Lemma 4 we have shown that the limits of the first and second derivatives do not depend on the value of* $\mathbf{c}$, *so the limiting* *distributions of* $n^{1/2+h/2}\left(\widehat{\beta}_n - \beta_n\right)$ *and* $n^{1/2-h}\left(\widehat{\pi}_n - \pi_n\right)$ *do not depend on the value of* $\mathbf{c}$ *either. Intuitively, it is because when* $\beta_n = n^{-h}b$ *and* $\pi_n = n^{-1/2+h}\mathbf{c}$*, the value of* π_n *is too small, and the effect of* X_t *on* y_t *is negligible. In Lemma 7 (in Appendix C) we have shown that* y_t *can be written as:*

$$y_t = \mu_X^\top \pi_n + \eta_t + \beta_n \xi_t \pi_n,$$

where

$$\begin{aligned} \eta_t &= \sum_{i=0}^{\infty} \left(1-\beta_n\right)^i \varepsilon_{t-i} = O_p\left(n^{1/2+h}\right), \quad \text{and} \\ \xi_t &= \sum_{i=0}^{\infty} \left(1-\beta_n\right)^i \left(X_{t-i} - \mu_X\right) = O_p\left(n^{1/2+h}\right). \end{aligned}$$

Therefore, by $\beta_n = n^{-h}b$ *and* $\pi_n = n^{-1/2+h}\mathbf{c}$*,* $\mu_X^\top \pi_n = O\left(n^{-1/2+h}\right)$*, and* $\beta_n \xi_t \pi_n = O_p\left(n^h\right)$*. Thus,*

$$n^{-1/2-h} y_t = n^{-1/2-h} \eta_t + o_p\left(1\right).$$

That is, the value of $\mathbf{c}$ *does not affect* y_t*. Again, the standardization factors in the close-to-zero* β_n *class* $\Gamma\left(h, b, \mathbf{c}\right)$ *(*n^{-h} *for* β_n*, and* $n^{-1/2+h}$ *for* π_n*) are chosen to bridge the distant-from-zero* β_n *class* $\Gamma\left(h, b\right)$ *and the local-to-zero* β_n *class* $\Gamma\left(1, b, \mathbf{c}\right)$*, and the standardization factors in* $\Gamma\left(1, b, \mathbf{c}\right)$ *(*n^{-1} *for* β_n*, and* $n^{1/2}$ *for* π_n*) are chosen to match the convergence or divergence rates of the estimators when the true value of* β_n *equal to zero. So the not-depending-on-*$\mathbf{c}$ *asymptotic distributions of* $n^{1/2+h/2}\left(\widehat{\beta}_n - \beta_n\right)$ *and* $n^{1/2-h}\left(\widehat{\pi}_n - \pi_n\right)$ *are not because of any arbitrary choice of the standardization factors, but a logical outcome of the convergence or divergence rates of the estimators when* $\beta_n = 0$*.*

When $\gamma_n \in \Gamma(h, b)$, Lemma 5 and Theorem 4 show that the asymptotic distribution of $n^{1/2} B(h) \left(\widehat{\theta}_n - \theta_n \right)$ is the same as the case when $\gamma_n \in \Gamma(h, b, \mathbf{c})$.

Lemma 5 *Suppose that Assumptions 1, 2 and 3 hold, and $\gamma_n \in \Gamma(h, b)$. Then*

1. $n^{1/2} B^{-1}(h) D_\theta Q_n(\theta_n) \Rightarrow \mathcal{G}^*(b; \varphi_0) \sim \mathcal{N}\left(\mathbf{0}_{(d_\pi+1)\times 1}, \sigma_\varepsilon^2 \mathcal{V}_h(b; \varphi_0) \right)$.
2. $B^{-1}(h) D_{\theta\theta^\top} Q_n(\theta_n) B_2^{-1}(h) \xrightarrow{p} \mathcal{V}_h(b; \varphi_0)$.

Theorem 4 *Suppose that Assumptions 1, 2 and 3 hold and $\gamma_n \in \Gamma(h, b)$. Then*

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

2.3 Sequences Drifting to Infinity

In this paper we use sequences drifting to $\pm\infty$ to mimic the true value of π. Specifically, when $\gamma_n \in \Gamma(1, b, \mathbf{c})$, $\pi_n = n^{1/2}\mathbf{c}$; when $\gamma_n \in \Gamma(h, b, \mathbf{c})$, $\pi_n = n^{-1/2+h}\mathbf{c}$. To the best of our knowledge, divergent drifting sequences for parameter values have never appeared in the literature. For example, in their study of weak instruments, Staiger and Stock (1997) assumed the parameter of interest to be fixed while the correlation between the endogenous variable and the instrument is drifting to zero.

In this paper we do not assume π to be fixed for a twofold reason. First, rather than any arbitrary artificial choice, the drifting-to-infinity sequences are logical outcomes of the convergence or 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 drifting sequences in the local-to-zero β_n class $\Gamma(1, b, \mathbf{c})$ ($\beta_n = n^{-1}b$ and $\pi_n = n^{1/2}\mathbf{c}$) are chosen to match the convergence or divergence rates in the benchmark scenario ($\beta_n = 0$). And the sequences in other two classes are chosen to bridge the ordinary case $\{\theta_n = \theta_0 \in \Theta^*\}$ and the class $\Gamma(1, b, \mathbf{c})$. If the true value of π is treated as a fixed value while β is assumed drifting to zero, the divergence of $\widehat{\pi}_n$ while $\beta_n = 0$ will be disregarded, and the convergence or divergence rates in the benchmark scenario ($\beta_n = 0$) are not matched.

Second, the drifting-to-infinity π_n, together with the drifting-to-zero β_n, gives the desired smooth transition in the asymptotic approximation in mimicking the finite-sample behavior. Our Monte Carlo simulation in Example 1 shows that our asymptotic approximations fit the finite-sample densities very well. As a contrast, to treat the true value of π as a fixed value does not give valid asymptotic approximations. It can be easily shown that if the true value of π_n is assumed to be fixed while β_n is approximated by a local-to-zero sequence ($\beta_n = b/n$), then

$$\begin{bmatrix} n\left(\widehat{\beta}_n - \beta_n\right) \\ n^{-1/2}\left(\widehat{\pi}_n - \pi_n\right) \end{bmatrix} \Rightarrow \widehat{\tau}(b, \mathbf{0}; \varphi_0) = \begin{bmatrix} \widehat{\lambda}_\beta\left(\widehat{\kappa}_\pi(b, \mathbf{0}; \varphi_0), b, \mathbf{0}; \varphi_0\right) - b \\ \widehat{\kappa}_\pi(b, \mathbf{0}; \varphi_0) \end{bmatrix}. \qquad (2.13)$$

where $\widehat{\lambda}_\beta$ and $\widehat{\kappa}_\pi$ are defined in Theorem 2. In Example 2 we show that the asymptotic approximations according to equation (2.13) do not fit the finite-sample densities.

Example 2 *Again, consider the following model as equation (2.10):*

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

where $x_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$*,* $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$*, and* $n = 100$*. Again, we consider* $\beta_n \in$

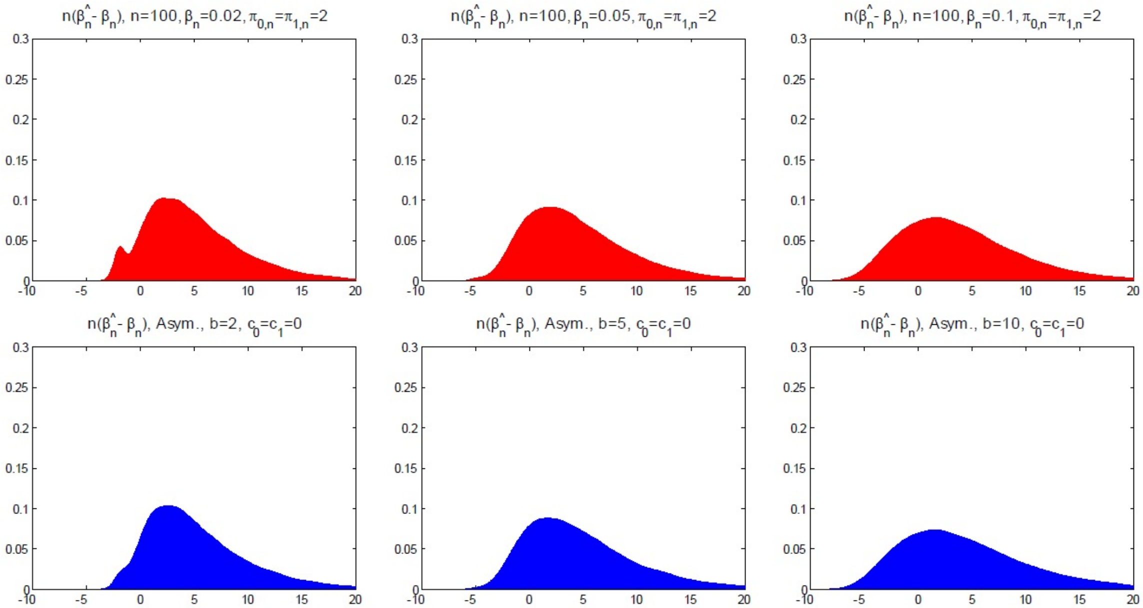

Figure 2.3: Finite-sample and asymptotic densities of $n(\widehat{\beta}_n - \beta_n)$, $\pi_{0,n} = \pi_{1,n} = 2$, Example 2
The first and rows are respectively the simulated finite-sample densities and the asymptotic densities of $n(\widehat{\beta}_n - \beta_n)$ with $\pi_{0,n} = \pi_{1,n} = 2$ in Example 2.

$\{0.02, 0.05, 0.1\}$ *and* $\pi_{0,n} = \pi_{1,n} = 2$.

According to equation (2.13), Figures 2.3 and 2.4 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 equation (2.13) do not fit the finite-sample densities.*

However, while β_n is approximated by a close-to-zero sequence ($\beta_n = n^{-h}b$), if the true value of π_n is assumed to be fixed, then the the asymptotic distributions of $n^{1/2+h/2}\left(\widehat{\beta}_n - \beta_n\right)$ and $n^{1/2-h}\left(\widehat{\pi}_n - \pi_n\right)$ remain the same as Theorem 3. Intuitively,

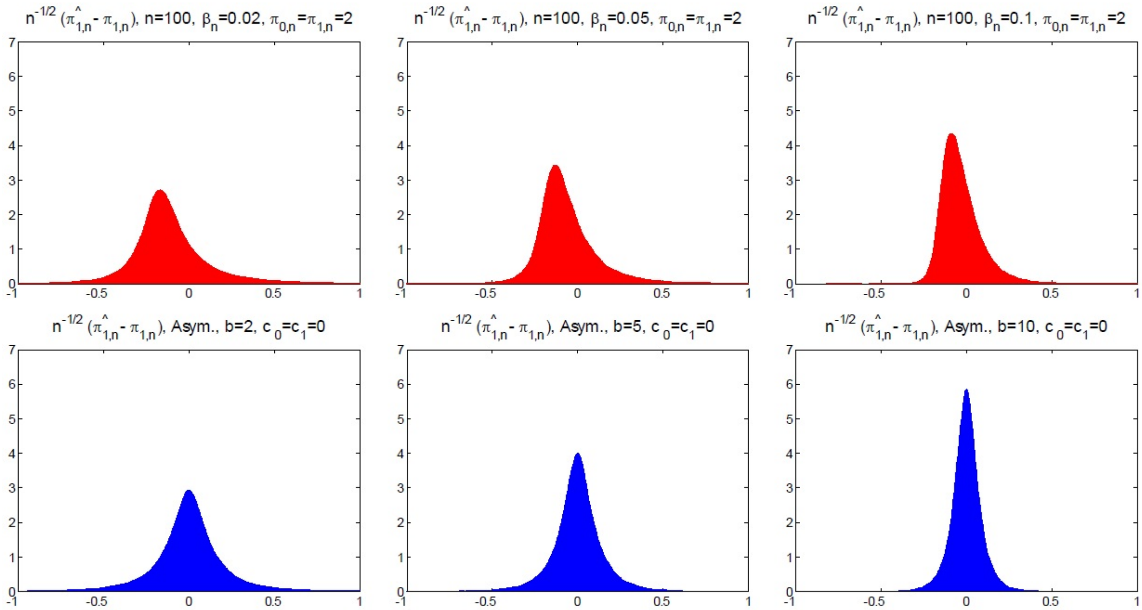

Figure 2.4: Finite-sample and asymptotic densities of $n^{-1/2}(\widehat{\pi}_{1,n}-\pi_{1,n})$, $\pi_{0,n}=\pi_{1,n}=2$, Example 2
The first and rows are respectively the simulated finite-sample densities and the asymptotic densities of $n^{-1/2}(\widehat{\pi}_{1,n}-\pi_{1,n})$ with $\pi_{0,n}=\pi_{1,n}=2$ in Example 2.

if π_n is assumed to be fixed, then by Lemma 7 (in Appendix C),

$$\eta_t = \sum_{i=0}^{\infty} (1-\beta_n)^i \varepsilon_{t-i} = O_p\left(n^{1/2+h}\right), \quad \text{and}$$
$$\xi_t = \sum_{i=0}^{\infty} (1-\beta_n)^i (X_{t-i}-\mu_X) = O_p\left(n^{1/2+h}\right),$$

and $\beta_n = n^{-h}b$, $n^{-1/2-h}y_t$ can be written as:

$$\begin{aligned} n^{-1/2-h}y_t &= n^{-1/2-h}\mu_X^\top \pi_n + n^{-1/2-h}\eta_t + n^{-1/2-h}\beta_n \xi_t \pi_n \\ &= O_p\left(n^{-1/2-h}\right) + n^{-1/2-h}\eta_t + O_p\left(n^{-h}\right) = n^{-1/2-h}\eta_t + o_p(1). \end{aligned}$$

Therefore, the value of $\mathbf{c}$ does not affect y_t, and the results of Theorem 3 remain.

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 $O_p\left(n^{-1}\right)$ and $O_p\left(n^{1/2}\right)$. Due to the difference in the convergence rates of estimators, we consider $\Gamma\left(1, b, \mathbf{c}\right)$, in which $n\beta_n = b$ and $n^{-1/2}\pi_n = \mathbf{c}$ to match the convergence or divergence rates, and use the other two classes to bridge $\{\theta_n = \theta_0 \in \Theta^*\}$ and $\Gamma\left(1, b, \mathbf{c}\right)$. In Theorems 2 and 3 we have already shown the necessity of the drifting-to-infinity assumption for inference about π_n.

In the current study, if we still use the same drifting sequences considered in Andrews and Cheng (2012) in their weak-identification scenario, it reduces to the case when $\gamma_n \in \Gamma\left(h, b, \mathbf{c}\right)$ with $h = 1/2$, in which β_n is a sequence drifting to zero with a standardization factor $n^{-1/2}$ $\left(n^{1/2}\beta_n = b\right)$ and π_n is a constant vector $\left(\pi_n = \mathbf{c}\right)$. We have already shown (in Theorem 3) that when $\gamma_n \in \Gamma\left(h, b, \mathbf{c}\right)$, the NLS estimator $\widehat{\pi}_n$ is asymptotically Gaussian distributed when $b \neq 0$, and is unidentifiable when $b = 0$ since $\text{Avar}\left(\widehat{\pi}_n\right) = \sigma_\varepsilon^2 b^{-2}\mathbf{M}_X^{-1} \to \infty$ when $b \to 0$. The desired smooth transition of the asymptotic approximation will be missing, due to the insufficient standardization factors not matching the convergence or 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 and the Wald test statistics, and discuss 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 $\mathbf{R} \in \mathrm{R}^{d_r \times (d_\pi + 1)}$, $\upsilon \in \mathrm{R}^{d_r}$ where $d_r \leq d_\pi + 1$, and $\mathrm{Rank}\,(\mathbf{R}) = d_r$.

3.1 t and Wald Test Statistics

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

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

$$W_n(\upsilon) = n\left[\mathbf{R}\widehat{\theta}_n - \upsilon\right]^\top \left[\widehat{\sigma}_n^2 \mathbf{R}\widehat{\mathbf{V}}_n^{-1}\mathbf{R}^\top\right]^{-1}\left[\mathbf{R}\widehat{\theta}_n - \upsilon\right], \tag{3.3}$$

where, by equation (2.9), $\widehat{\sigma}_n^2 = n^{-1}\sum_{t=1}^n \widehat{\varepsilon}_t^2$, and $\widehat{\mathbf{V}}_n$ is defined as:

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

Theorem 5 provides the asymptotic properties of $T_n(\upsilon_n)$ and $W_n(\upsilon_n)$, the t and Wald test statistics under the null $H_0 : \mathbf{R}\theta_n = \upsilon_n$, where υ_n denotes the true value of $\mathbf{R}\theta_n$.

Theorem 5 *Suppose that Assumptions 1, 2 and 3 hold.*

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

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

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

where $\widehat{\tau}(b, \mathbf{c}; \varphi_0)$ *is defined in Theorem 2.*

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

4. *When* $\gamma_n \in \Gamma(h, b)$, *i.e.,* $\beta_n = b/n^h$ *with* $0 < b < \infty$ *and* $h \in (0, 1/2]$, $T_n(\upsilon_n) \overset{A}{\sim} \mathcal{N}(0,1)$, *and* $W_n(\upsilon_n) \overset{A}{\sim} \chi^2(d_r)$.

$\mathcal{V}_1(b, \mathbf{c}; \varphi_0)$ is defined in details in Appendix A

Remark 3 *1. In Theorem 5 we obtain the asymptotic distribution of the t and the Wald statistics for all four cases we consider. When* $\theta_n = \theta_0 \in \Theta^*$, $\gamma_n \in \Gamma(h, b)$ *or* $\gamma_n \in \Gamma(h, b, \mathbf{c})$, $T_n(\upsilon_n)$ *and* $W_n(\upsilon_n)$ *have the standard and pivotal asymptotic Gaussian and* χ^2 *distributions. However, when* $\gamma_n \in \Gamma(1, b, \mathbf{c})$, *the asymptotic distribution of the* $T_n(\upsilon_n)$ *and* $W_n(\upsilon_n)$ *will depend on* $\widehat{\tau}(b, \mathbf{c}; \varphi_0)$ *and* $\mathcal{V}(b, \mathbf{c}; \varphi_0)$, *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* $\{b, \mathbf{c}\}$.

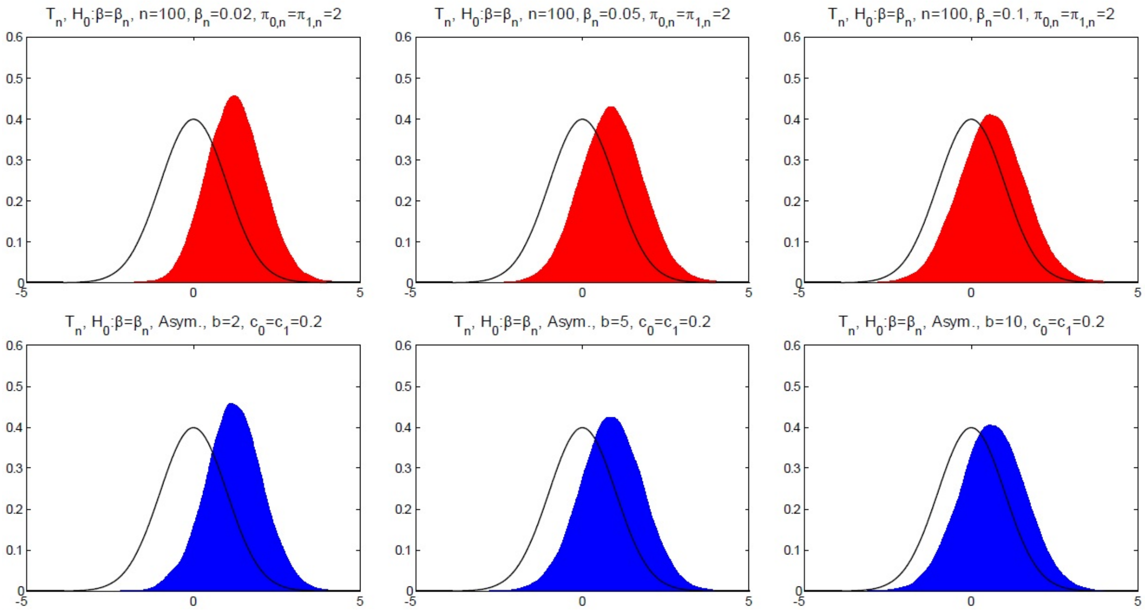

Figure 3.1: Finite-sample and asymptotic densities of T_n for $H_0 : \beta = \beta_n$, $\pi_{0,n} = \pi_{1,n} = 2$, Example 3
The first and second rows are respectively the simulated finite-sample densities and the asymptotic densities of T_n for $H_0 : \beta = \beta_n$ with $\pi_{0,n} = \pi_{1,n} = 2$ in Example 3.

2. *Similar to Mikusheva (2012), in this paper we only consider linear null hypotheses, e.g., $H_0 : \mathbf{R}\theta_n = \upsilon$. For the nonlinear null hypotheses, e.g., $H_0 : r(\theta_n) = \upsilon$ with a differentiable function $r : R^{(d_\pi+1)} \to R^{d_r}$, econometricians usually use the delta method to approximate the asymptotic variance of $r(\theta_n)$ by $\widehat{\sigma}_n^2 R^\top\left(\widehat{\theta}_n\right)\widehat{\mathbf{V}}_n^{-1}R\left(\widehat{\theta}_n\right)$, where $R(\theta) = D_\theta r(\theta)$ is the derivative of $r(\theta)$. When $\widehat{\theta}_n$ is a consistent estimator for θ_n, by the continuous mapping theorem, $R\left(\widehat{\theta}_n\right) \xrightarrow{p} R(\theta_n)$. For the problem we consider, however, we have shown that when $\gamma_n \in \Gamma(1, b, \mathbf{c})$ or $\gamma_n \in \Gamma(h, b, \mathbf{c})$ with $h \geq 1/2$, $\widehat{\pi}_n$ 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*

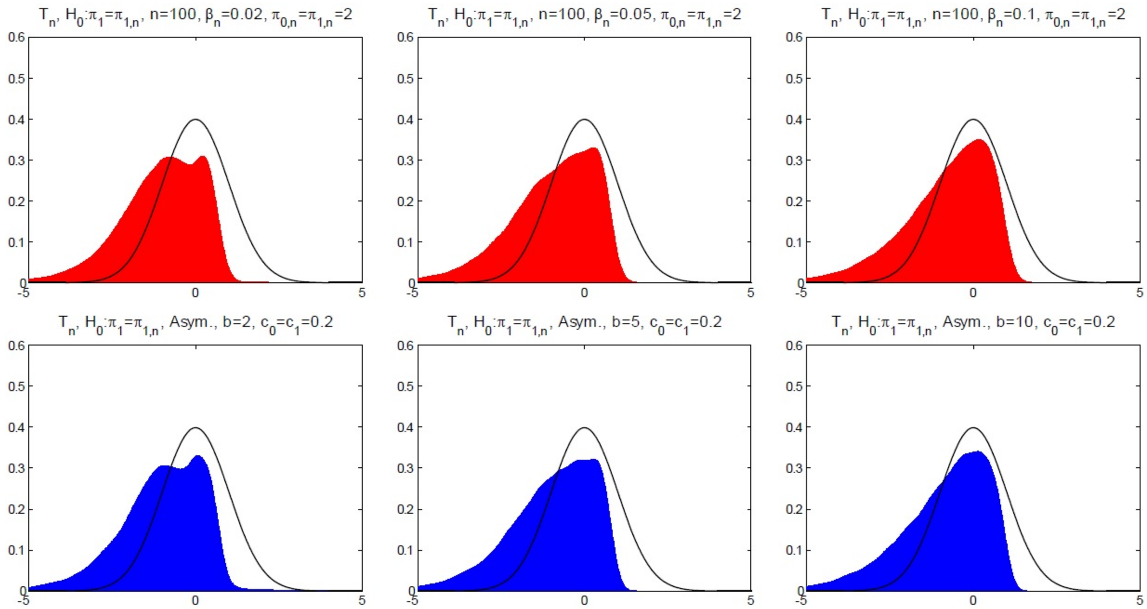

Figure 3.2: Finite-sample and asymptotic densities of T_n for $H_0 : \pi_1 = \pi_{1,n}$, $\pi_{0,n} = \pi_{1,n} = 2$, Example 3
The first and second rows are respectively the simulated finite-sample densities and the asymptotic densities of T_n for $H_0 : \beta = \beta_n$ with $\pi_{0,n} = \pi_{1,n} = 2$ in Example 3.

and Ham, 2013).

Again, when $\gamma_n \in \Gamma(1, b, \mathbf{c})$, the asymptotic distributions of T_n and 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 and 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 (2.10).*

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

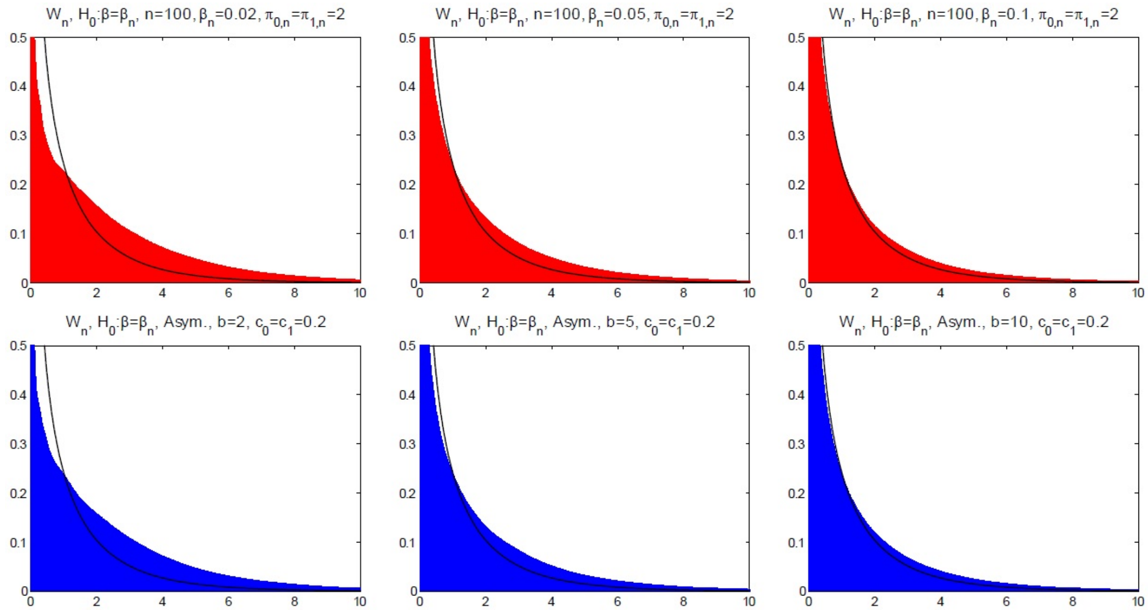

Figure 3.3: Finite-sample and asymptotic densities of W_n for $H_0 : \beta = \beta_n$, $\pi_{0,n} = \pi_{1,n} = 2$, Example 3
The first and second rows are respectively the simulated finite-sample densities and the asymptotic densities of W_n for $H_0 : \beta = \beta_n$ with $\pi_{0,n} = \pi_{1,n} = 2$ in Example 3.

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 *and* W_n *denote the t and 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 5, Figures 3.1 – 3.4 provide the simulated finite-sample and asymptotic densities of T_n *and* 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 5 fit the finite-sample densities very well.*

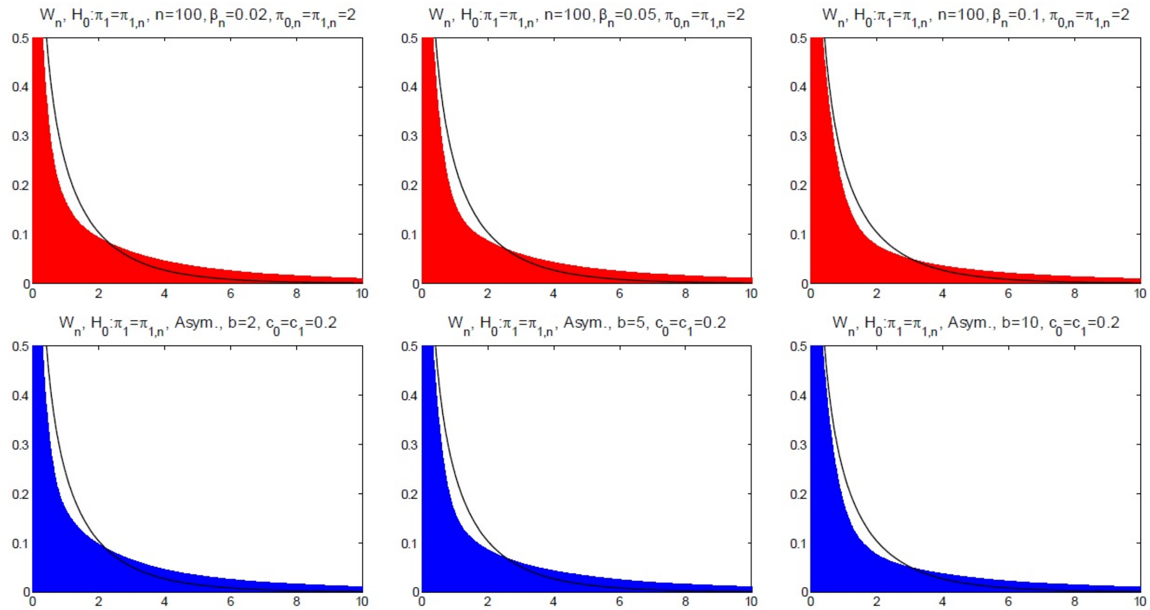

Figure 3.4: Finite-sample and asymptotic densities of W_n for $H_0 : \pi_1 = \pi_{1,n}$, $\pi_{0,n} = \pi_{1,n} = 2$, Example 3
The first and second rows are respectively the simulated finite-sample densities and the asymptotic densities of W_n for $H_0 : \beta = \beta_n$ with $\pi_{0,n} = \pi_{1,n} = 2$ in Example 3.

3.2 Robust Confidence Sets

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

The confidence sets when β_n is local-to-zero and not-local-to-zero are discussed separately. Let CS_n^L denotes the CS of $\mathbf{R}\theta_n$ when $\gamma_n \in \Gamma(1, b, \mathbf{c})$, and CS_n^D denotes the CS of $\mathbf{R}\theta_n$ when $\theta_n = \theta_0 \in \Theta^*$, $\gamma_n \in \Gamma(h, b)$ or $\gamma_n \in \Gamma(h, b, \mathbf{c})$, where L and D respectively represent 'local-to-zero β_n' and 'distant-from-zero β_n'.

The construction of CS_n^D of $\mathbf{R}\theta_n$, the confidence set when β_n is not-local-to-zero, is standard and simple. In Theorem 5 we have already show that when $\theta_n = \theta_0 \in$

Θ^*, $\gamma_n \in \Gamma(h, b)$ or $\gamma_n \in \Gamma(h, b, \mathbf{c})$, the Wald statistics is pivotally asymptotically $\chi^2(d_r)$-distributed, where $d_r = \text{Rank}(\mathbf{R})$. Let $\chi^2_{d_r, 1-\alpha}$ be the $(1-\alpha)$-quantile of $\chi^2(d_r)$. CS_n^D is simply the standard confidence set based on the asymptotic $\chi^2(d_r)$ distribution.

$$CS_n^D(\mathbf{R}\theta_n; 1-\alpha) = \left\{\upsilon : W_n(\upsilon) \le \chi^2_{d_r, 1-\alpha}\right\}. \tag{3.5}$$

Alternatively, since it is equivalent to consider if υ is in the $(1-\alpha)$-confidence set of $\mathbf{R}\theta_n$, or if the null hypothesis $H_0 : \mathbf{R}\theta_n = \upsilon$ can be accepted under the significant level $1-\alpha$, we can also interpret the construction of CS_n^D as inverting the Wald test. The steps to obtain CS_n^D can be written as:

1. For $H_0 : \mathbf{R}\theta_n = \upsilon$, obtain $\chi^2_{d_r, 1-\alpha}$, the $(1-\alpha)$-quantile of $\chi^2(d_r)$.
2. If $W_n(\upsilon) \le \chi^2_{d_r, 1-\alpha}$, then $\upsilon \in CS_n^D(\mathbf{R}\theta_n; 1-\alpha)$. If $W_n(\upsilon) > \chi^2_{d_r, 1-\alpha}$, then $\upsilon \notin CS_n^D(\mathbf{R}\theta_n; 1-\alpha)$.
3. Go back to step 1 and try another υ.

Since the asymptotic distribution of the Wald statistics $W_n(\upsilon)$ under the null $H_0 : \mathbf{R}\theta_n = \upsilon$ is standard and pivotal, for different υ in the null hypothesis the critical value $\chi^2_{d_r, 1-\alpha}$ remains the same.

For CS_n^L of $\mathbf{R}\theta_n$, the confidence set when β_n is local-to-zero, however, Theorem 5 shows that when $\gamma_n \in \Gamma(1, b, \mathbf{c})$, the Wald statistics has a nonstandard and non-pivotal asymptotic distribution $\mathcal{W}(b, \mathbf{c}; \varphi_0)$, which depends on the values of the localization parameters $\{b, \mathbf{c}\}$. Therefore, there will be a different critical value for the Wald test with every different $\{b, \mathbf{c}\}$. Since when $\gamma_n \in \Gamma(1, b, \mathbf{c})$, $\upsilon = \mathbf{R}\theta_n = \mathbf{R}\left[\beta_n, \pi_n^\top\right]^\top = \mathbf{R}\left[n^{-1}b, n^{1/2}\mathbf{c}^\top\right]^\top$, *i.e.*, the value of $\{b, \mathbf{c}\}$ depends on the value of υ, in this paper we impose the value of $\{b, \mathbf{c}\}$ implied by the null hypothesis into the asymptotic distribution of the Wald statistics $\mathcal{W}(b, \mathbf{c}; \varphi_0)$. Similar to equation (5.2) in Andrews and Cheng (2012), we define the 'null-imposition set' $\mathcal{H}(\mathbf{R}, \upsilon)$ for the localization parameters $\{b, \mathbf{c}\}$. $\mathcal{H}(\mathbf{R}, \upsilon)$ contains all possible values of the localization parameters $\{b, \mathbf{c}\}$ under the null $\mathbf{R}\theta_n = \upsilon$.

$$\mathcal{H}(\mathbf{R}, \upsilon) = \left\{ b, \mathbf{c} : \mathbf{R}\left[n^{-1}b, n^{1/2}\mathbf{c}^\top\right]^\top = \upsilon, \quad \left\{n^{-1}b, n^{1/2}\mathbf{c}^\top\right\} \in \Theta \right\}. \tag{3.6}$$

First we consider a simple case. Suppose that the null-imposition set $\mathcal{H}(\mathbf{R}, \upsilon)$ is a singleton for every υ. For example, suppose that $\mathbf{R} = \mathrm{I}_{d_\pi+1}$ and $\mathbf{R}\theta_n = \theta_n$, *i.e.*, we are interested in the confidence set of θ_n, then, by $\theta_n = \{\beta_n, \pi_n\} = \left\{n^{-1}b, n^{1/2}\mathbf{c}\right\}$, for any given null hypothesis $H_0 : \mathbf{R}\theta_n = \theta_n = \upsilon$, the values of the localization parameters are available under the null hypothesis. Let $\xi_{1-\alpha}(\mathcal{W}(b, \mathbf{c}; \varphi_0))$ be the $(1-\alpha)$-quantile of $\mathcal{W}(b, \mathbf{c}; \varphi_0)$, then under $\mathbf{R}\left[n^{-1}b_\upsilon, n^{1/2}\mathbf{c}_\upsilon^\top\right]^\top = \upsilon$,

$$CS_n^L(\mathbf{R}\theta_n; 1-\alpha, \varphi_0) = \{\upsilon : W_n(\upsilon) \le \xi_{1-\alpha}(\mathcal{W}(b_\upsilon, \mathbf{c}_\upsilon; \varphi_0))\}. \tag{3.7}$$

That is, to obtain the null-imposed $(1-\alpha)$-confidence set of $\mathbf{R}\theta_n$ when null-imposition set $\mathcal{H}(\mathbf{R}, \upsilon)$ is a singleton, we follow the steps below:

1. For $H_0 : \mathbf{R}\theta_n = \upsilon$, obtain $\{b_\upsilon, \mathbf{c}_\upsilon\} = \mathcal{H}(\mathbf{R}, \upsilon)$, *i.e.*, the value of $\{b, \mathbf{c}\}$ such that $\mathbf{R}\left[n^{-1}b, n^{1/2}\mathbf{c}^\top\right]^\top = \upsilon$.

2. For $\{b_\upsilon, \mathbf{c}_\upsilon\}$, obtain $\xi_{1-\alpha}(\mathcal{W}(b_\upsilon, \mathbf{c}_\upsilon; \varphi_0))$, the $(1-\alpha)$-quantile of $\mathcal{W}(b, \mathbf{c}; \varphi_0)$.

3. If $W_n(\upsilon) \le \xi_{1-\alpha}(\mathcal{W}(b_\upsilon, \mathbf{c}_\upsilon; \varphi_0))$, then $\upsilon \in CS_n^L(\mathbf{R}\theta_n; 1-\alpha, \varphi_0)$. If not, then $\upsilon \notin CS_n^L(\mathbf{R}\theta_n; 1-\alpha, \varphi_0)$.

4. Go back to step 1 and try another υ.

Since it is not practical to consider all possible values of υ, we propose the use of the grid method to test as many values of υ as possible.

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 $H_0 : \mathbf{R}\theta_n = \pi_n = \upsilon$, 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 $\gamma_n \in \Gamma(1, b, \mathbf{c})$ and the null-imposition set $\mathcal{H}(\mathbf{R}, \upsilon)$ is not a unit set, we consider three different methods to obtain the CS, the null-imposed least-favorable method (Andrews and Cheng, 2012), the projection-based method (Dufour, 1997), and the Bonferroni-based method. The confidence sets obtained by these three methods are accordingly $CS_n^{L,LF}$, $CS_n^{L,P}$, and $CS_n^{L,B}$.

The null-imposed least-favorable method establishes $CS_n^{L,LF}$ by selecting $\{b, \mathbf{c}\}$ with the greatest critical value among $\mathcal{H}(\mathbf{R}, \upsilon)$. Under $\mathbf{R}\left[n^{-1}b_\upsilon, n^{1/2}\mathbf{c}_\upsilon^\top\right]^\top = \upsilon$,

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

To be specific, $CS_n^{L,LF}$ is constructed by the following steps:

1. For $H_0 : \mathbf{R}\theta_n = \upsilon$, obtain all possible $\{b_\upsilon, \mathbf{c}_\upsilon\} \in \mathcal{H}(\mathbf{R}, \upsilon)$, *i.e.*, all possible $\{b, \mathbf{c}\}$ such that $\mathbf{R}\left[n^{-1}b, n^{1/2}\mathbf{c}^\top\right]^\top = \upsilon$.

2. For every $\{b_\upsilon, \mathbf{c}_\upsilon\} \in \mathcal{H}(\mathbf{R}, \upsilon)$, obtain $\xi_{1-\alpha}(\mathcal{W}(b_\upsilon, \mathbf{c}_\upsilon; \varphi_0))$, the $(1-\alpha)$-quantile of $\mathcal{W}(b_\upsilon, \mathbf{c}_\upsilon; \varphi_0)$.

3. If $W_n(\upsilon) \leq \sup_{\{b_\upsilon, \mathbf{c}_\upsilon\} \in \mathcal{H}(\mathbf{R}, \upsilon)} \xi_{1-\alpha}(\mathcal{W}(b_\upsilon, \mathbf{c}_\upsilon; \varphi_0))$, then υ is in the null-imposed least-favorable confidence set, *i.e.*, $\upsilon \in CS_n^{L,LF}(\mathbf{R}\theta_n; 1-\alpha, \varphi_0)$. If not, then $\upsilon \notin CS_n^{L,LF}(\mathbf{R}\theta_n; 1-\alpha, \varphi_0)$.

4. Go back to step 1 and try another υ.

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 $\xi_{1-\alpha}\left(\mathcal{W}\left(b, \mathbf{c}; \varphi_0\right)\right)$. $CS_n^{L,LF}$ is conservative since the greatest critical value is used. However, since the exact values of $\{b, \mathbf{c}\}$ are unknown to econometricians, by using the largest critical value, $CS_n^{L,LF}$ is robust to the risk of under-coverage. In practice, the grid method can be used to test as many values of υ as possible and to obtain the supremum of $\sup_{\{b_\upsilon, \mathbf{c}_\upsilon\}} \xi_{1-\alpha}$.

The projection-based method establishes $CS_n^{L,P}$ by projecting an $(d_\pi + 1)$-sphere to the R^{d_r}-space:

1. Let $\mathbf{R} = \mathbf{P}^P\mathbf{Q}^P$, where $\mathbf{P}^P \in \mathrm{R}^{d_r \times (d_\pi+1)}$ and $\mathbf{Q}^P \in \mathrm{R}^{(d_\pi+1)\times(d_\pi+1)}$ with $\operatorname{rank}\left(\mathbf{P}^P\right) = d_r$ and $\operatorname{rank}\left(\mathbf{Q}^P\right) = d_\pi + 1$. The matrices $\mathbf{P}^P$ and $\mathbf{Q}^P$ always exist since one can always select $\left\{\mathbf{P}^P, \mathbf{Q}^P\right\} = \{\mathbf{R}, \mathrm{I}_{d_\pi+1}\}$. Then the null hypothesis $H_0 : \mathbf{R}\theta_n = \upsilon$ can be written as
$$H_0 : \mathbf{P}^P\mathbf{Q}^P\theta_n = \mathbf{P}^P\varpi. \tag{3.9}$$

2. Consider another null hypothesis $H_0 : \mathbf{Q}^P\theta_n = \varpi$. Let $\mathcal{H}\left(\mathbf{Q}^P, \varpi\right)$ be the null-imposition set with respect to $\mathbf{Q}^P$ and ϖ. By $\operatorname{rank}\left(\mathbf{Q}^P\right) = d_\pi + 1$, $\mathcal{H}\left(\mathbf{Q}^P, \varpi\right)$ is a singleton for any given ϖ. :
$$\begin{aligned}\mathcal{H}\left(\mathbf{Q}^P, \varpi\right) &= \{b_\varpi, \mathbf{c}_\varpi\} \\ &= \left\{b, \mathbf{c} : \mathbf{Q}^P\left[n^{-1}b, n^{1/2}\mathbf{c}^\top\right]^\top = \varpi, \quad \left\{n^{-1}b, n^{1/2}\mathbf{c}^\top\right\} \in \Theta\right\}.\end{aligned} \tag{3.10}$$

3. Obtain the CS for $\mathbf{Q}^P\theta_n$ by equation (3.7). Because $\mathcal{H}\left(\mathbf{Q}^P, \varpi\right) = \{b_\varpi, \mathbf{c}_\varpi\}$ is a singleton, the critical value for testing $H_0 : \mathbf{Q}^P\theta_n = \varpi$ can be directly obtained

by imposing $\{b_\varpi, \mathbf{c}_\varpi\}$ into the asymptotic distribution of Wald statistics. Let $W_n(\varpi)$ denote the Wald statistics *w.r.t.* $H_0 : \mathbf{Q}^P\theta_n = \varpi$ and $\mathcal{W}(b_\varpi, \mathbf{c}_\varpi; \varphi_0)$ denote its limit. Under $\mathbf{Q}^P \left[n^{-1}b_\varpi, n^{1/2}\mathbf{c}_\varpi^\top\right]^\top = \varpi$,

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

4. The confidence set $CS_n^{L,P}$ for $\mathbf{R}\theta_n$ is obtained by projecting the confidence set for $\mathbf{Q}^P\theta_n$, $CS_n^L\left(\mathbf{Q}^P\theta_n; 1-\alpha, \varphi_0\right)$, the $(d_\pi+1)$-sphere, to the R^{d_r}-space.

$$CS_n^{L,P}(\mathbf{R}\theta_n; 1-\alpha, \varphi_0) = \left\{\upsilon : \upsilon = \mathbf{P}^P\varpi, \quad \varpi \in CS_n^L\left(\mathbf{Q}^P\theta_n; 1-\alpha, \varphi_0\right)\right\}. \tag{3.11}$$

For example, in the case when $\mathbf{R}\theta_n = \pi_n$, the projection-based method constructs the $CS_n^{L,P}$ by first, constructing the CS for $\theta_n = \{\beta_n, \pi_n\}$, and second, projecting the CS of θ_n to the R^{d_π}-space. $CS_n^{L,P}$ is also conservative since for any set $\mathcal{C} \subset \mathrm{R}^{(d_\pi+1)}$, the event $\left\{\mathbf{Q}^P\theta_n \in \mathcal{C}\right\}$ entails $\left\{\mathbf{P}^P\mathbf{Q}^P\theta_n \in \mathbf{P}^P\mathcal{C}\right\}$. Intuitively, the projection-based method uses all 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 Bonferroni-based method establishes $CS_n^{L,B}$ using the Bonferroni inequality:

1. For $H_0 : \mathbf{R}\theta_n = \upsilon$, let $\mathbf{Q}^B = \left[\mathbf{R}^\top, \mathbf{P}^{B\top}\right]^\top$, where $\mathbf{P}^B \in \mathrm{R}^{(d_\pi+1-d_r)\times(d_\pi+1)}$ and $\mathbf{Q}^B \in \mathrm{R}^{(d_\pi+1)\times(d_\pi+1)}$ with rank $\left(\mathbf{P}^B\right) = d_\pi + 1 - d_r$ and rank $\left(\mathbf{Q}^B\right) = d_\pi + 1$.

2. Consider a set of new null hypotheses $H_0 : \mathbf{R}\theta_n = \upsilon$ and $H_0 : \mathbf{P}^B\theta_n = \varsigma$, or $H_0 : \mathbf{Q}^B\theta_n = \left(\upsilon^\top, \varsigma^\top\right)^\top$. Let $\mathcal{H}\left(\mathbf{Q}^B, (\upsilon, \varsigma)\right)$ be the null-imposition set with respect to $\mathbf{Q}^B$ and (υ, ς). By rank $\left(\mathbf{Q}^B\right) = d_\pi + 1$, $\mathcal{H}\left(\mathbf{Q}^B, (\upsilon, \varsigma)\right)$ is a singleton

for any given (υ, ς):

$$\mathcal{H}\left(\mathbf{Q}^{B}, (\upsilon, \varsigma)\right) = \{b_{\upsilon\varsigma}, \mathbf{c}_{\upsilon\varsigma}\} \tag{3.12}$$
$$= \left\{b, \mathbf{c} : \mathbf{Q}^{B}\left[n^{-1}b, n^{1/2}\mathbf{c}^{\top}\right]^{\top} = \left(\upsilon^{\top}, \varsigma^{\top}\right)^{\top}, \quad \left\{n^{-1}b, n^{1/2}\mathbf{c}^{\top}\right\} \in \Theta\right\}.$$

3. For $(b_{\upsilon\varsigma}, \mathbf{c}_{\upsilon\varsigma}) = \mathcal{H}\left(\mathbf{Q}^{B}, (\upsilon, \varsigma)\right)$, let $W_{n,1}(\upsilon)$ and $W_{n,2}(\varsigma)$ denote the Wald statistics *w.r.t.* $H_0 : \mathbf{R}\theta_n = \upsilon$ and $H_0 : \mathbf{P}^{B}\theta_n = \varsigma$, and $\mathcal{W}_1(b_{\upsilon\varsigma}, \mathbf{c}_{\upsilon\varsigma}; \varphi_0)$ and $\mathcal{W}_2(b_{\upsilon\varsigma}, \mathbf{c}_{\upsilon\varsigma}; \varphi_0)$ be the corresponding limits. For the given confidence coefficient $1-\alpha$, let $\alpha = \alpha_1 + \alpha_2$, where $\alpha_1, \alpha_2 \geq 0$. The confidence set $CS_n^{L,B}$ is established by obtaining the $1-\alpha_2$ confidence set for $\mathbf{P}^{B}\theta_n$ and the $1-\alpha_1$ confidence set for $\mathbf{R}\theta_n$ simultaneously. Under $\mathbf{Q}^{B}\left[n^{-1}b_{\upsilon\varsigma}, n^{1/2}\mathbf{c}_{\upsilon\varsigma}^{\top}\right]^{\top} = \left(\upsilon^{\top}, \varsigma^{\top}\right)^{\top}$,

$$CS_n^{L,B}(\mathbf{R}\theta_n; 1-\alpha, \varphi_0) = \left\{\upsilon : \begin{array}{c} W_{n,1}(\upsilon) \leq \xi_{1-\alpha_1}\left(\mathcal{W}_1(b_{\upsilon\varsigma}, \mathbf{c}_{\upsilon\varsigma}; \varphi_0)\right), \\ W_{n,2}(\varsigma) \leq \xi_{1-\alpha_2}\left(\mathcal{W}_2(b_{\upsilon\varsigma}, \mathbf{c}_{\upsilon\varsigma}; \varphi_0)\right) \end{array}\right\}. \tag{3.13}$$

For example, in the case when $\mathbf{R}\theta_n = \pi_n$, the Bonferroni-based method constructs the $CS_n^{L,B}$ of π_n by first, selecting $\mathbf{P}^{B}\theta_n = \beta_n$, second, obtaining the $1-\alpha_2$ confidence set for β_n and the $1-\alpha_1$ confidence set for π_n at the same time, and third, using the $1-\alpha_1$ confidence set of π_n as the required $CS_n^{L,B}$. In practice, a simple choice for (α_1, α_2) is $\alpha_1 = \alpha_2 = \alpha/2$. $CS_n^{L,B}$ is also conservative since the Bonferroni inequality only guarantees the coverage probability to be greater than or equal to the given confidence coefficient $1-\alpha$. However, again, intuitively the Bonferroni-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.

Usually CS_n^{L} of $\mathbf{R}\theta_n$, the confidence set when β_n is local-to-zero, is more conservative, and CS_n^D, the confidence set when β_n is not-local-to-zero, is more informative. However, in practice we do not know if $\gamma_n \in \Gamma(1, b, \mathbf{c})$ or not, and without any prior knowledge about the class which γ_n belongs to, we do not know which confidence set we should use.

In this paper we propose an identification-category-selection (ICS) procedure similar to Andrews and Cheng (2012) in the construction of the robust confidence set. Since CS_n^L should be used when $n\beta_n = b = O(1)$, and CS_n^D should be used when $n\beta_n \to \infty$ as $n\beta_n \to \infty$, the ICS procedure uses the estimate of β_n. Let

$$A_n = \frac{\sqrt{n}\widehat{\beta}_n}{\sqrt{\widehat{\mathrm{Avar}}\left(\widehat{\beta}_n\right)}}, \tag{3.14}$$

in which, by equations (2.9) and (3.4),

$$\begin{aligned}
\widehat{\mathrm{Avar}}\left(\widehat{\beta}_n\right) &= \widehat{\sigma}_n^2 \left[1, \mathbf{0}_{1\times d_\pi}\right] \widehat{\mathbf{V}}_n^{-1} \left[1, \mathbf{0}_{1\times d_\pi}\right]^\top, \quad \text{where} \\
\widehat{\sigma}_n^2 &= n^{-1} \sum_{t=1}^{n} \widehat{\varepsilon}_t^2, \\
\widehat{\mathbf{V}}_n &= n^{-1} \sum_{t=1}^{n} \begin{bmatrix} \left(y_{t-1} - X_t^\top \widehat{\pi}_n\right)^2 & -\widehat{\beta}_n \left(y_{t-1} - X_t^\top \widehat{\pi}_n\right) X_t^\top \\ -\widehat{\beta}_n X_t \left(y_{t-1} - X_t^\top \widehat{\pi}_n\right) & \widehat{\beta}_n^2 X_t X_t^\top \end{bmatrix}.
\end{aligned}$$

Also let k_n be a sequence such that

$$k_n \to \infty, \quad \frac{k_n}{\sqrt{n}} \to 0. \tag{3.15}$$

The ICS procedure selects the confidence set according to the value of A_n and k_n:

$$CS_n^{ICS} = \begin{cases} CS_n^L, & \text{if } A_n \leq k_n, \\ CS_n^D, & \text{if } A_n > k_n. \end{cases}, \tag{3.16}$$

and $CS_n^{ICS,LF}$, $CS_n^{ICS,P}$ and $CS_n^{ICS,B}$ denote the robust confidence sets selected between CS_n^D and $CS_n^{L,LF}$, $CS_n^{L,P}$ and $CS_n^{L,B}$ respectively.

Theoretically k_n can be selected as any sequence such that $k_n \to \infty$ and $k_n / \sqrt{n} \to 0$. In this paper we select

$$k_n = c_k \log(n),$$

where $c_k > 0$ is a constant.

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

$$AsySz(CS_n) = \liminf_{n\to\infty} \inf_{\gamma_n \in \Gamma_n} \mathrm{P}(\mathbf{R}\theta_n \in CS_n). \tag{3.17}$$

Notice that in the definition of the asymptotic size (equation (3.17)) $\liminf_{n\to\infty}$ 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 6 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) and Mikusheva (2007, 2012). Theorem 6 shows the correctness of the asymptotic sizes of CS_n^{ICS} and $CS_n^{ICS,LF}$. The projection-based $CS_n^{ICS,P}$ and the Bonferroni-based $CS_n^{ICS,B}$, however, may be asymptotic oversized, *i.e.*, may have an asymptotic size higher than the required confidence coefficient $1-\alpha$.

Theorem 6 *Suppose that Assumptions 1, 2 and 3 hold and* $y_0 = o_p\left(n^{1/2}\right)$ *when* $\gamma_n \in \Gamma\left(1, b, \mathbf{c}\right)$.

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

Remark 4 1. *As in Andrews and Cheng (2012, 2013a, 2013b), we obtain the* CS *by inverting the* t *or Wald 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* γ_n *belongs to the distant-from-zero* β_n *class* $\Gamma\left(h, b\right)$ *or the close-to-zero* β_n *class* $\Gamma\left(h, b, \mathbf{c}\right)$*, the asymptotic distributions of* $\widehat{\theta}_n$ *will depend on the unknown values of* $\left\{h, b, \mathbf{c}\right\}$*. Therefore, we are not able to obtain the* CS *from the asymptotic distributions of* $\widehat{\theta}_n$ *directly. On the other hand, the* t *and Wald statistics will have standard and pivotal asymptotic distributions. Therefore, to consider the* t *and 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, in the simple case when the null-imposition set* $\mathcal{H}(\mathbf{R}, \upsilon)$ *is a singleton, we are able to plug in the values of* $\{b, \mathbf{c}\}$ *under the null hypothesis and to obtain the asymptotic distribution of the Wald statistics. When the null-imposition set* $\mathcal{H}(\mathbf{R}, \upsilon)$ *is not a singleton, we also use the onto mapping to obtain the conservative confidence set. For example, 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*. A wider and less informative confidence set may be obtained.*

Again, when $\gamma_n \in \Gamma(1, b, \mathbf{c})$, 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 $\widehat{\varphi}_n = \left\{\widehat{\mu}_{X,n}, \widehat{\mathbf{M}}_{X,n}, \widehat{\sigma}_n^2\right\}$, the CS can be obtained by replacing φ_0 with $\widehat{\varphi}_n$.

The following example shows the coverage probabilities of the null-imposed least-favorable CS ($CS_n^{L,LF}$), the projection-based CS ($CS_n^{L,P}$), the Bonferroni-based CS ($CS_n^{L,B}$), the CS from the standard (Newey and McFadden, 1994) based on the χ^2 distribution, and the identification-category-selection CS ($CS_n^{ICS,LF}$) with $k_n = c_k \log(n)$.

Example 4 (Example 1 continued) *Again, consider equation (2.10).*

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

where $x_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$*,* $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$*,* $\beta_n \in [0.02, 0.6]$*,* $\pi_{0,n} \in [0,2]$*,* $\pi_{1,n} = 2$*, and* $n = 100$ *or* 250*. 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 (*$CS_n^{L,LF}$*), the projection-based method (*$CS_n^{L,P}$*), the Bonferroni-based method (*$CS_n^{L,B}$*), the standard method (Newey and McFadden, 1994) based on the* $\chi^2(1)$ *distribution (*CS_n^{D}*), and the identification-category-selection (ICS) procedure between* $CS_n^{L,LF}$ *and* CS_n^{D} *(*$CS_n^{ICF,LF}$*).*

Figures 3.5 – 3.8 provide the simulated coverage probabilities of the CSs. For both cases $1-\alpha = 0.8$ *and* 0.9 *and for every values of* β_n *and* $\pi_{0,n}$*,* $CS_n^{L,LF}$*s,* $CS_n^{L,P}$*s and* $CS_n^{L,B}$*s have coverage probabilities greater than the confidence coefficient* $1-\alpha$*, while the coverage probabilities of the* $\chi^2(1)$ *CSs are seriously downward biased, especially when* β_n *is close to zero. Under most circumstances* $CS_n^{L,LF}$*s have coverage probabilities closer to* $1-\alpha$ *than* $CS_n^{L,P}$ *and* $CS_n^{L,B}$*. especially when* β_n *is close to zero. However, when* β_n *is not close to zero,* $CS_n^{L,P}$ *and* $CS_n^{L,B}$ *may have better coverage probabilities. When the sample size* n *increases from* 100 *to* 250*, all three conservative CSs have coverage probabilities closer to* $1-\alpha$*. Both* $CS_n^{ICS,LF}$*s with* $c_k = 1$ *and* 2 *have coverage probabilities closer to* $1-\alpha$ *than* $CS_n^{L,LF}$*. When the sample size* $n = 100$*, the coverage probabilities of the* CS_n^{ICS}*s are downward biased when* β_n *is not close to zero., but the bias is much smaller when* n *increases from* 100 *to* 250*.*

For all results $50,000$ *simulation repetitions are used. For values of parameters,* $1,230$ *grids are generated in the true parameter space* $\Theta^* = [0, 0.6] \times [0, 2]$*, where grids for* β_n *and* $\pi_{0,n}$ *are respectively of size* 0.02 *and* 0.05*.*

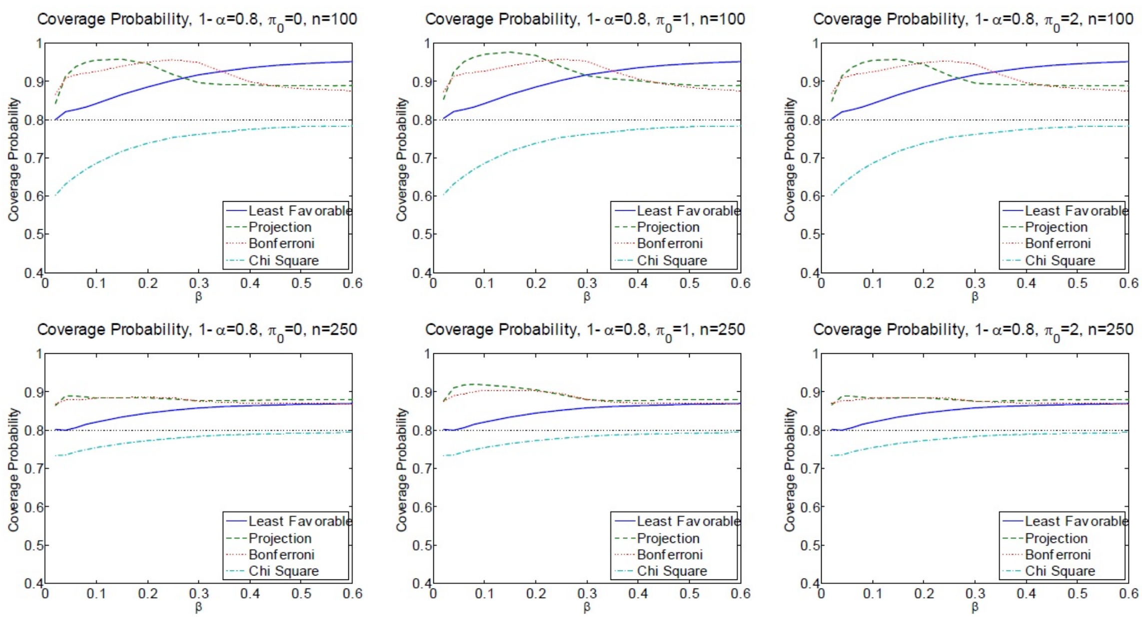

Figure 3.5: Coverage probabilities of $CS_n^{L,LF}$, $CS_n^{L,P}$, $CS_n^{L,B}$ and $\chi^2(1)$ CS for $\pi_{1,n} = 2$, $1-\alpha = 0.8$, $n = 100$ or 250, Example 4

The first row is the simulated coverage probabilities of the least-favorable confidence sets $CS_n^{L,LF}$, the projection-based confidence sets $CS_n^{L,P}$, the Bonferroni-based confidence sets $CS_n^{L,B}$ and the standard confidence sets based on the $\chi^2(1)$ distribution of Example 4 with $1-\alpha = 0.8$, $\beta_n \in [0.02, 0.6]$, $\pi_{0,n} = 0$, 1 and 2, $\pi_{1,n} = 2$ and $n = 100$. The second row is the coverage probabilities with $n = 250$.

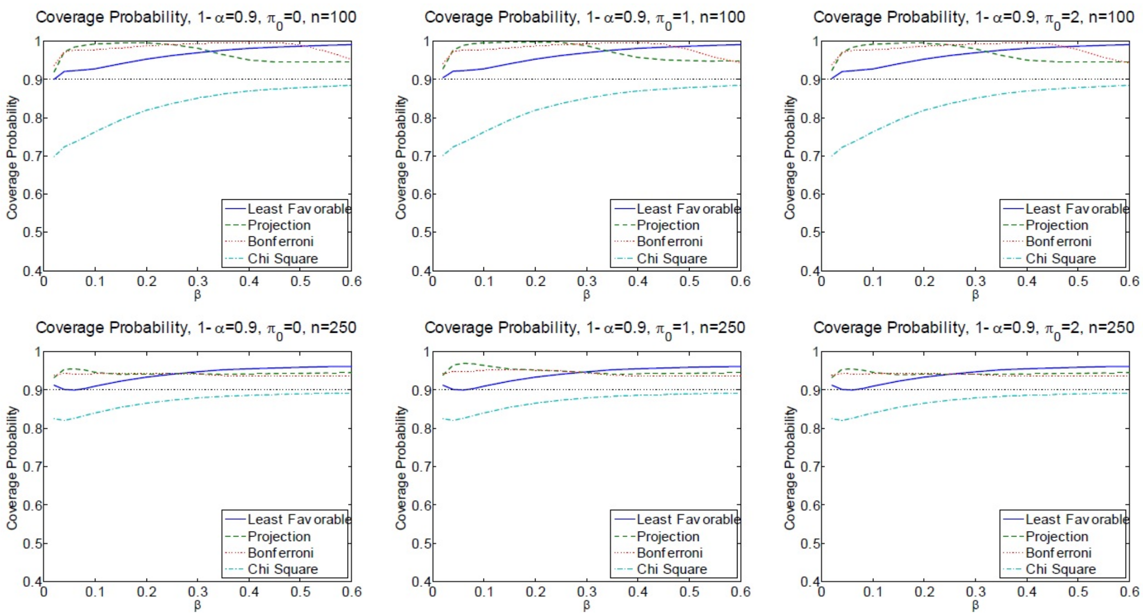

Figure 3.6: Coverage probabilities of $CS_n^{L,LF}$, $CS_n^{L,P}$, $CS_n^{L,B}$ and $\chi^2(1)$ CS for $\pi_{1,n} = 2$, $1-\alpha = 0.9$, $n = 100$ or 250, Example 4
The first row is the simulated coverage probabilities of the least-favorable confidence sets $CS_n^{L,LF}$, the projection-based confidence sets $CS_n^{L,P}$, the Bonferroni-based confidence sets $CS_n^{L,B}$ and the standard confidence sets based on the $\chi^2(1)$ distribution of Example 4 with $1-\alpha = 0.9$, $\beta_n \in [0.02, 0.6]$, $\pi_{0,n} = 0$, 1 and 2, $\pi_{1,n} = 2$ and $n = 100$. The second row is the coverage probabilities with $n = 250$.

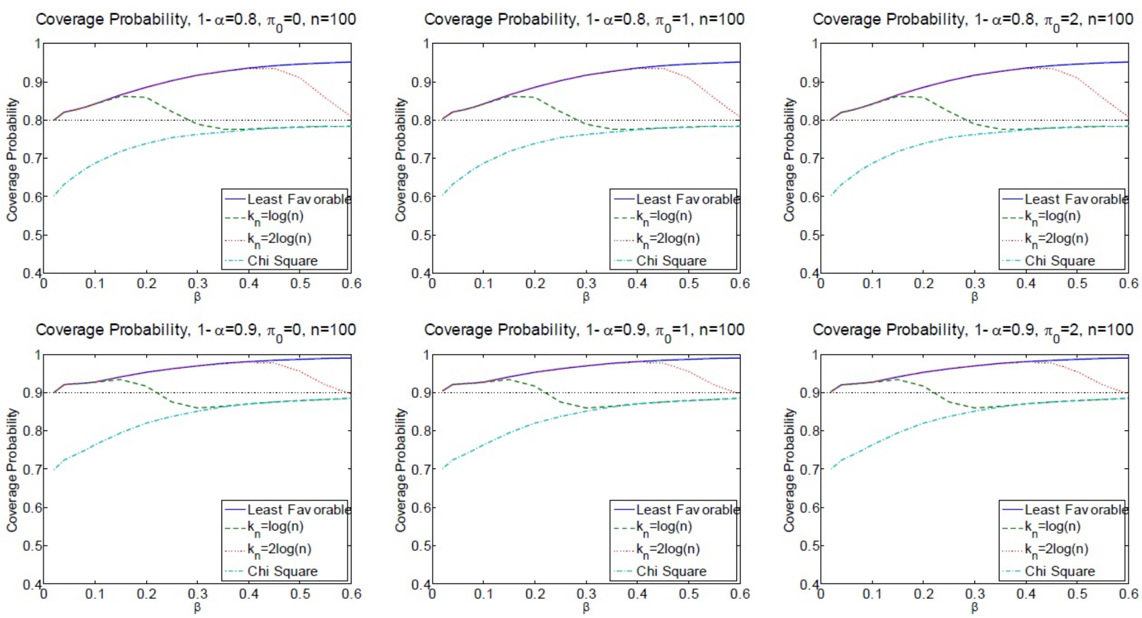

Figure 3.7: Coverage probabilities of $CS_n^{L,LF}$, $\chi^2(1)$ CS and $CS_n^{L,ICS}$ with $c_k = 1$ or 2 for $\pi_{1,n} = 2$, $1 - \alpha = 0.8$, $n = 100$ or 250, Example 4
The first row is the simulated coverage probabilities of the least-favorable confidence sets $CS_n^{L,LF}$, the standard confidence sets based on the $\chi^2(1)$ distribution and the identification-category-selection confidence sets ($CS_n^{L,ICS}$) with $c_k = 1$ or 2 of Example 4 with $1 - \alpha = 0.8$, $\beta_n \in [0.02, 0.6]$, $\pi_{0,n} = 0$, 1 and 2, $\pi_{1,n} = 2$ and $n = 100$. The second row is the coverage probabilities with $n = 250$.

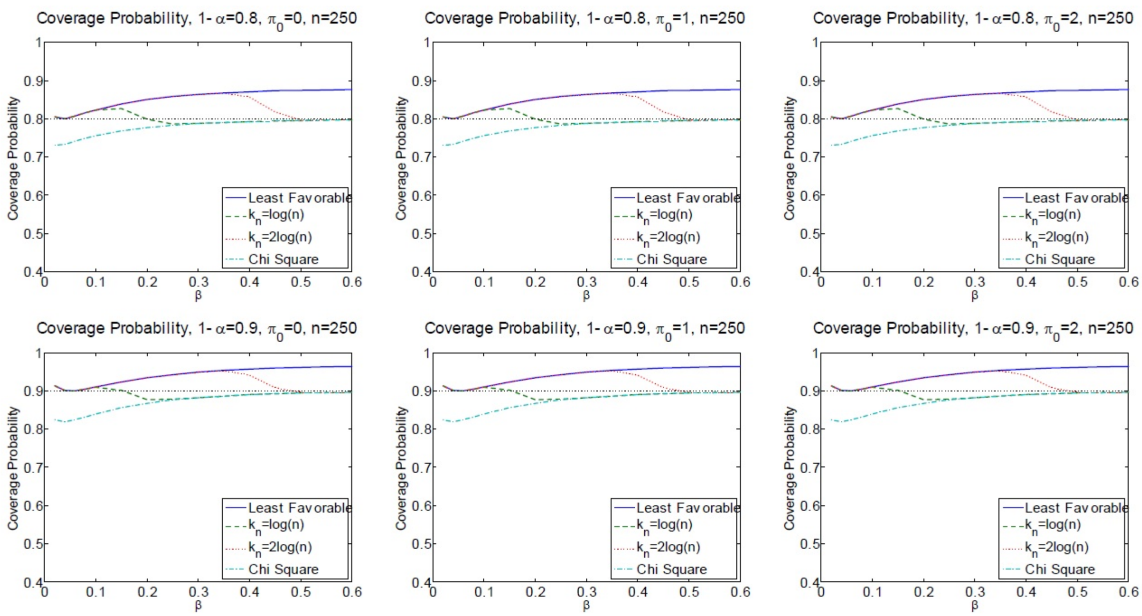

Figure 3.8: Coverage probabilities of $CS_n^{L,LF}$, $\chi^2(1)$ CS and $CS_n^{L,ICS}$ with $c_k = 1$ or 2 for $\pi_{1,n} = 2$, $1 - \alpha = 0.9$, $n = 100$ or 250, Example 4

The first row is the simulated coverage probabilities of the least-favorable confidence sets $CS_n^{L,LF}$, the standard confidence sets based on the $\chi^2(1)$ distribution and the identification-category-selection confidence sets ($CS_n^{L,ICS}$) with $c_k = 1$ or 2 of Example 4 with $1 - \alpha = 0.9$, $\beta_n \in [0.02, 0.6]$, $\pi_{0,n} = 0$, 1 and 2, $\pi_{1,n} = 2$ and $n = 100$. The second row is the coverage probabilities with $n = 250$.

4. EMPIRICAL APPLICATION: U.S.'S FORECAST-BASED MPRF

In this paper we reexamine the empirical findings of Clarida *et al.* (2000) with more recent real-time data. In their seminal paper, Clarida *et al.* (2000) estimated the monetary policy reaction function of U.S. for 1960:1 – 1996:4 by GMM, using the lags of $\{i_t, \dot{p}_{t,k}, x_{t,k}\}$ as IV.

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

However, many empirical studies (*e.g.*, Inoue and Rossi, 2011; Mavroeidis, 2004, 2010) suggested that the lags of $\{i_t, \dot{p}_{t,k}, x_{t,k}\}$ are only weakly correlated to $\{\dot{p}_{t,k}, x_{t,k}\}$. 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 the annualized inflations and the average output gaps ($\{\mathrm{E}_t\dot{p}_{t,k}, \mathrm{E}_t x_{t,k}\}$) of the Federal Reserve. As the real-time data is used, the model (equation (4.1)) can be estimated by NLS without using any IV.

According to our asymptotic theory, we construct the 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). Our confidence sets are robust to the value of the smoothing coefficient $\rho = 1-\beta$. The null-imposed least-favorable confidence sets ($CS_n^{L,LF}$) will have correct asymptotic sizes, while the projection-based confidence sets ($CS_n^{L,P}$) and

Bonferroni-based confidence sets ($CS_n^{L,B}$) are asymptotically over-coverage but may be more informative. We use the identification-category-selection (ICS) procedure in equation (3.16) to select the appropriate CS between $CS_n^{L,LF}$ and the standard one (CS_n^D), which is based on $\chi^2(2)$ distribution.

The real-time data is available in the Federal Reserve Bank of Philadelphia for 1987:3–2007:4, *i.e.*, $n = 82$[1]. We consider the forecast horizons $k = 0$ or 1. As in Orphanides (2001), for the *ex ante* forecasts of the annualized inflations and the average output gaps ($\{\mathrm{E}_t\dot{p}_{t,k}, \mathrm{E}_t x_{t,k}\}$) of the Federal Reserve, we use the forecasts corresponding to the FOMC meeting closest to the middle of the quarter. In the period relevant for this study (1987:3–2007:4), the FOMC had eight meetings per year, typically in February, March, May, July, August, September, November, and December. In this paper we use the forecasts corresponding to the February, May, August, and November meetings. For the interest rates ($\{i_t\}$), as in Nikolsko-Rzhevskyy (2011), we use the average of effective federal funds target rates at the last month of each quarter, giving the Fed time to respond to intra-quarter news. Figure 4.1 provides the plots of the data. Table 4.1 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 4.2 reports the $CS_n^{L,LF}$ and the standard CS based on $\chi^2(2)$ distribution. When the forecast horizon $k = 0$, the standard CSs contain some values of $\{\pi_{\dot{p}}, \pi_x\}$ not in

[1]Both expected inflations and output gaps are from the Real-Time Data Research Center in Fed Philadelphia. The expected inflations are from the Philadelphia Fed's Greenbook Data Set http://www.phil.frb.org/research-and-data/real-time-center/greenbook-data/philadelphia-data-set.cfm, and the expected output gaps are from the Output Gap and Financial Assumptions from the Board of Governors http://www.phil.frb.org/research-and-data/real-time-center/greenbook-data/gap-and-financial-data-set.cfm.

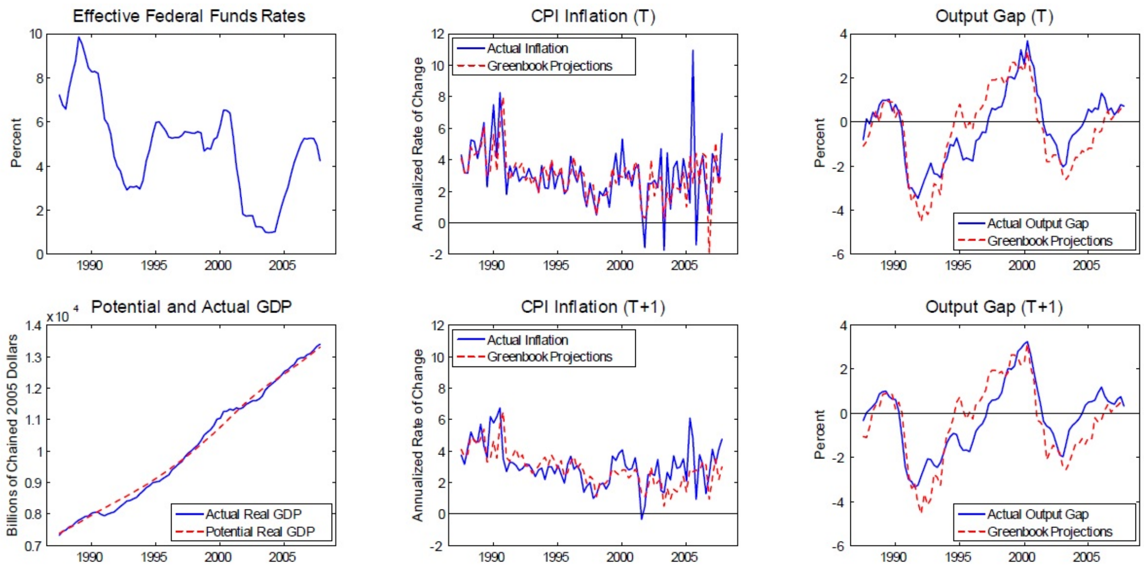

Figure 4.1: Federal funds target rates, inflations and output gaps

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.

Table 4.1: 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)		

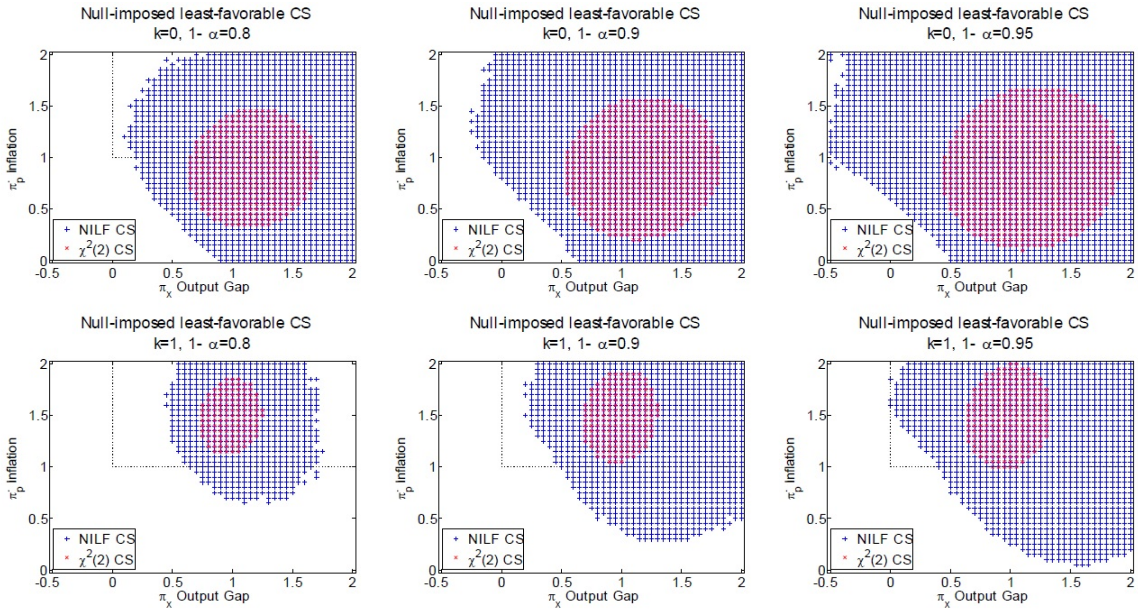

Figure 4.2: Least-favorable CSs for the reaction coefficients
The first and second rows are respectively for $k = 0$ and 1. The first to third panels are respectively for $1 - \alpha = 0.8$, 0.9 and 0.95. The dot line denotes the determinacy region $\mathcal{DR} = \{\pi_{\dot{p}} > 1, \pi_x > 0\}$.

the determinacy region $\mathcal{DR} = \{\pi_{\dot{p}} > 1, \pi_x > 0\}$, while when $k = 1$ the standard CSs are contained in the region $\mathcal{DR}$. 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 ($CS_n^{L,P}$) and the Bonferroni-based CS ($CS_n^{L,B}$) of $\{\pi_{\dot{p}}, \pi_x\}$. Figures 4.3 and 4.4 report the $CS_n^{L,P}$s and $CS_n^{L,B}$s. For all cases $CS_n^{L,P}$s and $CS_n^{L,B}$s contain many values not in $\mathcal{DR}$.

To decide to use the standard and more informative CS or the conservative $CS_n^{L,LF}$, we consider our identification-category-selection procedure. Let $k_n = \log(n)$. Since $n = 82$, $\log(n) = \log(82) = 4.41$. For the case with forecast horizon $k = 0$, $A_n = 3.63 < 4.41$. Therefore we select $CS_n^{L,LF}$s, which contain many values not in

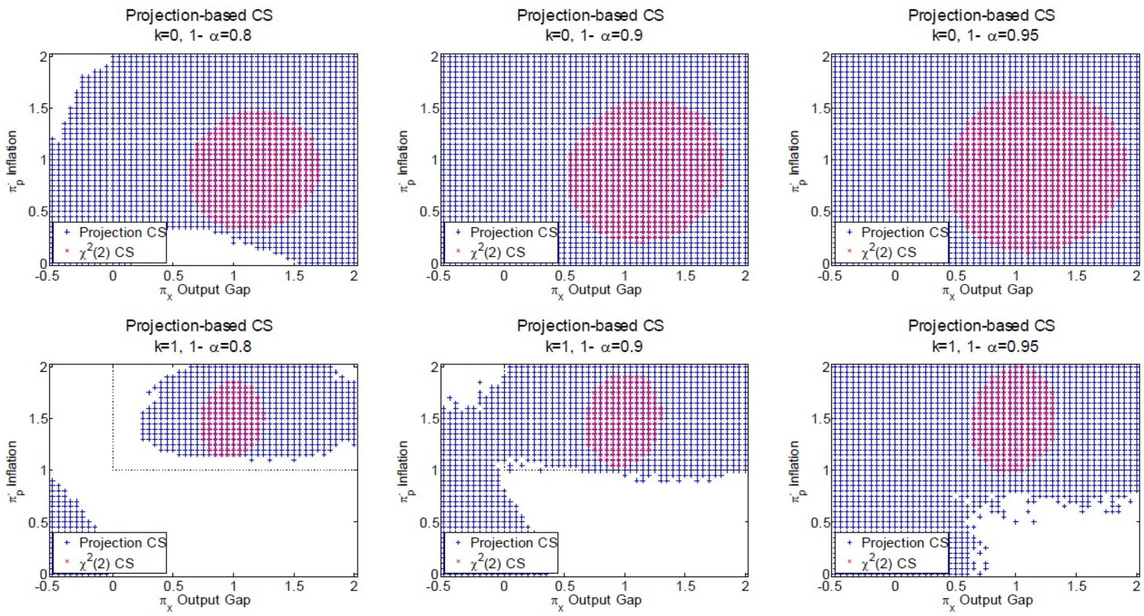

Figure 4.3: Projection-based CSs for the reaction coefficients
The first and second rows are respectively for $k = 0$ and 1. The first to third panels are respectively for $1 - \alpha = 0.8$, 0.9 and 0.95. The dot line denotes the determinacy region $\mathcal{DR} = \{\pi_{\dot{p}} > 1, \pi_x > 0\}$.

the region $\mathcal{DR} = \{\pi_{\dot{p}} > 1, \pi_x > 0\}$ for all three confidence coefficients $1 - \alpha = 0.8$, 0.9 and 0.95. And for the case $k = 1$, $A_n = 5.71 > 4.41$. Therefore we select the standard CSs, which are contained in the region $\mathcal{DR} = \{\pi_{\dot{p}} > 1, \pi_x > 0\}$ when $1 - \alpha = 0.8$, 0.9 and 0.95. Our empirical application suggests that for the case $k = 0$, the NLS estimates for $\{\pi_{\dot{p}}, \pi_x\}$ are not accurate sufficiently to rule out the possibility of indeterminacy. But in the case $k = 1$, the possibility of indeterminacy may be rule out.

For all results $5,000$ simulation repetitions are used. For values of parameters, grids are generated in the true parameter space $\Theta^* = [0, 0.3] \times [-1, 3]^3$, where grids for β, π_α, $\pi_{\dot{p}}$ and π_x are respectively of size 0.02, 0.05, 0.05 and 0.05.

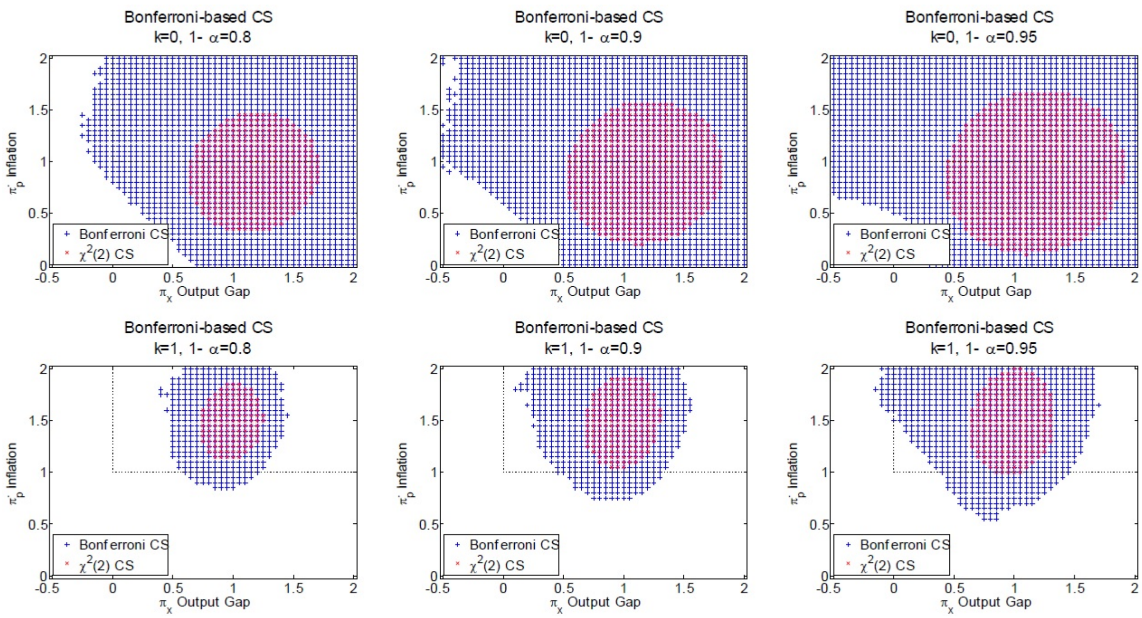

Figure 4.4: Bonferroni-based CSs for the reaction coefficients
The first and second rows are respectively for $k = 0$ and 1. The first to third panels are respectively for $1 - \alpha = 0.8$, 0.9 and 0.95. The dot line denotes the determinacy region $\mathcal{DR} = \{\pi_{\dot{p}} > 1, \pi_x > 0\}$.

5. CONCLUDING REMARKS

In this paper we modify the method of Andrews and Cheng (2012) on inference with weak and semi-strong identification and establish the asymptotic distributions of the NLS estimator and tests for the forecast-based monetary policy reaction function ($MPRF$) with a close-to-unity smoothing coefficient. Conservative confidence sets with correct or 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 application suggests that for the case with forecast horizon $k = 0$, 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. But in the case $k = 1$, the possibility of indeterminacy may be rule out.

REFERENCES

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

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.

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

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

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

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.

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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.

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

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.

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

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

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

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

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

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

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.

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.

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

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

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

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

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

Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65, 557–586.

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

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

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.

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

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

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

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

APPENDIX A

LEMMA 2, THEOREM 2 AND THEOREM 5

In this appendix we provide complete versions of Lemma 2, Theorem 2 and Theorem 5. Proofs are collected in Appendix B.

Lemma 2. Suppose that Assumptions 1, 2 and 3 hold, $\gamma_n \in \Gamma(1, b, \mathbf{c})$, 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 $n \to \infty$,

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

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

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

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

2. $n^{-1}\left[\partial^2 Q_n(0,\pi)/\partial\beta^2\right] \Rightarrow \mathcal{H}(\kappa_\pi, b, \mathbf{c}; \varphi_0)$, 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 2. Suppose that Assumptions 1, 2 and 3 hold, $\gamma_n \in \Gamma(1,b,\mathbf{c})$, and $y_0 = o_p\left(n^{1/2}\right)$. Then

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

where

$$\begin{aligned}
\widehat{\kappa}_\pi(b,\mathbf{c};\varphi_0) &= \frac{\sigma_\varepsilon^2 \int_0^1 \mathcal{W}_\varepsilon^2(r)\,dr \sigma_\varepsilon \mathbf{M}_X^{1/2}\mathcal{Z} - \sigma_\varepsilon^2 \int_0^1 \mathcal{W}_\varepsilon(r)\,d\mathcal{W}_\varepsilon(r)\,\mu_X \sigma_\varepsilon \int_0^1 \mathcal{W}_\varepsilon(r)\,dr}{\mu_X \sigma_\varepsilon \int_0^1 \mathcal{W}_\varepsilon(r)\,dr \cdot \sigma_\varepsilon \mathbf{M}_X^{1/2}\mathcal{Z} - \sigma_\varepsilon^2 \int_0^1 \mathcal{W}_\varepsilon(r)\,d\mathcal{W}_\varepsilon(r) \cdot \mathbf{M}_X}, \\
\widehat{\lambda}_\beta(\widehat{\kappa}_\pi, b, \mathbf{c}; \varphi_0) &= \frac{\sigma_\varepsilon \mathbf{M}_X^{1/2}\mathcal{Z} \cdot \widehat{\kappa}_\pi - \sigma_\varepsilon^2 \int_0^1 \mathcal{W}_\varepsilon(r)\,d\mathcal{W}_\varepsilon(r)}{\mathbf{M}_X \cdot \widehat{\kappa}_\pi^2 - 2\mu_X \sigma_\varepsilon \int_0^1 \mathcal{W}_\varepsilon(r)\,dr \cdot \widehat{\kappa}_\pi + \sigma_\varepsilon^2 \int_0^1 \mathcal{W}_\varepsilon^2(r)\,dr}.
\end{aligned}$$

Theorem 5. Suppose that Assumptions 1, 2 and 3 hold.

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

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

$y_0 = o_p\left(n^{1/2}\right)$,

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

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 $\gamma_n \in \Gamma(h,b,\mathbf{c})$, i.e., $\beta_n = b/n^h$ with $0 < b < \infty$ and $\pi_n = n^{-1/2+h}\mathbf{c}$,

where $h \in [1/2, 1)$, $T_n \overset{A}{\sim} \mathcal{N}(0,1)$, and $W_n \overset{A}{\sim} \chi^2(d_r)$.

4. When $\gamma_n \in \Gamma(h,b)$, i.e., $\beta_n = b \big/ n^h$ with $0 < b < \infty$ and $h \in (0, 1/2]$, $T_n \overset{A}{\sim} \mathcal{N}(0,1)$, and $W_n \overset{A}{\sim} \chi^2(d_r)$.

APPENDIX B

PROOFS OF THEOREMS AND LEMMAS

Proof. (Theorem 1) The proof directly follows Theorems 2.7 and 3.1 of Newey and McFadden (1994). For the consistency, let

$$\begin{aligned} Q_0(\theta) &= \frac{1}{2}\mathrm{E}\left\{[y_t - m(y_{t-1}, X_t; \theta)]^2\right\} \\ &= \frac{1}{2}\mathrm{E}\left\{\left[y_t - (1-\beta) y_{t-1} - \beta X_t^\top \pi\right]^2\right\}. \end{aligned}$$

By Assumption 3, $m(y_{t-1}, X_t; \theta_0) = \mathrm{E}(y_t | y_{t-1}, X_t)$. By the fact that the mean square error has a unique minimum at the conditional mean, $Q_0(\theta)$ is uniquely minimized at θ_0. By Assumption 1, θ_0 is an element of the interior of the convex set Θ^* and $Q_n(\theta)$ is concave. By Assumptions 1, 2 and 3, the law of number for stationary ergodic sequences (White, 2001, Theorem 3.34, p. 44), and the law of number for martingale difference sequences (White, 2001, Theorem 3.76, p. 60), $Q_n(\theta) \overset{a.s.}{\to} Q_0(\theta)$ for all $\theta \in \Theta^*$. Therefore, by Theorem 2.7 of Newey and McFadden (1994), $\widehat{\theta}_n \overset{p}{\to} \theta_n = \theta_0$.

For the asymptotic normality, we have already shown that $\widehat{\theta}_n \overset{p}{\to} \theta_n = \theta_0$. By Assumption 1, $\theta_0 \in \mathit{interior}(\Theta^*)$. $Q_n(\theta)$ is clearly twice continuously differentiable with

$$\nabla_\theta Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \begin{bmatrix} y_{t-1} - X_t^\top \pi \\ X_t \end{bmatrix} \left[y_t - (1-\beta) y_{t-1} - \beta X_t^\top \pi\right],$$

$$\nabla_{\theta\theta^\top} Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \begin{bmatrix} \left(y_{t-1} - X_t^\top \pi\right)^2 & -\beta\left(y_{t-1} - X_t^\top \pi\right) X_t^\top \\ -\beta X_t \left(y_{t-1} - X_t^\top \pi\right) & \beta^2 X_t X_t^\top \end{bmatrix}.$$

Let

$$\mathcal{V}(\theta;\gamma_n) = \mathrm{E}\left\{\left[\begin{array}{cc} \left(y_{t-1} - X_t^\top \pi\right)^2 & -\beta\left(y_{t-1} - X_t^\top \pi\right) X_t^\top \\ -\beta X_t \left(y_{t-1} - X_t^\top \pi\right) & \beta^2 X_t X_t^\top \end{array}\right]\right\}.$$

and $\mathcal{V}_0(\gamma_n) = \mathcal{V}(\theta_0;\gamma_n) = \mathrm{E}\nabla_{\theta\theta^\top} Q_n(\theta_0)$. Clearly $\mathcal{V}(\theta;\gamma_n)$ is continuous with respect to θ and nonsingular. Also by Assumptions 1, 2 and 3, the law of number for stationary ergodic sequences (White, 2001, Theorem 3.34, p. 44), and the law of number for martingale difference sequences (White, 2001, Theorem 3.76, p. 60), $\nabla_{\theta\theta^\top} Q_n(\theta) \xrightarrow{p} \mathcal{V}(\theta;\gamma_n)$. Furthermore, by Assumptions 1, 2 and 3, and the central limit theorem for martingale difference sequences (White, 2001, Theorem 5.24, p. 133),

$$\sqrt{n}\nabla_\theta Q_n(\theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[\begin{array}{c} y_{t-1} - X_t^\top \pi_0 \\ X_t \end{array}\right] \varepsilon_t \overset{A}{\sim} \mathcal{N}\left(\mathbf{0}_{(d_\pi+1)\times 1}, \sigma_\varepsilon^2 \mathcal{V}_0(\gamma_n)\right).$$

Therefore, by Theorem 3.1 of Newey and McFadden (1994),

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

■

Proof. (Lemma 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 martingale difference sequences (White, 2001, Theorem

5.24, p. 133), and Lemma 6 with $b \to 0$,

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

Then by the first order condition of equation (2.2),

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

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

■

Proof. (Lemma 2)

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

martingale difference sequences (White, 2001, Theorem 5.24, p. 133),

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

2. $\left(n^{-1}\left[\partial^2 Q_n(0,\pi)/\partial\beta^2\right]\right)$ By Lemma 6 and the law of large number for stationary

ergodic sequences (White, 2001, Theorem 3.34, p. 44),

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

■

Proof. (Theorem 2) For notational simplicity, let $\widehat{\lambda}_\beta$ denote $\widehat{\lambda}_\beta(\widehat{\kappa}_\pi(b,\mathbf{c};\varphi_0),b,\mathbf{c};\varphi_0)$, $\widehat{\kappa}_\pi$ 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

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

Then by equations (2.2), (2.4), (2.7) and (2.8), $\left\{\widehat{\lambda}_{\beta,n},\widehat{\kappa}_{\pi,n}\right\}$ and $\left\{\widehat{\lambda}_\beta,\widehat{\kappa}_\pi\right\}$ are respectively the unique minimizers of $q_n(\lambda_\beta,\kappa_\pi)$ and $q(\lambda_\beta,\kappa_\pi)$ in $\mathrm{R}^{d_\pi+1}$, *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 (2.6), 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 R when $n \to \infty$.

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

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

By the concavity of $q_n(\lambda_\beta, \kappa_\pi)$ and $q(\lambda_\beta, \kappa_\pi)$, $\left\{\widehat{\lambda}^*_{\beta,n}, \widehat{\kappa}^*_{\pi,n}\right\}$ and $\left\{\widehat{\lambda}^*_{\beta}, \widehat{\kappa}^*_{\pi}\right\}$ are unique. If $\left\{\widehat{\lambda}^*_{\beta,n}, \widehat{\kappa}^*_{\pi,n}\right\}$ are tight for every $n \in \mathrm{N}$, by the compactness of 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 $n \to \infty$,

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

Since C is arbitrary, the desired results directly follow. That is, for any compact subset $\mathrm{C} \subset \mathrm{R}$ 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 $n \to \infty$,

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

It only remains to show the tightness of $\left\{\widehat{\lambda}_{\beta,n},\widehat{\kappa}_{\pi,n}\right\}=\left\{n^{-1}\widehat{\beta}_n,n^{-1/2}\widehat{\pi}_n\right\}$. By the first order condition of equation (2.2), the law of number for stationary ergodic sequences (White, 2001, Theorem 3.34, p. 44), the central limit theorem for martingale difference sequences (White, 2001, Theorem 5.24, p. 133), and Lemma 6,

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

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

■

Proof. (Proposition 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 law of number for martingale difference sequences

(White, 2001, Theorem 3.76, p. 60),

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

■

Proof. (Lemma 4)

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

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

By Lemma 7, and the central limit theorem for martingale difference sequences (White, 2001, Theorem 5.24, p. 133),

$$\begin{aligned} & n^{-1/2-h/2}\sum_{t=1}^{n}\left(y_{t-1} - X_t^\top\pi_n\right)\varepsilon_t \overset{A}{\sim} \mathcal{N}\left(0, (2b)^{-1}\sigma_\varepsilon^4\right), \\ & -\beta_n n^{-1/2+h}\sum_{t=1}^{n} X_t\varepsilon_t \overset{A}{\sim} \mathcal{N}\left(0, \sigma_\varepsilon^2 b^2 \mathbf{M}_X\right).\end{aligned}$$

By Assumption 3, ε_t is independent to (y_{t-1}, X_t). Thus by Lemma 7,

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

Therefore

$$n^{1/2}B^{-1}(h)\,D_\theta Q_n(\theta_n) \overset{A}{\sim} \mathcal{N}\left(\mathbf{0}_{(d_\pi+1)\times 1}, \begin{bmatrix} (2b)^{-1}\sigma_\varepsilon^4 & \mathbf{0}_{1\times d_\pi} \\ \mathbf{0}_{d_\pi\times 1} & \sigma_\varepsilon^2 b^2 \mathbf{M}_X \end{bmatrix}\right).$$

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

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

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

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

and

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

Therefore,

$$B^{-1}(h)\, D_{\theta\theta^{\top}} Q_n(\theta_n)\, B^{-1}(h) \xrightarrow{p} \begin{bmatrix} (2b)^{-1}\sigma_{\varepsilon}^2 & \\ \mathbf{0}_{d_\pi \times 1} & b^2 \mathbf{M}_X \end{bmatrix}.$$

■

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 equation (2.1), the first order conditions of equation (2.2) can be written as:

$$\begin{aligned} 0 = &\sum_{t=1}^{n} \left[(y_{t-1} - \pi_n X_t) - X_t(\widehat{\pi}_n - \pi_n)\right] \\ &\left\{\left(\widehat{\beta}_n - \beta_n\right)\left[(y_{t-1} - \pi_n X_t) - X_t(\widehat{\pi}_n - \pi_n)\right] - \beta_n(\widehat{\pi}_n - \pi_n) X_t + \varepsilon_t\right\}, \\ 0 = &\sum_{t=1}^{n} X_t \left\{\left(\widehat{\beta}_n - \beta_n\right)\left[(y_{t-1} - \pi_n X_t) - X_t(\widehat{\pi}_n - \pi_n)\right] - \beta_n(\widehat{\pi}_n - \pi_n) X_t + \varepsilon_t\right\}, \end{aligned}$$

that is,

$$\begin{aligned} 0 = &\left(\widehat{\beta}_n - \beta_n\right)\left[\sum_{t=1}^{n} (y_{t-1} - \pi_n X_t)^2\right] \\ &- 2\left(\widehat{\beta}_n - \beta_n\right)(\widehat{\pi}_n - \pi_n)\left[\sum_{t=1}^{n} X_t(y_{t-1} - \pi_n X_t)\right] \\ &+ \left(\widehat{\beta}_n - \beta_n\right)(\widehat{\pi}_n - \pi_n)^2\left(\sum_{t=1}^{n} X_t^2\right) - \beta_n(\widehat{\pi}_n - \pi_n)\left[\sum_{t=1}^{n} X_t(y_{t-1} - \pi_n X_t)\right] \\ &+ \beta_n(\widehat{\pi}_n - \pi_n)^2\left(\sum_{t=1}^{n} X_t^2\right) + \left[\sum_{t=1}^{n} (y_{t-1} - \pi_n X_t)\varepsilon_t\right] - (\widehat{\pi}_n - \pi_n)\left(\sum_{t=1}^{n} X_t \varepsilon_t\right), \\ 0 = &\left(\widehat{\beta}_n - \beta_n\right)\left[\sum_{t=1}^{n} X_t(y_{t-1} - \pi_n X_t)\right] - \left(\widehat{\beta}_n - \beta_n\right)(\widehat{\pi}_n - \pi_n)\left(\sum_{t=1}^{n} X_t^2\right) \\ &- \beta_n(\widehat{\pi}_n - \pi_n)\left(\sum_{t=1}^{n} X_t^2\right) + \left(\sum_{t=1}^{n} X_t \varepsilon_t\right). \end{aligned}$$

Therefore,

$$\widehat{\beta}_n - \beta_n = \frac{\beta_n \left(\widehat{\pi}_n - \pi_n\right) \left(\sum_{t=1}^n X_t^2\right) - \left(\sum_{t=1}^n X_t \varepsilon_t\right)}{\left[\sum_{t=1}^n X_t \left(y_{t-1} - \pi_n X_t\right)\right] - \left(\widehat{\pi}_n - \pi_n\right) \left(\sum_{t=1}^n X_t^2\right)},$$

and

$$\begin{aligned}
0 = {} & \beta_n \left(\widehat{\pi}_n - \pi_n\right) \left[\sum_{t=1}^n \left(y_{t-1} - \pi_n X_t\right)^2\right] \left(\sum_{t=1}^n X_t^2\right) \\
& - \left[\sum_{t=1}^n \left(y_{t-1} - \pi_n X_t\right)^2\right] \left(\sum_{t=1}^n X_t \varepsilon_t\right) \\
& - 2\beta_n \left(\widehat{\pi}_n - \pi_n\right)^2 \left[\sum_{t=1}^n X_t \left(y_{t-1} - \pi_n X_t\right)\right] \left(\sum_{t=1}^n X_t^2\right) \\
& + 2 \left(\widehat{\pi}_n - \pi_n\right) \left[\sum_{t=1}^n X_t \left(y_{t-1} - \pi_n X_t\right)\right] \left(\sum_{t=1}^n X_t \varepsilon_t\right) + \beta_n \left(\widehat{\pi}_n - \pi_n\right)^3 \left(\sum_{t=1}^n X_t^2\right)^2 \\
& - \left(\widehat{\pi}_n - \pi_n\right)^2 \left(\sum_{t=1}^n X_t^2\right) \left(\sum_{t=1}^n X_t \varepsilon_t\right) - \beta_n \left(\widehat{\pi}_n - \pi_n\right) \left[\sum_{t=1}^n X_t \left(y_{t-1} - \pi_n X_t\right)\right]^2 \\
& + \beta_n \left(\widehat{\pi}_n - \pi_n\right)^2 \left[\sum_{t=1}^n X_t \left(y_{t-1} - \pi_n X_t\right)\right] \left(\sum_{t=1}^n X_t^2\right) \\
& + \beta_n \left(\widehat{\pi}_n - \pi_n\right)^2 \left[\sum_{t=1}^n X_t \left(y_{t-1} - \pi_n X_t\right)\right] \left(\sum_{t=1}^n X_t^2\right) - \beta_n \left(\widehat{\pi}_n - \pi_n\right)^3 \left(\sum_{t=1}^n X_t^2\right)^2 \\
& + \left[\sum_{t=1}^n X_t \left(y_{t-1} - \pi_n X_t\right)\right] \left[\sum_{t=1}^n \left(y_{t-1} - \pi_n X_t\right) \varepsilon_t\right] \\
& - \left(\widehat{\pi}_n - \pi_n\right) \left[\sum_{t=1}^n \left(y_{t-1} - \pi_n X_t\right) \varepsilon_t\right] \left(\sum_{t=1}^n X_t^2\right) \\
& - \left(\widehat{\pi}_n - \pi_n\right) \left[\sum_{t=1}^n X_t \left(y_{t-1} - \pi_n X_t\right)\right] \left(\sum_{t=1}^n X_t \varepsilon_t\right) \\
& + \left(\widehat{\pi}_n - \pi_n\right)^2 \left(\sum_{t=1}^n X_t^2\right) \left(\sum_{t=1}^n X_t \varepsilon_t\right),
\end{aligned}$$

or

$$
\begin{aligned}
0 = (\widehat{\pi}_n - \pi_n) \Bigg\{ & \beta_n \left[\sum_{t=1}^{n} (y_{t-1} - \pi_n X_t)^2 \right] \left(\sum_{t=1}^{n} X_t^2 \right) - \beta_n \left[\sum_{t=1}^{n} X_t (y_{t-1} - \pi_n X_t) \right]^2 \\
& + \left[\sum_{t=1}^{n} X_t (y_{t-1} - \pi_n X_t) \right] \left(\sum_{t=1}^{n} X_t \varepsilon_t \right) - \left[\sum_{t=1}^{n} (y_{t-1} - \pi_n X_t) \varepsilon_t \right] \left(\sum_{t=1}^{n} X_t^2 \right) \Bigg\} \\
& - \left[\sum_{t=1}^{n} (y_{t-1} - \pi_n X_t)^2 \right] \left(\sum_{t=1}^{n} X_t \varepsilon_t \right) \\
& + \left[\sum_{t=1}^{n} X_t (y_{t-1} - \pi_n X_t) \right] \left[\sum_{t=1}^{n} (y_{t-1} - \pi_n X_t) \varepsilon_t \right].
\end{aligned}
$$

Therefore,

$$
\widehat{\pi}_n - \pi_n = \frac{\left\{ \begin{array}{c} \left[\sum_{t=1}^{n} (y_{t-1} - \pi_n X_t)^2 \right] \left(\sum_{t=1}^{n} X_t \varepsilon_t \right) \\ - \left[\sum_{t=1}^{n} X_t (y_{t-1} - \pi_n X_t) \right] \left[\sum_{t=1}^{n} (y_{t-1} - \pi_n X_t) \varepsilon_t \right] \end{array} \right\}}{\left\{ \begin{array}{c} \beta_n \left[\sum_{t=1}^{n} (y_{t-1} - \pi_n X_t)^2 \right] \left(\sum_{t=1}^{n} X_t^2 \right) \\ - \beta_n \left[\sum_{t=1}^{n} X_t (y_{t-1} - \pi_n X_t) \right]^2 \\ + \left[\sum_{t=1}^{n} X_t (y_{t-1} - \pi_n X_t) \right] \left(\sum_{t=1}^{n} X_t \varepsilon_t \right) \\ - \left[\sum_{t=1}^{n} (y_{t-1} - \pi_n X_t) \varepsilon_t \right] \left(\sum_{t=1}^{n} X_t^2 \right) \end{array} \right\}}.
$$

By the law of number for stationary ergodic sequences (White, 2001, Theorem 3.34, p. 44), the central limit theorem for martingale difference sequences (White, 2001, Theorem 5.24, p. 133),

$$
n^{-1/2} \sum_{t=1}^{n} X_t \varepsilon_t \xrightarrow{d} \sigma_\varepsilon \mathbf{M}_X^{1/2} \mathcal{Z} \sim \mathcal{N}\left(0, \sigma_\varepsilon^2 \mathbf{M}_X\right), \quad n^{-1} \sum_{t=1}^{n} X_t^2 \xrightarrow{p} \mathbf{M}_X.
$$

And by Lemma 7, let

$$n^{-1/2-h/2}\sum_{t=1}^{n}\left(y_{t-1}-X_t^\top\pi_n\right)\varepsilon_t\xrightarrow{d}(2b)^{-1/2}\sigma_\varepsilon^2\mathcal{Z}_1\sim\mathcal{N}\left(0,(2b)^{-1}\sigma_\varepsilon^4\right),$$
$$n^{-1/2-h}\sum_{t=1}^{n}X_t\left(y_{t-1}-X_t^\top\pi_n\right)\xrightarrow{d}b^{-1}\sigma_\varepsilon\mu_X\mathcal{Z}_2\sim\mathcal{N}\left(0,b^{-2}\sigma_\varepsilon^2\mu_X^2\right),$$

and

$$n^{-1-h}\sum_{t=1}^{n}\left(y_{t-1}-X_t^\top\pi_n\right)^2\xrightarrow{p}(2b)^{-1}\sigma_\varepsilon^2.$$

Therefore,

$$\begin{aligned}n^{1/2-h}\left(\widehat{\pi}_n-\pi_n\right)&=\frac{\left[n^{-1-h}\sum_{t=1}^n\left(y_{t-1}-\pi_nX_t\right)^2\right]\left(n^{-1/2}\sum_{t=1}^nX_t\varepsilon_t\right)+o_p(1)}{b\left[n^{-1-h}\sum_{t=1}^n\left(y_{t-1}-\pi_nX_t\right)^2\right]\left(n^{-1}\sum_{t=1}^nX_t^2\right)+o_p(1)}\\&=O_p(1).\end{aligned}$$

For $\widehat{\beta}_n$, since

$$\begin{aligned}\widehat{\beta}_n-\beta_n&=\frac{\beta_n\left(\widehat{\pi}_n-\pi_n\right)\left(\sum_{t=1}^nX_t^2\right)-\left(\sum_{t=1}^nX_t\varepsilon_t\right)}{\left[\sum_{t=1}^nX_t\left(y_{t-1}-\pi_nX_t\right)\right]-\left(\widehat{\pi}_n-\pi_n\right)\left(\sum_{t=1}^nX_t^2\right)}\\&=\frac{\left\{\begin{array}{c}-\beta_n\left[\sum_{t=1}^nX_t\left(y_{t-1}-\pi_nX_t\right)\right]\left[\sum_{t=1}^n\left(y_{t-1}-\pi_nX_t\right)\varepsilon_t\right]\left(\sum_{t=1}^nX_t^2\right)\\+\beta_n\left[\sum_{t=1}^nX_t\left(y_{t-1}-\pi_nX_t\right)\right]^2\left(\sum_{t=1}^nX_t\varepsilon_t\right)\\-\left[\sum_{t=1}^nX_t\left(y_{t-1}-\pi_nX_t\right)\right]\left(\sum_{t=1}^nX_t\varepsilon_t\right)^2\\+\left[\sum_{t=1}^n\left(y_{t-1}-\pi_nX_t\right)\varepsilon_t\right]\left(\sum_{t=1}^nX_t^2\right)\left(\sum_{t=1}^nX_t\varepsilon_t\right)\end{array}\right\}}{\left\{\begin{array}{c}\beta_n\left[\sum_{t=1}^n\left(y_{t-1}-\pi_nX_t\right)^2\right]\left[\sum_{t=1}^nX_t\left(y_{t-1}-\pi_nX_t\right)\right]\left(\sum_{t=1}^nX_t^2\right)\\-\beta_n\left[\sum_{t=1}^nX_t\left(y_{t-1}-\pi_nX_t\right)\right]^3\\+\left[\sum_{t=1}^nX_t\left(y_{t-1}-\pi_nX_t\right)\right]^2\left(\sum_{t=1}^nX_t\varepsilon_t\right)\\-\left[\sum_{t=1}^n\left(y_{t-1}-\pi_nX_t\right)^2\right]\left(\sum_{t=1}^nX_t^2\right)\left(\sum_{t=1}^nX_t\varepsilon_t\right)\end{array}\right\}}.\end{aligned}$$

Again, by the law of number for stationary ergodic sequences (White, 2001, Theorem

3.34, p. 44), the central limit theorem for martingale difference sequences (White, 2001, Theorem 5.24, p. 133), and Lemma 7,

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

Then we show that $R(\theta^*)=o_p\left(n^{-1}\right)$. By equation (2.11),

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

where θ^* is in between $\widehat{\theta}_n$ and θ_n. Therefore, $\beta^*=\left(\beta^*-\beta_n\right)+\beta_n=O_p\left(n^{-1/2-h/2}\right)+O\left(n^{-h}\right)$.

Since $\partial^3 Q_n\left(\theta^*\right)/\partial\beta^3=0$, and $\partial^3 Q_n\left(\theta^*\right)/\partial\pi^3=\mathbf{0}_{d_\pi^3\times 1}$,

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

in which

$$\frac{\partial^3 Q_n(\theta^*)}{\partial\beta^2\partial\pi} = -2n^{-1}\sum_{t=1}^{n}\left[y_{t-1} - X_t^\top\pi_n - X_t^\top(\pi^* - \pi_n)\right]X_t^\top,$$
$$\frac{\partial^3 Q_n(\theta^*)}{\partial\beta\partial\pi\partial\pi^\top} = 2\beta^* n^{-1}\sum_{t=1}^{n} X_t X_t^\top.$$

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

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

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

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

Let

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

Then by equation (2.11), Lemma 4, and the fact that $R(\theta^*) = o_p(n^{-1})$,

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

By definition (equation (2.2)), $\widehat{\theta}_n$ is the minimizer of $Q_n(\theta) - Q_n(\theta_n)$, and therefore $\Delta_n^*\left(\widehat{\theta}_n\right)$ is the minimizer of $q_n(\Delta_n^*(\theta))$, *i.e.*,

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

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

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

■

Proof. (Lemma 5)

1. $(n^{1/2} B^{-1}(h) D_\theta Q_n(\theta_n))$

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

By Lemma 8 and the central limit theorem for martingale difference sequences

(White, 2001, Theorem 5.24, p. 133),

$$\begin{aligned}
&n^{-1/2-h/2}\sum_{t=1}^{n}\left(y_{t-1}-X_t^\top \pi_n\right)\varepsilon_t \overset{A}{\sim} \mathcal{N}\left(0,(2b)^{-1}\sigma_\varepsilon^4\right), \\
&-\beta_n n^{-1/2+h}\sum_{t=1}^{n} X_t\varepsilon_t \overset{A}{\sim} \mathcal{N}\left(0,\sigma_\varepsilon^2 b^2 \mathbf{M}_X\right).
\end{aligned}$$

By Assumption 3, ε_t is independent to (y_{t-1}, X_t). Thus by Lemma 8,

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

Therefore

$$n^{1/2}B^{-1}(h)\,D_\theta Q_n(\theta_n)\overset{A}{\sim}\mathcal{N}\left(\mathbf{0}_{(d_\pi+1)\times 1},\begin{bmatrix}(2b)^{-1}\sigma_\varepsilon^4 & \mathbf{0}_{1\times d_\pi}\\ \mathbf{0}_{d_\pi\times 1} & \sigma_\varepsilon^2 b^2\mathbf{M}_X\end{bmatrix}\right).$$

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

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

By Lemma 8 and the law of large number for stationary ergodic sequences

(White, 2001, Theorem 3.34, p. 44),

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

and

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

Therefore,

$$B^{-1}(h) D_{\theta\theta^\top} Q_n(\theta_n) B^{-1}(h) \xrightarrow{p} \begin{bmatrix} (2b)^{-1} \sigma_\varepsilon^2 & \\ \mathbf{0}_{d_\pi \times 1} & b^2 \mathbf{M}_X \end{bmatrix}.$$

■

Proof. (Theorem 4) Since the results of Lemma 5 are the same as the ones of Lemma 4, the proof of Theorem 4 directly follows the proof of Theorem 3. ■

Proof. (Theorem 5) For 1. it suffices to show $\widehat{\sigma}_n^2 \xrightarrow{p} \sigma_\varepsilon^2$ and $\widehat{\mathbf{V}}_n \xrightarrow{p} \mathcal{V}_0(\gamma_n)$. By Theorem 1 and the law of number for martingale difference sequences (White, 2001, Theorem 3.76, p. 60),

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

And by Theorem 1 and the law of large number for stationary ergodic sequences (White, 2001, Theorem 3.34, p. 44),

$$\begin{aligned}
\widehat{\mathbf{V}}_n &= n^{-1}\sum_{t=1}^{n}\begin{bmatrix} \left(y_{t-1} - X_t^\top\widehat{\pi}_n\right)^2 & -\widehat{\beta}_n\left(y_{t-1} - X_t^\top\widehat{\pi}_n\right)X_t^\top \\ -\widehat{\beta}_n X_t\left(y_{t-1} - X_t^\top\widehat{\pi}_n\right) & \widehat{\beta}_n^2 X_t X_t^\top \end{bmatrix} \\
&= n^{-1}\sum_{t=1}^{n}\begin{bmatrix} \left(y_{t-1} - X_t^\top\pi_0\right)^2 & \\ -\beta_0 X_t\left(y_{t-1} - X_t^\top\pi_0\right) & \beta_0^2 X_t X_t^\top \end{bmatrix} + O_p\left(n^{-1/2}\right) \xrightarrow{p} \mathcal{V}_0\left(\gamma_n\right).
\end{aligned}$$

For 2. by equations (3.4) and (2.12),

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

By Lemma 6, the law of large number for stationary ergodic sequences (White, 2001, Theorem 3.34, p. 44) and Theorem 2:

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

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

and

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

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 law of number for martingale difference sequences (White, 2001, Theorem 3.76, p. 60),

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

For $B^{-1}(h)\widehat{\mathbf{V}}_n B^{-1}(h)$, by equations (3.4) and (2.12),

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

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),

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

And the remains directly follow by Theorem 3.

For 4., it directly follows the proof of 3. and Theorem 4. ■

Proof. (Theorem 6) We first prove 2., *i.e.*, $AsySz\left(CS_n^{ICS,LF}(\mathbf{R}\theta_n;1-\alpha,\varphi_0)\right) = 1-\alpha$. Notice that $CS_n^{ICS,LF}(\mathbf{R}\theta_n;1-\alpha,\varphi_0)$ can be written as

$$CS_n^{ICS,LF}(\mathbf{R}\theta_n;1-\alpha,\varphi_0) = \{\upsilon : W_n(\upsilon)\le c\},$$

where

$$c = \begin{cases} c_L = \sup_{\{b_\upsilon,\mathbf{c}_\upsilon\}\in\mathcal{H}(\mathbf{R},\upsilon)}\xi_{1-\alpha}\left(\mathcal{W}(b_\upsilon,\mathbf{c}_\upsilon;\varphi_0)\right), & \text{if } A_n\le k_n, \\ c_D = \chi^2_{d_r,1-\alpha}, & \text{if } A_n>k_n. \end{cases}.$$

First we show that $c\xrightarrow{p}c_L$ when $\gamma_n\in\Gamma(1,b,\mathbf{c})$ and $c\xrightarrow{p}c_D$ when $\theta_n=\theta_0\in\Theta^*$,

$\gamma_n \in \Gamma(h,b)$ or $\gamma_n \in \Gamma(h,b,\mathbf{c})$. Since $k_n = c_k \log(n)$, it suffices to show $A_n = \sqrt{n}\widehat{\beta}_n \Big/ \sqrt{\widehat{\text{Avar}}\left(\widehat{\beta}_n\right)} = O_p(1)$ when $\gamma_n \in \Gamma(1,b,\mathbf{c})$, and $k_n^{-1}A_n \xrightarrow{p} \infty$ when $\theta_n = \theta_0 \in \Theta^*$, $\gamma_n \in \Gamma(h,b)$ or $\gamma_n \in \Gamma(h,b,\mathbf{c})$.

By Theorems 2 and 5, when $\gamma_n \in \Gamma(1,b,\mathbf{c})$,

$$A_n = \frac{n\widehat{\beta}_n}{\sqrt{n}\sqrt{\widehat{\text{Avar}}\left(\widehat{\beta}_n\right)}} \Rightarrow \frac{\widehat{\lambda}_\beta\left(\widehat{\kappa}_\pi(b,\mathbf{c};\varphi_0),b,\mathbf{c};\varphi_0\right)}{\left[\sigma_\varepsilon^2\left[1,\mathbf{0}_{1\times d_\pi}\right]\mathcal{V}_1^{-1}(b,\mathbf{c};\varphi_0)\left[1,\mathbf{0}_{1\times d_\pi}\right]^\top\right]^{1/2}} = O_p(1).$$

And when $\theta_n = \theta_0 \in \Theta^*$, $\gamma_n \in \Gamma(h,b)$ or $\gamma_n \in \Gamma(h,b,\mathbf{c})$, by Theorem 5, $\beta_n = n^{-h}b$, and $h < 1$,

$$\begin{aligned} k_n^{-1}A_n &= k_n^{-1}\left(\frac{\sqrt{n}\left(\widehat{\beta}_n - \beta_n\right)}{\sqrt{\widehat{\text{Avar}}\left(\widehat{\beta}_n\right)}} + \frac{n^{1/2-h/2}b}{n^{h/2}\sqrt{\widehat{\text{Avar}}\left(\widehat{\beta}_n\right)}}\right) \\ &= k_n^{-1}\left(O_p(1) + O_p\left(n^{1/2-h/2}\right)\right) \xrightarrow{p} \infty. \end{aligned}$$

Then it suffices to verify the Assumption ACP in Andrews and Cheng (2012). Let the coverage probability $CP_n^{ICS,LF} = \mathrm{P}\left(W_n(\upsilon_n) \le c\right)$, where $\upsilon_n = \mathbf{R}\theta_n$ denotes the true value of $\mathbf{R}\theta_n$. We would like to show

(i). For any $\gamma_n \in \Gamma(1,b,\mathbf{c})$, $CP_n \to CP_{LF}(b,\mathbf{c};\varphi_0)$ for some $CP_{LF}(b,\mathbf{c};\varphi_0) \in [0,1]$.

(ii). For any $\theta_n = \theta_0 \in \Theta^*$, $\gamma_n \in \Gamma(h,b)$ or $\gamma_n \in \Gamma(h,b,\mathbf{c})$, $\liminf_{n\to\infty} CP_n \ge CP_\infty$ for some $CP_\infty \in [0,1]$.

(iii). For some $\theta_n = \theta_0 \in \Theta^*$, $\gamma_n \in \Gamma(h,b)$ or $\gamma_n \in \Gamma(h,b,\mathbf{c})$, $CP_n \to CP_\infty$.

(iv). For some $\delta_1 > 0$ and $\delta_2 > 0$, $\gamma = \{\beta,\pi,\varphi\} \in \Gamma = \Theta^* \times \Phi$ with $\beta < \delta_1$ and $\|\pi\| < \delta_2$ implies $\widetilde{\gamma} = \left\{\widetilde{\beta},\widetilde{\pi},\varphi\right\} \in \Gamma$ with $\widetilde{\beta} < \delta_1$ and $\|\widetilde{\pi}\| < \delta_2$.

For (i), by Theorem 5, when $\gamma_n \in \Gamma(1, b, \mathbf{c})$,

$$\begin{aligned}
CP_n &= \mathrm{P}\left(W_n\left(\upsilon_n\right) \leq c_L\right) \\
&\to \mathrm{P}\left(\mathcal{W}\left(b, \mathbf{c}; \varphi_0\right) \leq \sup_{\{b_\upsilon, \mathbf{c}_\upsilon\} \in \mathcal{H}(\mathbf{R}, \upsilon)} \xi_{1-\alpha}\left(\mathcal{W}\left(b_\upsilon, \mathbf{c}_\upsilon; \varphi_0\right)\right)\right) \\
&:= CP_{LF}\left(b, \mathbf{c}; \varphi_0\right) \in [0, 1].
\end{aligned}$$

Specifically, by construction,

$$\inf_{\{b, \mathbf{c}\} \in \mathcal{H}(\mathbf{R}, \upsilon)} CP_{LF}\left(b, \mathbf{c}; \varphi_0\right) = 1 - \alpha.$$

For (ii) and (iii), by Theorem 5, when $\theta_n = \theta_0 \in \Theta^*$, $\gamma_n \in \Gamma(h, b)$ or $\gamma_n \in \Gamma(h, b, \mathbf{c})$,

$$CP_n = \mathrm{P}\left(W_n\left(\upsilon_n\right) \leq c_D\right) \to F_{\chi^2(d_r)}\left(\chi^2_{d_r, 1-\alpha}\right) = 1 - \alpha := CP_\infty,$$

where $F_{\chi^2(d_r)}$ denotes the *cdf* of $\chi^2(d_r)$ distribution. And (iv) follows by the convexity of Θ^*. Therefore, by Lemma 2.1 of Andrews and Cheng (2012),

$$\begin{aligned}
& AsySz\left(CS_n^{ICS,LF}\left(\mathbf{R}\theta_n; 1-\alpha, \varphi_0\right)\right) \\
&= \liminf_{n\to\infty} \inf_{\gamma_n \in \Gamma_n} \mathrm{P}\left(\mathbf{R}\theta_n \in CS_n^{ICS,LF}\left(\mathbf{R}\theta_n; 1-\alpha, \varphi_0\right)\right) \\
&= \min\left(\inf_{\{b, \mathbf{c}\} \in \mathcal{H}(\mathbf{R}, \upsilon)} CP_{LF}\left(b, \mathbf{c}; \varphi_0\right), CP_\infty\right) = 1 - \alpha.
\end{aligned}$$

1. directly follows 2 since 1. is a special case of 2. For 3. and 4. since both the projection-based method and the Bonferroni-based method are conservative, when $\gamma_n \in \Gamma(1, b, \mathbf{c})$, the corresponding coverage probabilities $CP_P\left(b, \mathbf{c}; \varphi_0\right)$ and

$CP_B\left(b,\mathbf{c};\varphi_0\right)$ are greater than or equal to $1-\alpha$. Therefore,

$$\min\left(\inf_{\{b,\mathbf{c}\}\in\mathcal{H}(\mathbf{R},\upsilon)} CP_P\left(b,\mathbf{c};\varphi_0\right), CP_\infty\right) \geq 1-\alpha,$$
$$\min\left(\inf_{\{b,\mathbf{c}\}\in\mathcal{H}(\mathbf{R},\upsilon)} CP_B\left(b,\mathbf{c};\varphi_0\right), CP_\infty\right) \geq 1-\alpha.$$

■

APPENDIX C

SUPPLEMENTARY RESULTS AND PROOFS

This appendix 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 $\gamma_n \in \Gamma\left(1, b, \mathbf{c}\right)$. Let $\mathcal{W}_\varepsilon\left(\cdot\right)$ and $\mathcal{W}_X\left(\cdot\right)$ be two standard Wiener processes (one-dimensional and d_π-dimensional, respectively), and $\mathcal{J}_{b,\varepsilon}\left(\cdot\right)$ and $\mathcal{J}_{b,X}\left(\cdot\right)$ be an Ornstein–Uhlenbeck process. For any $r \in [0,1]$, when $n \to \infty$,*

$$n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor}\varepsilon_t \Rightarrow \sigma_\varepsilon \mathcal{W}_\varepsilon\left(r\right), \quad n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor}\left(X_{t-i}-\mu_X\right)^\top \Rightarrow \mathbf{\Sigma}_X^{1/2}\mathcal{W}_X\left(r\right),$$

$$\mathcal{J}_{b,\varepsilon}\left(r\right) = \int_0^r \exp\left(-b\left(r-s\right)\right)d\mathcal{W}_\varepsilon\left(s\right) \quad \text{and}$$

$$\mathcal{J}_{b,X}\left(r\right) = \int_0^r \exp\left(-b\left(r-s\right)\right)d\mathcal{W}_X\left(s\right)$$

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

1. $n^{-1/2}y_{\lfloor nr\rfloor} \Rightarrow \sigma_\varepsilon\mathcal{J}_{b,\varepsilon}\left(r\right) + \mathbf{c}^\top\mu_X\left(1-\exp\left(-br\right)\right)$.

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

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

4. $n^{-1}\sum_{t=1}^n y_{t-1}\varepsilon_t \Rightarrow \sigma_\varepsilon^2\int_0^1 \mathcal{J}_{b,\varepsilon}\left(r\right)d\mathcal{W}_\varepsilon\left(r\right)$
   $+\sigma_\varepsilon\mathbf{c}^\top\mu_X\left(\int_0^1\left(1-\exp\left(-br\right)\right)d\mathcal{W}_\varepsilon\left(r\right)\right)$.

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

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

Proof. For 1., under Assumption 1, equation (2.1) can be written as:

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

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 $n \to \infty$,

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

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

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

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

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

■

Lemma 7 *Suppose that Assumptions 1, 2 and 3 hold and $\gamma_n \in \Gamma(h, b, \mathbf{c})$. Then as $n \to \infty$:*

1. $n^{-1/2-h/2}\sum_{t=1}^{n} \left(y_{t-1} - X_t^\top \pi_n\right) \varepsilon_t \overset{A}{\sim} \mathcal{N}\left(0, (2b)^{-1}\sigma_\varepsilon^4\right)$.

2. $n^{-1-h}\sum_{t=1}^{n}\left(y_{t-1}-X_t^\top\pi_n\right)^2 \xrightarrow{p} (2b)^{-1}\sigma_\varepsilon^2$.

3. $n^{-1/2-h}\sum_{t=1}^{n}X_t\left(y_{t-1}-X_t^\top\pi_n\right) \overset{A}{\sim} \mathcal{N}\left(0, b^{-2}\sigma_\varepsilon^2\mu_X\mu_X^\top\right)$.

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 (2.1) can be written as:

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

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

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

Therefore,

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

$$
\begin{aligned}
&n^{-1-h}\sum_{t=1}^{n}\left(y_{t-1}-X_t^{\top}\pi_n\right)^2 = n^{-1-h}\sum_{t=1}^{n}\left(y_{t-1}-\mu_X^{\top}\pi_n+\mu_X^{\top}\pi_n-X_t^{\top}\pi_n\right)^2 \\
&= n^{-1-h}\sum_{t=1}^{n}\left[\eta_{t-1}+n^{-1/2}b\mathbf{c}^{\top}\xi_{t-1}-n^{-1/2+h}\mathbf{c}^{\top}\left(X_t-\mu_X\right)\right]^2 \\
&= n^{-1-h}\sum_{t=1}^{n}\eta_{t-1}^2 + o_p(1) \xrightarrow{p} \frac{\sigma_\varepsilon^2}{2b},
\end{aligned}
$$

and

$$
\begin{aligned}
&n^{-1/2-h}\sum_{t=1}^{n}X_t\left(y_{t-1}-X_t^{\top}\pi_n\right)\\
&= n^{-1/2-h}\mu_X\sum_{t=1}^{n}\left(y_{t-1}-\mu_X^{\top}\pi_n\right)-n^{-1/2-h}\sum_{t=1}^{n}\left(X_t-\mu_X\right)\left(y_{t-1}-\mu_X^{\top}\pi_n\right)\\
&- n^{-1/2-h}\mu_X\pi_n^{\top}\sum_{t=1}^{n}\left(X_t-\mu_X\right)+n^{-1/2-h}\sum_{t=1}^{n}\left(X_t-\mu_X\right)\left(X_t-\mu_X\right)^{\top}\pi_n\\
&= n^{-1/2-h}\mu_X\sum_{t=1}^{n}\eta_{t-1}+n^{-1-h}b\mathbf{c}^{\top}\mu_X\sum_{t=1}^{n}\xi_{t-1}\\
&- n^{-1/2-h}\sum_{t=1}^{n}\left(X_t-\mu_X\right)\eta_{t-1}-n^{-1-h}b\mathbf{c}^{\top}\sum_{t=1}^{n}\left(X_t-\mu_X\right)\xi_{t-1}\\
&- n^{-1}\mu_X\mathbf{c}^{\top}\sum_{t=1}^{n}\left(X_t-\mu_X\right)+n^{-1}\sum_{t=1}^{n}\left(X_t-\mu_X\right)\left(X_t-\mu_X\right)^{\top}\mathbf{c}\\
&= n^{-1/2-h}\mu_X\sum_{t=1}^{n}\eta_{t-1}+o_p\left(1\right)\overset{A}{\sim}\mathcal{N}\left(0,b^{-2}\sigma_{\varepsilon}^2\mu_X\mu_X^{\top}\right).
\end{aligned}
$$

■

Lemma 8 *Suppose that Assumptions 1, 2 and 3 hold and* $\gamma_n\in\Gamma\left(h,b\right)$. *Then as* $n\to\infty$*:*

1. $n^{-1/2-h/2}\sum_{t=1}^{n}\left(y_{t-1}-X_t^{\top}\pi_n\right)\varepsilon_t\overset{A}{\sim}\mathcal{N}\left(0,\left(2b\right)^{-1}\sigma_{\varepsilon}^4\right)$.

2. $n^{-1-h}\sum_{t=1}^{n}\left(y_{t-1}-X_t^{\top}\pi_n\right)^2\xrightarrow{p}\left(2b\right)^{-1}\sigma_{\varepsilon}^2$.

3. $n^{-1}\sum_{t=1}^{n}X_t\left(y_{t-1}-X_t^{\top}\pi_n\right)\xrightarrow{p}\mathbf{\Sigma}_X\pi_n$.

Proof. By the proof of Lemma 7, $y_t = \mu_X^\top \pi_n + \eta_t + \beta_n \xi_t \pi_n$,

$$\begin{aligned}
& n^{-1/2-h/2} \sum_{t=1}^{n} \left(y_{t-1} - X_t^\top \pi_n \right) \varepsilon_t \\
& = n^{-1/2-h/2} \sum_{t=1}^{n} \left(y_{t-1} - \mu_X^\top \pi_n \right) \varepsilon_t - n^{-1/2-h/2} \pi_n^\top \sum_{t=1}^{n} \left(X_t - \mu_X \right) \varepsilon_t \\
& = n^{-1/2-h/2} \sum_{t=1}^{n} \eta_{t-1} \varepsilon_t + n^{-1/2-3h/2} b \pi_n^\top \sum_{t=1}^{n} \xi_{t-1} \varepsilon_t - n^{-1/2-h/2} \pi_n^\top \sum_{t=1}^{n} \left(X_t - \mu_X \right) \varepsilon_t \\
& = n^{-1/2-h/2} \sum_{t=1}^{n} \eta_{t-1} \varepsilon_t + o_p(1) \overset{A}{\sim} \mathcal{N}\left(0, \frac{\sigma_\varepsilon^4}{2b} \right),
\end{aligned}$$

$$\begin{aligned}
& n^{-1-h} \sum_{t=1}^{n} \left(y_{t-1} - X_t^\top \pi_n \right)^2 = n^{-1-h} \sum_{t=1}^{n} \left(y_{t-1} - \mu_X^\top \pi_n + \mu_X^\top \pi_n - X_t^\top \pi_n \right)^2 \\
& = n^{-1-h} \sum_{t=1}^{n} \left[\eta_{t-1} + n^{-h} b \pi_n^\top \xi_{t-1} - \pi_n^\top \left(X_t - \mu_X \right) \right]^2 \\
& = n^{-1-h} \sum_{t=1}^{n} \eta_{t-1}^2 + o_p(1) \xrightarrow{p} \frac{\sigma_\varepsilon^2}{2b},
\end{aligned}$$

and

$$
\begin{aligned}
&n^{-1}\sum_{t=1}^{n} X_t\left(y_{t-1}-X_t^{\top}\pi_n\right)\\
&=n^{-1}\mu_X\sum_{t=1}^{n}\left(y_{t-1}-\mu_X^{\top}\pi_n\right)-n^{-1}\sum_{t=1}^{n}\left(X_t-\mu_X\right)\left(y_{t-1}-\mu_X^{\top}\pi_n\right)\\
&-n^{-1}\mu_X\pi_n^{\top}\sum_{t=1}^{n}\left(X_t-\mu_X\right)+n^{-1}\sum_{t=1}^{n}\left(X_t-\mu_X\right)\left(X_t-\mu_X\right)^{\top}\pi_n\\
&=n^{-1}\mu_X\sum_{t=1}^{n}\eta_{t-1}+n^{-1-h}b\pi_n^{\top}\mu_X\sum_{t=1}^{n}\xi_{t-1}\\
&-n^{-1}\sum_{t=1}^{n}\left(X_t-\mu_X\right)\eta_{t-1}-n^{-1-h}b\pi_n^{\top}\sum_{t=1}^{n}\left(X_t-\mu_X\right)\xi_{t-1}\\
&-n^{-1}\mu_X\pi_n^{\top}\sum_{t=1}^{n}\left(X_t-\mu_X\right)+n^{-1}\sum_{t=1}^{n}\left(X_t-\mu_X\right)\left(X_t-\mu_X\right)^{\top}\pi_n\\
&=n^{-1}\sum_{t=1}^{n}\left(X_t-\mu_X\right)\left(X_t-\mu_X\right)^{\top}\pi_n+o_p\left(1\right)\xrightarrow{p}\mathbf{\Sigma}_X\pi_n.
\end{aligned}
$$

■