Testing for Regime Switching in State Space Models*

Fan Zhuo†

Boston University

November 10, 2015

Abstract

This paper develops a modified likelihood ratio (MLR) test for detecting regime switching in state space models. I apply the filtering algorithm introduced in Gordon and Smith (1988) to construct a modified likelihood function under the alternative hypothesis of two regimes and extend the analysis in Qu and Zhuo (2015) to establish the asymptotic distribution of the MLR statistic under the null hypothesis. I also present a practical application of the test using U.S. unemployment rates. This paper is the first to develop a test for detecting regime switching in state space models that is based on the likelihood ratio principle.

Keywords: Hypothesis testing, likelihood ratio, state space model, Markov switching.

*I am deeply indebted to my advisor Zhongjun Qu for his invaluable advice and support for my research. I also wish to thank Pierre Perron, Hiroaki Kaido, Ivan Fernandez-Val, and seminar participants at Boston University for valuable suggestions.

†Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (zhuo@bu.edu).

1 Introduction

Economists have long recognized the possibility that model parameters may not be constant through time, and that instead there can be variations in model structure. If these variations are temporary and recurrent, then the Markov regime switching model can offer a natural modeling choice. Hamilton (1989) makes a seminal contribution that not only introduces a framework with Markov regime switching for describing economic growth, but also provides a general algorithm for filtering, smoothing, and maximum likelihood estimation. A survey of the literature on regime switching models can be found in Hamilton (2008). Meanwhile, state space models are widely used in economics and finance to study time series with latent state variables. Harvey (1981) and Meinhold and Singpurwalla (1983) introduced economists to the use of the Kalman (1960) filter for constructing likelihood functions through the prediction error decomposition. Latent state variables and regime switching can arise at the same time, which poses a challenge for modeling them jointly.

However, as Hamilton (1990) pointed out, conducting formal tests for the presence of Markov switching is challenging. There are generally three approaches for detecting regime switching. The first approach tests for parameter homogeneity versus heterogeneity. Early contributions include Neyman and Scott (1996), Chesher (1984), Lancaster (1984) and Davidson and MacKinnon (1991). Recently, Carrasco, Hu and Ploberger (2014) further developed this approach and proposed a class of optimal tests for the constancy of parameters in random coefficients models where the parameters are weakly dependent under the alternative hypothesis. The second approach, from Hamilton (1996), offers a series of specification tests of regime switching in time series models. These tests only need researchers to estimate the model under the null hypothesis and have powers against a wide range of alternative models. However, their powers can be lower than what is achievable if the parameters indeed follow a finite state Markov chain. The third approach is based on the (quasi) likelihood ratio principle. Several important advances have been made by Hansen (1992), Garcia (1998), Cho and White (2007), and Carter and Steigerwald (2012). Qu and Zhuo (2015) is a recent development, which analyzes likelihood ratio based tests for Markov regime switching allowing for multiple switching parameters. The purpose of the present paper is to detect the regime switching in state space models.

The likelihood function for a state space model with regime switching is hard to construct, as discussed in Kim and Nelson (1999). Different approximations to the likelihood function have been considered in the literature, such as in Gordon and Smith (1988) and Highfield (1990). This paper uses the approximation applied in Gordon and Smith (1988). Based on this approximation, I develop a modified likelihood ratio (MLR) test.

I extend the techniques developed in Qu and Zhuo (2015) to handle the nonstandard features associated with the MLR test. These nonstandard features include the following: (1) Some nuisance parameters are unidentified under the null hypothesis, which violates the standard conditions that yield the chi-squared asymptotic distribution for the test statistic. This gives rise to the Davies (1977) problem. (2) The null hypothesis yields a local optimum (c.f. Hamilton, 1990), making the score function identically zero at the null parameter estimates. Consequently, a second order Taylor approximation of the likelihood ratio is insufficient to study its asymptotic properties. (3) Conditional regime probabilities follow stochastic processes that can only be constructed recursively. Moreover, this paper tackles an additional difficulty introduced by the latent state variables when expanding the MLR. The asymptotic distribution of the MLR test statistic is analyzed in five steps.

1. I describe the algorithm used to construct the modified likelihood function for Markov switching state space models introduced in Gordon and Smith (1988).

2. I characterize the conditional regime probability, the filtered latent state, the mean squared error of the filtered latent state, and their high order derivatives with respect to the model parameters.

3. I first fix p and q and derive a fourth order Taylor approximation to the MLR. Then, I view the MLR as an empirical process indexed by p and q, and derive its asymptotic distribution.

4. While the above limiting distributions are adequate for a broad class of models, they can lead to over-rejections in some situations specified later. To resolve the issue of over-rejection, the higher order terms in the likelihood expansion are incorporated into the asymptotic distribution to safe guard against their effects.

5. I apply a unified algorithm proposed in Qu and Zhuo (2015) to simulate the above refined asymptotic distribution.

Three Monte Carlo experiments are conducted to examine the MLR statistic. The first experiment checks the improvement introduced by the refined asymptotic distribution. The second and third experiments check the size and power of the MLR statistic. I also apply my method to study changes in U.S. unemployment rates and find strong evidence favoring the regime switching specification.

This paper is the first to develop a likelihood ratio based test for detecting regime switching in general state space models and contributes to the literature in several ways. First, I demonstrate the construction of the modified likelihood function under a two regimes specification for general state space models. Next, I study the Taylor expansion of the MLR when some regularity conditions fail to hold. Finally, I apply my method to an empirical example and find the comovement between the U.S. business cycle and changes in monthly U.S. unemployment rates.

The paper is structured as follows. In Section 2, I provide the general model, the basic filter, and the hypotheses. Section 3 introduces the test statistic. Section 4 studies asymptotic properties of the MLR for prespecified p and q. Section 5 provides the limiting distribution of the MLR test statistic and introduces a finite sample refinement. Section 6 examines the finite sample properties of the test statistic. Section 7 considers an empirical application to the U.S. unemployment rate. Section 8 concludes. All proofs are collected in the appendix.

The following notation is used. $||x||$ is the Euclidean norm of a vector x. $||X||$ is the vector induced norm for a matrix X. $x^{\otimes k}$ and $X^{\otimes k}$ denote the k-fold Kronecker product of x and X, respectively. The expression $\text{vec}(A)$ stands for the vectorization of a k dimensional array A. For example, for a three dimensional array A with n elements along each dimension, $\text{vec}(A)$ returns a n^3-vector whose $(i+(j-1)n+(k-1)n^2)$-th element equals $A(i,j,k)$. $\mathbf{1}_{\{\cdot\}}$ is the indicator function. For a scalar valued function $f(\theta)$, let $\theta \in R^p$, $\nabla_\theta f(\theta_0)$ denotes a $p \times 1$ vector of partial derivatives with respect to θ and evaluated at θ_0. $\nabla_{\theta'} f(\theta_0)$ equals the transpose of $\nabla_\theta f(\theta_0)$ and $\nabla_{\theta_j} f(\theta_0)$ denotes its j-th element. For a matrix function $P(\theta)$, $\nabla_{\theta_j} P(\theta)$ denotes the derivative of $P(\theta)$ with respect to the j-th element in θ. The symbols "$\Rightarrow$", "$\rightarrow^d$" and "$\rightarrow^p$" denote weak convergence under the Skorohod topology, convergence in distribution and in probability, respectively. $O_p(\cdot)$ and $o_p(\cdot)$ are the usual notations for the orders of stochastic magnitude.

2 Model and hypotheses

This section presents the model and hypotheses. The discussion consists of the following: the model, the log likelihood function under the null hypothesis (i.e., one regime), the modified log likelihood function under the alternative hypothesis (i.e., two regimes), and some assumptions related to these three aspects.

2.1 The model

Consider the following state space representation of a dynamic linear model with switching in both transition and measurement equations:

$$x_t = G_{s_t} + F_{s_t} x_{t-1} + u_t, \tag{2.1}$$

$$y_t = H'_{s_t} x_t + A'_{s_t} z_t, \tag{2.2}$$

$$u_t \sim N\left(0, Q_{s_t}\right). \tag{2.3}$$

The transition equation (2.1) describes the dynamics of the unobserved state vector x_t as a function of a $J \times 1$ vector of shocks u_t and x_{t-1}. The measurement equation (2.2) describes the evolution of an observed scalar time series as a function of x_t and a $K \times 1$ vector of weakly exogenous variables z_t. The measurement error, normally included in (2.2), is treated as a latent variable in x_t. F_{s_t} is of dimension $J \times J$, G_{s_t} is of dimension $J \times 1$, H_{s_t} is of dimension $J \times 1$, and A_{s_t} is of dimension $K \times 1$. Q_{s_t} is a positive semidefinite symmetric matrix of dimension $J \times J$.

The subscripts in F_{s_t}, G_{s_t}, H_{s_t}, A_{s_t}, and Q_{s_t} imply that some of the parameters in these matrices are dependent on an unobserved binary variable s_t whose value determines the regime at time t. The regimes are Markovian, i.e., $p(s_t = 1|s_{t-1} = 1) = p$ and $p(s_t = 2|s_{t-1} = 2) = q$. The resulting stationary (or invariant) probability for $s_t = 1$ is given by

$$\xi_*(p, q) = \frac{1-q}{2-p-q}. \tag{2.4}$$

In the subsequent analysis, $\xi_*(p, q)$ is abbreviated as ξ_*. Because this paper seeks to test regime switching in state space models based on the likelihood ratio principle, subsections 2.2-2.3 will focus on constructing the likelihood function under the two regimes specification, and the likelihood function under one regime specification is a by-product of the standard Kalman filter.

2.2 Modified Kalman filter

When constructing the likelihood function for a general state space model with regime switching, each iteration of the Kalman filter produces a two-fold increase in the number of cases to consider under a two regimes specification, as noted by Gordon and Smith (1988) and Harrison and Stevens (1976). This means there can be more than 1000 components in the likelihood function for a sample of size

$T = 10$. This makes studying this likelihood function and its expansion infeasible. Therefore, an approximation is considered here to "collapse" the filtered states when $s_t = 1$ and $s_t = 2$ to a single filtered state at each t, as in Gordon and Smith (1988).

Define the information set at time $t-1$ as

$$\Omega_{t-1} = \sigma\text{-field}\{..., z'_{t-1}, y_{t-2}, z'_t, y_{t-1}\}.$$

Suppose the model parameters are known. The modified Kalman filter algorithm, conditional on $s_t = i$, is given by:

$$x^{(i)}_{t|t-1} := G_i + F_i x_{t-1|t-1}, \tag{2.5}$$

$$P^{(i)}_{t|t-1} := F_i P_{t-1|t-1} F'_i + Q_i, \tag{2.6}$$

$$\mu^{(i)}_{t|t-1} := y_t - H'_i x^{(i)}_{t|t-1} - A'_i z_t, \tag{2.7}$$

$$C^{(i)}_{t|t-1} := H'_i P^{(i)}_{t|t-1} H_i, \tag{2.8}$$

$$x^{(i)}_{t|t} := x^{(i)}_{t|t-1} + P^{(i)}_{t|t-1} H_i [C^{(i)}_{t|t-1}]^{-1} \mu^{(i)}_{t|t-1}, \tag{2.9}$$

$$P^{(i)}_{t|t} := (I - P^{(i)}_{t|t-1} H_i [C^{(i)}_{t|t-1}]^{-1} H'_i) P^{(i)}_{t|t-1}, \tag{2.10}$$

where $x_{t-1|t-1}$ is an estimate of x_{t-1} based on information up to time $t-1$; $x^{(i)}_{t|t-1}$ is an estimate of x_t based on information up to time $t-1$ given $s_t = i$; $P^{(i)}_{t|t-1}$ is an estimate of the mean squared error of $x^{(i)}_{t|t-1}$; $\mu^{(i)}_{t|t-1}$ estimates the conditional forecast error of y_t based on information up to time $t-1$ given $s_t = i$; and $C^{(i)}_{t|t-1}$ estimates the conditional variance of the forecast error $\mu^{(i)}_{t|t-1}$.

Let $\xi_{t|t}$ be an estimate of $Pr(s_t = 1|\Omega_t)$. The "collapse" step combines the two filtered states $x^{(1)}_{t|t}$ and $x^{(2)}_{t|t}$ into a single estimate of x_t based on Ω_t by

$$x_{t|t} := \xi_{t|t} x^{(1)}_{t|t} + (1 - \xi_{t|t}) x^{(2)}_{t|t}. \tag{2.11}$$

Then, the mean squared error of $x_{t|t}$ can be computed as:

$$P_{t|t} := \xi_{t|t} P^{(1)}_{t|t} + (1 - \xi_{t|t}) P^{(2)}_{t|t} + \xi_{t|t}(1 - \xi_{t|t}) \left(x^{(1)}_{t|t} - x^{(2)}_{t|t}\right) \left(x^{(1)}_{t|t} - x^{(2)}_{t|t}\right)'. \tag{2.12}$$

At the end of each iteration, equations (2.11) and (2.12) are employed to collapse the two filtered states into one filtered state $x_{t|t}$ and calculate the mean squared error of $x_{t|t}$, i.e. $P_{t|t}$.

2.3 Modified Markov switching filter

To complete the modified Kalman filter, we need to calculate $\xi_{t|t}$ for $t = 1, 2, ..., T$. The calculation of $\xi_{t|t}$ is based on the Markov regime switching filter introduced in Hamilton (1989) and conducted in three steps.

1. At the beginning of the t-th iteration, given $\xi_{t-1|t-1}$, we have

$$\xi_{t|t-1} := p\xi_{t-1|t-1} + (1-q)(1-\xi_{t-1|t-1}). \tag{2.13}$$

2. An estimate of the density of y_t is obtained by

$$f(y_t|\Omega_{t-1}) := \xi_{t|t-1} f(y_t|s_t = 1, \Omega_{t-1}) + (1-\xi_{t|t-1}) f(y_t|s_t = 2, \Omega_{t-1}),$$

where the conditional density satisfies

$$f(y_t|s_t = i, \Omega_{t-1}) := \left[2\pi C^{(i)}_{t|t-1}\right]^{-1/2} \exp\left\{ -\frac{\left[\mu^{(i)}_{t|t-1}\right]^2}{2C^{(i)}_{t|t-1}} \right\}, \ (i = 1, 2) \tag{2.14}$$

where $\mu^{(i)}_{t|t-1}$ and $C^{(i)}_{t|t-1}$ are given in (2.7) and (2.8).

3. Once y_t is observed, we can update the modified conditional regime probability

$$\xi_{t|t} := \frac{\left[p\xi_{t-1|t-1} + (1-q)(1-\xi_{t-1|t-1})\right] f(y_t|s_t = 1, \Omega_{t-1})}{f(y_t|s_t = 2, \Omega_{t-1}) + \left[p\xi_{t-1|t-1} + (1-q)(1-\xi_{t-1|t-1})\right] \left[f(y_t|s_t = 1, \Omega_{t-1}) - f(y_t|s_t = 2, \Omega_{t-1})\right]}. \tag{2.15}$$

Figure 1 presents a flowchart for the filter described in subsections 2.2-2.3. The modified log likelihood function under the two regimes specification, i.e. $\sum_{t=1}^{T} \log\left[f(y_t|\Omega_{t-1})\right]$, is given as a by-product of the filter. The initial values $\xi_{0|0}$, $x_{0|0}$ and $P_{0|0}$ will be discussed in section 4.

2.4 Hypotheses

Let δ represent parameters that are affected by regime switching, taking a value of δ_1 in regime 1 and δ_2 in regime 2. Let β represent parameters that remain constant across the regimes. Then, for any

$$\xi_{0|0},\ x_{0|0},\ P_{0|0}$$

$$\Downarrow$$

<u>Kalman Filter</u>

$$x^{(i)}_{t|t-1},\ P^{(i)}_{t|t-1},\ \eta^{(i)}_{t|t-1},\ C^{(i)}_{t|t-1},\ x^{(i)}_{t|t},\ P^{(i)}_{t|t}$$

$$\Downarrow$$

<u>Markov Switching Filter</u>

$$\xi_{t|t-1} := p\xi_{t-1|t-1} + (1-q)(1-\xi_{t-1|t-1})$$

$$f(y_t|\Omega_{t-1}) := \xi_{t|t-1} f(y_t|s_t = 1, \Omega_{t-1}) + (1-\xi_{t|t-1}) f(y_t|s_t = 2, \Omega_{t-1})$$

$$\xi_{t|t} := \frac{\left[p\xi_{t-1|t-1}+(1-q)(1-\xi_{t-1|t-1})\right] f(y_t|s_t=1,\Omega_{t-1})}{f(y_t|s_t=2,\Omega_{t-1})+\left[p\xi_{t-1|t-1}+(1-q)(1-\xi_{t-1|t-1})\right]\left[f(y_t|s_t=1,\Omega_{t-1})-f(y_t|s_t=2,\Omega_{t-1})\right]}$$

$$\Downarrow$$

<u>Aggregation</u>

$$x_{t|t} := \xi_{t|t} x^{(1)}_{t|t} + (1-\xi_{t|t}) x^{(2)}_{t|t}$$

$$P_{t|t} := \xi_{t|t} P^{(1)}_{t|t} + (1-\xi_{t|t}) P^{(2)}_{t|t} + \xi_{t|t}(1-\xi_{t|t}) \left(x^{(1)}_{t|t} - x^{(2)}_{t|t}\right)\left(x^{(1)}_{t|t} - x^{(2)}_{t|t}\right)' \qquad t-1=t$$

$$\Downarrow$$

$$l(\theta) = \sum_{t=1}^{T} \ln\left(f(y_t|\Omega_{t-1})\right)$$

Figure 1: Flowchart for the filter: state space models with regime switching

prespecified $0 < p, q < 1$, the modified log likelihood function is given by

$$
\begin{aligned}
&\mathcal{L}^A(p,q,\beta,\delta_1,\delta_2) \\
=&\sum_{t=1}^{T}\log\left\{f_{1t}(p,q,\beta,\delta_1,\delta_2)\xi_{t|t-1}(p,q,\beta,\delta_1,\delta_2)+f_{2t}(p,q,\beta,\delta_1,\delta_2)(1-\xi_{t|t-1}(p,q,\beta,\delta_1,\delta_2))\right\}, \quad (2.16)
\end{aligned}
$$

where

$$
f_{it}(p,q,\beta,\delta_1,\delta_2) = f(y_t|s_t = i, \Omega_{t-1}),
$$

which is defined in (2.14). When $\delta_1 = \delta_2 = \delta$, the modified log likelihood function reduces to

$$
\begin{aligned}
\mathcal{L}^N(\beta,\delta) &= \sum_{t=1}^{T}\log f_{1t}(p,q,\beta,\delta,\delta) \qquad (2.17)\\
&:= \sum_{t=1}^{T}\log f_t(\beta,\delta),
\end{aligned}
$$

which can be computed using the standard Kalman filter. This paper studies a test statistic based on (2.17) and (2.16) for the one regime specification versus the two regimes specification. To start, I impose the following restrictions on the DGP and the parameter space, following Assumption 1-3 in Qu and Zhuo (2015).

Assumption 1. *(i) The random vector* (z_t', y_t) *is strict stationary, ergodic and* β*-mixing with the mixing coefficient* β_τ *satisfying* $\beta_\tau \le c\rho^\tau$ *for some* $c > 0$ *and* $\rho \in [0, 1)$*. (ii) Under the null hypothesis,* y_t *is generated by* $f(\cdot|\Omega_{t-1}; \beta_*, \delta_*)$ *where* β_* *and* δ_* *are interior points of* $\Theta \subset \mathbb{R}^{n_\beta}$ *and* $\Delta \subset \mathbb{R}^{n_\delta}$ *with* Θ *and* Δ *being compact.*

Assumption 2. *Under the null hypothesis: (i)* (β_*, δ_*) *uniquely solves* $\max_{(\beta,\delta)\in\Theta\times\Delta} E\left[\mathcal{L}^N(\beta,\delta)\right]$*; (ii) for any* $0 < p, q < 1$*,* $(\beta_*, \delta_*, \delta_*)$ *uniquely solves* $\max_{(\beta,\delta_1,\delta_2)\in\Theta\times\Delta\times\Delta} E\left[\mathcal{L}^A(\beta,\delta_1,\delta_2)\right]$*.*

Assumption 3. *Under the null hypothesis, we have: (i)* $T^{-1}\left[\mathcal{L}^N(\beta,\delta) - E\mathcal{L}^N(\beta,\delta)\right] = o_p(1)$ *holds uniformly over* $(\beta,\delta) \in \Theta\times\Delta$*, with* $T^{-1}\sum_{t=1}^{T}\left(\nabla_{(\beta',\delta')'}\log f_t(\beta,\delta)\right)\left(\nabla_{(\beta',\delta')}\log f_t(\beta,\delta)\right)$ *being positive definite over an open neighborhood of* (β_*, δ_*) *for sufficiently large* T*; (ii) for any* $0 < p, q < 1$*,* $T^{-1}\left[\mathcal{L}^A(\beta,\delta_1,\delta_2) - E\mathcal{L}^A(\beta,\delta_1,\delta_2)\right] = o_p(1)$ *holds uniformly over* $(\beta,\delta_1,\delta_2) \in \Theta\times\Delta\times\Delta$*.*

Using the above notation, the null and alternative hypotheses can be more formally stated as:

$$H_0 : \delta_1 = \delta_2 = \delta_* \text{ for some unknown } \delta_*,$$

$$H_1 : (\delta_1, \delta_2) = (\delta_1^*, \delta_2^*) \text{ for some unknown } \delta_1^* \neq \delta_2^* \text{ and } (p,q) \in (0,1) \times (0,1).$$

For the remainder of this paper, I use an $ARMA(K, L)$ model to illustrate the main results.

The illustrative model. Let us consider a general $ARMA(K, L)$ model:

$$m_t = \sum_{k=1}^{K} \phi_{k,s_t} m_{t-k} + \varepsilon_t + \sum_{l=1}^{L} \theta_{l,s_t} \varepsilon_{t-l}, \ \varepsilon_t \sim i.i.d. \ N(0, \sigma_{s_t}^2), \tag{2.18}$$

$$y_t = \alpha_{s_t} + m_t, \tag{2.19}$$

where only y_t is observable but not m_t or s_t. In this illustrative model, some or all of the model parameters can be affected by s_t. This $ARMA$ model can be written into the setting in (2.1)-(2.3) as follows: Define $n_r = \max\{K, L+1\}$. Interpret $\phi_{j,s_t} = 0$ for $j > K$ and $\theta_{j,s_t} = 0$ for $j > L$. Let

$$F_{s_t} = \begin{bmatrix} \phi_{1,s_t} & \phi_{2,s_t} & \cdots & \phi_{n_r-1,s_t} & \phi_{n_r,s_t} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \quad G_{s_t} = 0,$$

$$u_t = \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \sim N\left(0, Q_{s_t}\right), \quad Q_{s_t} = \begin{bmatrix} \sigma_{s_t}^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix},$$

$$H_{s_t} = \begin{bmatrix} 1 \\ \theta_{1,s_t} \\ \vdots \\ \theta_{n_r-1,s_t} \end{bmatrix}, \quad A_{s_t} = \alpha_{s_t}, \text{ and } z_t = 1.$$

3 The test statistic

This section proposes a test statistic based on the MLR. Let $\tilde{\beta}$ and $\tilde{\delta}$ denote the maximizer of the log likelihood function under null hypothesis:

$$(\tilde{\beta}, \tilde{\delta}) = \arg\max_{\beta,\delta} \mathcal{L}^N(\beta, \delta). \tag{3.1}$$

The MLR evaluated at some $0 < p, q < 1$ then equals

$$MLR(p,q) = 2\left[\max_{\beta,\delta_1,\delta_2} \mathcal{L}^A(p,q,\beta,\delta_1,\delta_2) - \mathcal{L}^N(\tilde{\beta}, \tilde{\delta})\right]. \tag{3.2}$$

It is natural to consider the following test statistic:

$$SupMLR(\Lambda_\epsilon) = \sup_{(p,q)\in\Lambda_\epsilon} MLR(p,q),$$

where Λ_ϵ is a compact set to be specified later. Similar test statistics have been studied by Hansen (1992), Garcia (1998) and Qu and Zhuo (2015).

4 MLR under prespecified p and q

This section studies the MLR under a given $(p,q) \in \Lambda_\epsilon$. The choice of Λ_ϵ will be discussed in the next section.

4.1 Conditional regime probability

Let us first study the conditional regime probability $\xi_{t+1|t}(p,q,\beta,\delta_1,\delta_2)$ as well as its derivatives with respect to β, δ_1 and δ_2 because the results will be needed to develop the expansion of the modified log likelihood function. Combining equations (2.13) and (2.15) gives a recursive formula to calculate $\xi_{t+1|t}(p,q,\beta,\delta_1,\delta_2)$:

$$\begin{aligned}&\xi_{t+1|t}(p,q,\beta,\delta_1,\delta_2)\\ =&p + (p+q-1)\frac{f_{2t}(p,q,\beta,\delta_1,\delta_2)(\xi_{t|t-1}(p,q,\beta,\delta_1,\delta_2)-1)}{f_{1t}(p,q,\beta,\delta_1,\delta_2)\xi_{t|t-1}(p,q,\beta,\delta_1,\delta_2) + f_{2t}(p,q,\beta,\delta_1,\delta_2)(1-\xi_{t|t-1}(p,q,\beta,\delta_1,\delta_2))},\end{aligned} \tag{4.1}$$

where

$$f_{it}(p,q,\beta,\delta_1,\delta_2) = f(y_t|s_t = i, \Omega_{t-1}) \tag{4.2}$$

as in (2.14). This recursive formula implies that the derivatives of $\xi_{t+1|t}$ with respect to the model parameters must also follow first order difference equations. Because the asymptotic expansions are considered around the estimates under the null hypothesis, it is sufficient to analyze $\xi_{t+1|t}(p,q,\beta,\delta_1,\delta_2)$ and its derivatives at $\delta_1 = \delta_2 = \delta$ for an arbitrary value of δ in Δ.

Let $\theta = (\beta', \delta_1', \delta_2')'$ be an augmented parameter vector. We then three sets of integers (they index the elements in β, δ_1 and δ_2, respectively):

$$I_0 = \{1, ..., n_\beta\},\ I_1 = \{n_\beta + 1, ..., n_\beta + n_\delta\},\ I_2 = \{n_\beta + n_\delta + 1, ..., n_\beta + 2n_\delta\}.$$

Let "$\bar{g}$" denote that $g(\beta,\delta_1,\delta_2)$ is evaluated at some β and $\delta_1 = \delta_2 = \delta$, i.e., $\bar{\xi}_{t+1|t}$ and $\bar{f}_t$ denote that $\xi_{t+1|t}(p,q,\beta,\delta_1,\delta_2)$ and $f_{1t}(p,q,\beta,\delta_1,\delta_2)$ (or $f_{2t}(p,q,\beta,\delta_1,\delta_2)$) are evaluated at some β and $\delta_1 = \delta_2 = \delta$. Let $\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_k}}\bar{\xi}_{t|t-1}$, $\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_k}}\bar{f}_{1t}$ and $\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_k}}\bar{f}_{2t}$ denote the k-th order derivatives of $\xi_{t|t-1}(p,q,\beta,\delta_1,\delta_2)$, $f_{1t}(p,q,\beta,\delta_1,\delta_2)$ and $f_{2t}(p,q,\beta,\delta_1,\delta_2)$ with respect to the $(j_1,...,j_k)$-th elements of θ, evaluated at some β and $\delta_1 = \delta_2 = \delta$. Also let "$\bar{F}$" denote the matrix F_i ($i = 1$ or 2) evaluated at some β and $\delta_1 = \delta_2 = \delta$ and $\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_k}}\bar{F}_i$ denote the k-th order derivatives of the parameter matrix F_i evaluated at some β and $\delta_1 = \delta_2 = \delta$. By definition, the following relationships hold: $\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_k}}\bar{f}_{1t} = \nabla_{\theta_{j_1}}...\nabla_{\theta_{j_k}}\bar{f}_{2t}$ if $j_1,...,j_k$ all belong to I_0. The next lemma is parallel to Lemma 1 in Qu and Zhuo (2015), which characterizes the properties of $\xi_{t+1|t}(p,q,\beta,\delta_1,\delta_2)$ and its derivatives when $\delta_1 = \delta_2 = \delta$.

Lemma 1. *Let $\xi_{0|0} = \xi_*$, $\rho = p + q - 1$ and $r = \rho\xi_*(1-\xi_*)$ with ξ_* defined in (2.4). Then, for $t \geq 1$, we have:*

1. $\bar{\xi}_{t+1|t} = \xi_*$.

2. $\nabla_{\theta_j}\bar{\xi}_{t+1|t} = \rho\nabla_{\theta_j}\bar{\xi}_{t|t-1} + \bar{\mathcal{E}}_{j,t}$, *where*

$$\bar{\mathcal{E}}_{j,t} = r\left[\frac{\nabla_{\theta_j}\bar{f}_{1t}}{\bar{f}_t} - \frac{\nabla_{\theta_j}\bar{f}_{2t}}{\bar{f}_t}\right],$$

with $j \in \{I_0, I_1, I_2\}$.

3. $\nabla_{\theta_j}\nabla_{\theta_k}\bar{\xi}_{t+1|t} = \rho\nabla_{\theta_j}\nabla_{\theta_k}\bar{\xi}_{t|t-1} + \bar{\mathcal{E}}_{jk,t}$, *where $\bar{\mathcal{E}}_{jk,t}$ are given by (Let (I_a,I_b) denote the situation with $j \in I_a$ and $k \in I_b$; $a,b = 0,1,2,$):*

$(I_0, I_0):\ 0$

(I_0, I_1) *or* $(I_0, I_2): r\left(\frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{1t}}{\bar{f}_t} - \frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}}{\bar{f}_t} + \frac{\nabla_{\theta_j}\bar{f}_{2t}}{\bar{f}_t}\frac{\nabla_{\theta_k}\bar{f}_{2t}}{\bar{f}_t} - \frac{\nabla_{\theta_j}\bar{f}_{2t}}{\bar{f}_t}\frac{\nabla_{\theta_k}\bar{f}_{1t}}{\bar{f}_t}\right)$

(I_1, I_1) *or* (I_1, I_2) *or* (I_2, I_2) :

$$r\left(\frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{1t}}{\bar{f}_t} - \frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}}{\bar{f}_t}\right) + \rho(1-2\xi_*)\left[\left(\frac{\nabla_{\theta_j}\bar{f}_{1t}}{\bar{f}_t} - \frac{\nabla_{\theta_j}\bar{f}_{2t}}{\bar{f}_t}\right)\nabla_{\theta_k}\bar{\xi}_{t|t-1} + \left(\frac{\nabla_{\theta_k}\bar{f}_{1t}}{\bar{f}_t} - \frac{\nabla_{\theta_k}\bar{f}_{2t}}{\bar{f}_t}\right)\nabla_{\theta_j}\bar{\xi}_{t|t-1}\right]$$
$$-\left[2r\left(\xi_*\frac{\nabla_{\theta_j}\bar{f}_{1t}}{\bar{f}_t}\frac{\nabla_{\theta_k}\bar{f}_{1t}}{\bar{f}_t} - (1-\xi_*)\frac{\nabla_{\theta_j}\bar{f}_{2t}}{\bar{f}_t}\frac{\nabla_{\theta_k}\bar{f}_{2t}}{\bar{f}_t}\right)\right] + r(2\xi_*-1)\left[\frac{\nabla_{\theta_j}\bar{f}_{1t}}{\bar{f}_t}\frac{\nabla_{\theta_k}\bar{f}_{2t}}{\bar{f}_t} + \frac{\nabla_{\theta_j}\bar{f}_{2t}}{\bar{f}_t}\frac{\nabla_{\theta_k}\bar{f}_{1t}}{\bar{f}_t}\right].$$

4. $\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_k}\bar{\xi}_{t+1|t} = \rho\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_k}\bar{\xi}_{t|t-1} + \bar{\mathcal{E}}_{jkl,t}$*, where* $\bar{\mathcal{E}}_{jkl,t}$ *are given in the appendix with* $j,k,l \in \{I_a, I_b, I_c\}$ *and* $a,b,c = 0,1,2$*.*

I now discuss the first order derivatives of f_{it} appearing in the lemma. By (2.5), (2.6), (2.7), (2.8), (2.14), and (4.2), we have:

$$f_{it} = \left[2\pi H_i'\left(F_iP_{t-1|t-1}F_i' + Q_i\right)H_i\right]^{-1/2}\exp\left\{-\frac{\left[y_t - H_i'\left(G_i + F_ix_{t-1|t-1}\right) - A_i'z_t\right]^2}{2H_i'\left(F_iP_{t-1|t-1}F_i' + Q_i\right)H_i}\right\}.$$

The first order derivative of f_{it} with respect to the j-th component in θ is as follows:

$$\begin{aligned}\nabla_{\theta_j}f_{it} = {} & f_{it}\left\{\frac{\left[y_t - H_i'\left(G_i + F_ix_{t-1|t-1}\right) - A_i'z_t\right]}{H_i'\left(F_iP_{t-1|t-1}F_i' + Q_i\right)H_i}\right\} \\ & \times\left\{\nabla_{\theta_j}H_i'\left(G_i + F_ix_{t-1|t-1}\right) + \nabla_{\theta_j}A_i'z_t + H_i'\left(\nabla_{\theta_j}G_i + \nabla_{\theta_j}F_ix_{t-1|t-1} + F_i\nabla_{\theta_j}x_{t-1|t-1}\right)\right\} \\ & + f_{it}\left(\frac{1}{2H_i'\left(F_iP_{t-1|t-1}F_i' + Q_i\right)H_i}\right)\left\{\frac{\left[y_t - H_i'\left(G_i + F_ix_{t-1|t-1}\right) - A_i'z_t\right]^2}{H_i'\left(F_iP_{t-1|t-1}F_i' + Q_i\right)H_i} - 1\right\} \\ & \times\left\{\nabla_{\theta_j}H_i'\left(F_iP_{t-1|t-1}F_i' + Q_i\right)H_i + H_i'\left(F_iP_{t-1|t-1}F_i' + Q_i\right)\nabla_{\theta_j}H_i\right. \\ & \left.+H_i'\left(\nabla_{\theta_j}F_iP_{t-1|t-1}F_i' + F_i\nabla_{\theta_j}P_{t-1|t-1}F_i' + F_iP_{t-1|t-1}\nabla_{\theta_j}F_i' + \nabla_{\theta_j}Q_i\right)H_i\right\}.\end{aligned} \tag{4.3}$$

The exact expressions for the second order derivatives of f_{it} are included in the appendix. The properties of $x_{t|t}$ and $P_{t|t}$ and their derivatives will be studied in the next subsection (see Lemma 3 below). Note that, for some β and $\delta_1 = \delta_2$,

$$
\begin{aligned}
&\nabla_{\theta_j}\bar{f}_{1t} - \nabla_{\theta_j}\bar{f}_{2t} \\
&= \bar{f}_t \left\{ \frac{\left[y_t - \bar{H}'\left(\bar{G} + \bar{F}\bar{x}_{t-1|t-1}\right) - \bar{A}'z_t\right]}{\bar{H}'\left(\bar{F}\bar{P}_{t-1|t-1}\bar{F}' + \bar{Q}\right)\bar{H}} \right\} \\
&\times \left\{ \left(\nabla_{\theta_j}\bar{H}_1' - \nabla_{\theta_j}\bar{H}_2'\right)\left(\bar{G} + \bar{F}\bar{x}_{t-1|t-1}\right) + \left(\nabla_{\theta_j}\bar{A}_1' - \nabla_{\theta_j}\bar{A}_2'\right) z_t \right. \\
&\left. + \bar{H}'\left[\left(\nabla_{\theta_j}\bar{G}_1 - \nabla_{\theta_j}\bar{G}_2\right) + \left(\nabla_{\theta_j}\bar{F}_1 - \nabla_{\theta_j}\bar{F}_2\right)\bar{x}_{t-1|t-1}\right]\right\} \\
&+ \bar{f}_t \left(\frac{1}{2\bar{H}'\left(\bar{F}\bar{P}_{t-1|t-1}\bar{F}' + \bar{Q}\right)\bar{H}} \right) \left\{ \frac{\left[y_t - \bar{H}'\left(\bar{G} + \bar{F}\bar{x}_{t-1|t-1}\right) - \bar{A}'z_t\right]^2}{\bar{H}'\left(\bar{F}\bar{P}_{t-1|t-1}\bar{F}' + \bar{Q}\right)\bar{H}} - 1 \right\} \\
&\times \left\{ \left(\nabla_{\theta_j}\bar{H}_1' - \nabla_{\theta_j}\bar{H}_2'\right)\left(\bar{F}\bar{P}_{t-1|t-1}\bar{F}' + \bar{Q}\right)\bar{H} + \bar{H}'\left(\bar{F}\bar{P}_{t-1|t-1}\bar{F}' + \bar{Q}\right)\left(\nabla_{\theta_j}\bar{H}_1 - \nabla_{\theta_j}\bar{H}_2\right) \right. \\
&\left. + \bar{H}'\left[\left(\nabla_{\theta_j}\bar{F}_1 - \nabla_{\theta_j}\bar{F}_2\right)\bar{P}_{t-1|t-1}\bar{F}' + \bar{F}\bar{P}_{t-1|t-1}\left(\nabla_{\theta_j}\bar{F}_1' - \nabla_{\theta_j}\bar{F}_2'\right) + \left(\nabla_{\theta_j}\bar{Q}_1 - \nabla_{\theta_j}\bar{Q}_2\right)\right]\bar{H}\right\},
\end{aligned}
$$

in which the $\nabla_{\theta_j}\bar{x}_{t-1|t-1}$ and $\nabla_{\theta_j}\bar{P}_{t-1|t-1}$ terms are canceled out. Consequently, these two quantities are not needed for calculating $\nabla_{\theta_j}\bar{\xi}_{t+1|t}$. Similarly, $\nabla_{\theta_j}\nabla_{\theta_k}\bar{x}_{t-1|t-1}$ and $\nabla_{\theta_j}\nabla_{\theta_k}\bar{P}_{t-1|t-1}$ are not needed for calculating $\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{1t} - \nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}$. This is also true for higher order derivatives of $(f_{1t} - f_{2t})$ when evaluated at $\delta_1 = \delta_2$. I now use an example to illustrate Lemma 1.

The illustrative model (cont'd). Consider the illustrative example (2.18)-(2.19) and assume that both α_{s_t} and $\sigma^2_{s_t}$ are affected by regime switching. Lemma 1 implies that

$$
\begin{aligned}
\nabla_{\alpha_1}\bar{\xi}_{t+1|t} &= \rho\nabla_{\alpha_1}\bar{\xi}_{t|t-1} + r\frac{1}{\bar{\sigma}^2}\left[y_t - \bar{\alpha} - \bar{H}'\bar{F}\bar{x}_{t-1|t-1}\right], \\
\nabla_{\sigma_1^2}\bar{\xi}_{t+1|t} &= \rho\nabla_{\sigma_1^2}\bar{\xi}_{t|t-1} + r\frac{1}{2\bar{\sigma}^2}\left[\frac{\left(y_t - \bar{\alpha} - \bar{H}'\bar{F}\bar{x}_{t-1|t-1}\right)^2}{\bar{\sigma}^2} - 1\right].
\end{aligned}
$$

Because the filter described in subsections 2.2-2.3 reduces to the standard Kalman filter when $\delta_1 = \delta_2$, $\nabla_{\alpha_1}\bar{\xi}_{t+1|t}$ and $\nabla_{\sigma_1^2}\bar{\xi}_{t+1|t}$ both reduce to stationary $AR(1)$ processes with mean zero when evaluated at the true parameter values under the null hypothesis. Their variances are finite and satisfy

$$
E(\nabla_{\alpha_1}\bar{\xi}_{t+1|t})^2 = \frac{r^2}{(1-\rho^2)\sigma_*^2} \quad \text{and} \quad E(\nabla_{\sigma_1^2}\bar{\xi}_{t+1|t})^2 = \frac{r^2}{2(1-\rho^2)\sigma_*^4},
$$

where σ_*^2 denotes the true value of $\sigma^2_{s_t}$ under the null hypothesis.

4.2 Filtered state and its mean squared error

Since both the filtered state in (2.11) and its mean squared error in (2.12) are components in the modified log likelihood function, it is important to study these functions and their derivatives with respect to β, δ_1 and δ_2. As in the previous subsection, it is sufficient to study these quantities when $\delta_1 = \delta_2$.

Under Assumptions 1-3 and when $\delta_1 = \delta_2$, x_t in (2.1) is stationary. The unconditional mean of x_t can be employed as the initial value, denoted by $x_{0|0}$. The unconditional mean of x_t satisfies $E(x_t) = \bar{G} + \bar{F}E(x_{t-1})$. This implies

$$x_{0|0} = (I - \bar{F})^{-1}\bar{G}.$$

The next lemma provides the initial value for $P_{t|t}$ when $\delta_1 = \delta_2$, denoted by $P_{0|0}$. Its results are also used later to study the properties of $x_{t|t}$ and $P_{t|t}$.

Lemma 2. *Let $\bar{F}$ have all it eigenvalues inside the unit circle. Set $P_{0|0} = \bar{P}_*$, where $\bar{P}_*$ solves*

$$\bar{P}_* = \left[I - \left(\bar{F}\bar{P}_*\bar{F}' + \bar{Q}\right)\bar{H}\left[\bar{H}'\left(\bar{F}\bar{P}_*\bar{F}' + \bar{Q}\right)\bar{H}\right]^{-1}\bar{H}'\right]\left(\bar{F}\bar{P}_*\bar{F}' + \bar{Q}\right). \tag{4.4}$$

Then, under the null hypothesis and Assumption 1-3,

$$\bar{P}_{t|t} = \bar{P}_*,\ \bar{P}^{(i)}_{t|t} = \bar{P}_* \text{ and } \bar{P}_{t|t-1} = \bar{F}\bar{P}_*\bar{F}' + \bar{Q},$$

for all $t = 1, ..., T$.

In the subsequent analysis, $\bar{P}_{t|t-1}$ is abbreviated as $\bar{P}$. The following assumption, which is analogous to Proposition 13.2 in Hamilton (1994), ensures that $\bar{P}_*$ and $\bar{P}$ are unique.

Assumption 4. *The eigenvalues of*

$$\left(I - \frac{\bar{P}\bar{H}\bar{H}'}{\bar{H}'\bar{P}\bar{H}}\right)\bar{F}$$

are all inside the unit circle.

I use the *ARMA* model in (2.18)-(2.19) to illustrate this assumption.

The illustrative model (cont'd.) Consider the model in (2.18)-(2.19). Then, $\bar{P}_*$ and $\bar{P}$ are equal to 0 and $\bar{Q}$ respectively. The quantity in Assumption 4 is given by:

$$\left(I - \frac{\bar{P}\bar{H}\bar{H}'}{\bar{H}'\bar{P}\bar{H}}\right)\bar{F} = \begin{bmatrix} -\bar{\theta}_1 & -\bar{\theta}_2 & \cdots & -\bar{\theta}_{n_r-1} & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix},$$

where $\bar{\theta}_j$ denotes that the parameter θ_{j,s_t} is evaluated at some β and $\delta_1 = \delta_2 = \delta$. For the $ARMA(1,1)$ model, Assumption 4 is equivalent to $-1 < \bar{\theta}_1 < 1$.

The next lemma contains the details on the first and second order derivatives of the filtered state and its mean squared error when evaluated at $\delta_1 = \delta_2 = \delta$.

Lemma 3. *Under the null hypothesis and Assumptions 1-4, we have:*

1. For any $j \in I_a$, $a = 0, 1, 2$,

$$vec\left(\nabla_{\theta_j}\bar{P}_{t|t}\right) = \left[\left(I - \frac{\bar{P}\bar{H}\bar{H}'}{\bar{H}'\bar{P}\bar{H}}\right)\bar{F}\right]^{\otimes 2} vec\left(\nabla_{\theta_j}\bar{P}_{t-1|t-1}\right) + vec\left(\bar{\mathcal{P}}_{j,t}\right),$$

where

$$\begin{aligned}
\bar{\mathcal{P}}_{j,t} &= \xi_*\left(I - \frac{\bar{P}\bar{H}\bar{H}'}{\bar{H}'\bar{P}\bar{H}}\right)\left[\nabla_{\theta_j}\bar{F}_1\bar{P}_*\bar{F}' + \bar{F}\bar{P}_*\nabla_{\theta_j}\bar{F}_1' + \nabla_{\theta_j}Q_1\right]\left(I - \frac{\bar{H}\bar{H}'\bar{P}}{\bar{H}'\bar{P}\bar{H}}\right) \\
&+ (1-\xi_*)\left(I - \frac{\bar{P}\bar{H}\bar{H}'}{\bar{H}'\bar{P}\bar{H}}\right)\left[\nabla_{\theta_j}\bar{F}_2\bar{P}_*\bar{F}' + \bar{F}\bar{P}_*\nabla_{\theta_j}\bar{F}_2' + \nabla_{\theta_j}Q_2\right]\left(I - \frac{\bar{H}\bar{H}'\bar{P}}{\bar{H}'\bar{P}\bar{H}}\right) \\
&- \xi_*\left(\frac{\bar{P}\left(\nabla_{\theta_j}\bar{H}_1\bar{H}' + \bar{H}\nabla_{\theta_j}\bar{H}_1'\right)}{\bar{H}'\bar{P}\bar{H}} - \frac{\left(\bar{P}\bar{H}\bar{H}'\right)\left(\nabla_{\theta_j}\bar{H}_1'\bar{P}\bar{H} + \bar{H}'\bar{P}\nabla_{\theta_j}\bar{H}_1\right)}{\left(\bar{H}'\bar{P}\bar{H}\right)^2}\right)\left[\bar{F}\bar{P}_*\bar{F}' + \bar{Q}\right] \\
&- (1-\xi_*)\left(\frac{\bar{P}\left(\nabla_{\theta_j}\bar{H}_2\bar{H}' + \bar{H}\nabla_{\theta_j}\bar{H}_2'\right)}{\bar{H}'\bar{P}\bar{H}} - \frac{\left(\bar{P}\bar{H}\bar{H}'\right)\left(\nabla_{\theta_j}\bar{H}_2'\bar{P}\bar{H} + \bar{H}'\bar{P}\nabla_{\theta_j}\bar{H}_2\right)}{\left(\bar{H}'\bar{P}\bar{H}\right)^2}\right)\left[\bar{F}\bar{P}_*\bar{F}' + \bar{Q}\right].
\end{aligned}$$

2. For any $j \in I_a$, $a = 0, 1, 2$,

$$\nabla_{\theta_j}\bar{x}_{t|t} = \left[\left(I - \frac{\bar{P}\bar{H}\bar{H}'}{\bar{H}'\bar{P}\bar{H}}\right)\bar{F}\right]\nabla_{\theta_j}\bar{x}_{t-1|t-1} + \bar{\mathcal{X}}_{j,t},$$

where the expression of $\bar{\mathcal{X}}_{j,t}$ are given in the appendix.

3. For any $j \in I_a$ and $k \in I_b$, $a, b = 0, 1, 2$,

$$vec\left(\nabla_{\theta_j}\nabla_{\theta_k}\bar{P}_{t|t}\right) = \left[\left(I - \frac{\bar{P}\bar{H}\bar{H}'}{\bar{H}'\bar{P}\bar{H}}\right)\bar{F}\right]^{\otimes 2} vec\left(\nabla_{\theta_j}\nabla_{\theta_k}\bar{P}_{t-1|t-1}\right) + vec\left(\bar{\mathcal{P}}_{jk,t}\right),$$

where the expression of $\bar{\mathcal{P}}_{jk,t}$ *are given in the appendix.*

4. For any $j \in I_a$ *and* $k \in I_b$*,* $a, b = 0, 1, 2$*,*

$$\nabla_{\theta_j}\nabla_{\theta_k}\bar{x}_{t|t} = \left[\left(I - \frac{\bar{P}\bar{H}\bar{H}'}{\bar{H}'\bar{P}\bar{H}}\right)\bar{F}\right]\nabla_{\theta_j}\nabla_{\theta_k}\bar{x}_{t-1|t-1} + \bar{\mathcal{X}}_{jk,t},$$

where the expression of $\bar{\mathcal{X}}_{jk,t}$ *are given in the appendix.*

Lemma 3 shows that the first and second order derivatives of the filtered state and its mean squared error all follow first order linear difference equations and the lagged coefficient matrices for them always include $\left[I - \bar{P}\bar{H}(\bar{H}'\bar{P}\bar{H})^{-1}\bar{H}'\right]\bar{F}$. The recursive structures implied by Lemma 3 suggest that we apply a similar strategy to analyze $x_{t|t}$ and $P_{t|t}$, as we are studying the properties of the higher order derivatives of $\xi_{t+1|t}$. The following example illustrates the results in Lemma 3 with an $ARMA(1,1)$ model.

The illustrative model (cont'd.) Consider the $ARMA(1,1)$ model in (2.18)-(2.19) and assume only α_{s_t} switches. Lemma 3 implies that $\bar{P}_*$ and $\bar{P}$ are equal to 0 and $\bar{Q}$, respectively. Meanwhile, further calculations show that $\nabla_{\theta_j}\bar{P}_{t|t} = 0$ for $j \in \{1, ..., n_\beta + 2\}$ and

$$\nabla_{\theta_j}\bar{x}_{t|t} = \begin{cases} \begin{bmatrix} 0 & 0 \end{bmatrix}' & j \in \{1, ..., n_\beta\}, \\ \begin{bmatrix} -\frac{\xi_*}{1+\bar{\theta}_1} & 0 \end{bmatrix}' & j = n_\beta + 1, \\ \begin{bmatrix} -\frac{1-\xi_*}{1+\bar{\theta}_1} & 0 \end{bmatrix}' & j = n_\beta + 2. \end{cases}$$

The second order derivatives of $P_{t|t}$ and $x_{t|t}$, with respect to α_1, satisfy

$$\nabla^2_{\alpha_1}\bar{P}_{t|t} = 2\xi_*(1-\xi_*)\left(\frac{1}{1-\bar{\theta}_1^2}\right)\begin{bmatrix} 1 & -\bar{\theta}_1 \\ -\bar{\theta}_1 & 1 \end{bmatrix}$$

and

$$\begin{aligned}\nabla^2_{\alpha_1}\bar{x}_{t|t} &= \begin{bmatrix} -\bar{\theta}_1 & 0 \\ 1 & 0 \end{bmatrix}\nabla^2_{\alpha_1}\bar{x}_{t-1|t-1} - 2\nabla_{\alpha_1}\bar{\xi}_{t|t}\begin{bmatrix} 1 & 0 \end{bmatrix}' \\ &\quad + \begin{bmatrix} -\bar{\phi}_1\bar{\theta}_1 & \bar{\phi}_1 \end{bmatrix}'\left(\frac{y_t - \bar{\alpha} - \bar{H}'\bar{F}\bar{x}_{t-1|t-1}}{\bar{\sigma}^2}\right)2\xi_*(1-\xi_*),\end{aligned}$$

where

$$\nabla_{\alpha_1}\bar{\xi}_{t|t} = \rho\nabla_{\alpha_1}\bar{\xi}_{t-1|t-1} + (1-\xi_*)\xi_*\left[\frac{\nabla_{\alpha_1}\bar{f}_{1t}}{\bar{f}_t} - \frac{\nabla_{\alpha_1}\bar{f}_{2t}}{\bar{f}_t}\right].$$

In this case, when $\delta_1 = \delta_2$, $\nabla^2_{\alpha_1}\bar{P}_{t|t}$ is a constant matrix, while $\nabla^2_{\alpha_1}\bar{x}_{t|t}$ depends on $\nabla^2_{\alpha_1}\bar{x}_{t-1|t-1}$, $\nabla_{\alpha_1}\bar{\xi}_{t|t}$, and the prediction error, $y_t - \bar{\alpha} - \bar{H}'\bar{F}\bar{x}_{t-1|t-1}$, at time t.

4.3 Modified log likelihood function and its expansion

Because of the multiple local maxima in $\mathcal{L}^A(p,q,\beta,\delta_1,\delta_2)$, it is difficult to directly expand this function around the null estimates $(\tilde{\beta},\tilde{\delta},\tilde{\delta})$. Both Cho and White (2007) and Qu and Zhuo (2015) suggest to work with the concentrated likelihood function. To derive the concentrated likelihood function, β and δ_1 are treated as functions of δ_2 and the dependence between (β,δ_1) and δ_2 is quantified using the first order conditions that define the concentrated and modified log likelihood function (see Lemma A.3 in the appendix). This effectively removes β and δ_1 from the subsequent analysis and allows us to work with the concentrated, modified log likelihood function, which is only a function of δ_2. Therefore, we can expand the concentrated, modified log likelihood function around $\delta_2 = \tilde{\delta}$ (see Lemma 4 below) to obtain an approximation for $MLR(p,q)$.

For any $\delta_2 \in \Delta$, we can write

$$\mathcal{L}(p,q,\delta_2) = \max_{\beta,\delta_1}\mathcal{L}^A(p,q,\beta,\delta_1,\delta_2)$$

and

$$\left(\hat{\beta}(\delta_2),\hat{\delta}_1(\delta_2)\right) = \arg\max_{\beta,\delta_1}\mathcal{L}^A(p,q,\beta,\delta_1,\delta_2).$$

Then,

$$MLR(p,q) = 2\max_{\delta_2}[\mathcal{L}(p,q,\delta_2) - \mathcal{L}(p,q,\tilde{\delta})].$$

For $k \geq 1$, let $\mathcal{L}^{(k)}_{i_1\ldots i_k}(p,q,\delta_2)$ $(i_1,\ldots i_k \in \{1,\ldots,n_\delta\})$ denote the k-th order derivative of $\mathcal{L}(p,q,\delta_2)$ with respect to the $(i_1,\ldots i_k)$-th elements of δ_2. Let d_j $(j \in \{1,\ldots,n_\delta\})$ denote the j-th element of

$(\delta_2 - \tilde{\delta})$. Then, a fourth order Taylor expansion of $\mathcal{L}(p,q,\delta_2)$ around $\tilde{\delta}$ is given by

$$\begin{aligned}\mathcal{L}(p,q,\delta_2) - \mathcal{L}(p,q,\tilde{\delta}) &= \sum_{j=1}^{n_\delta} \mathcal{L}_j^{(1)}(p,q,\tilde{\delta})d_j + \frac{1}{2!}\sum_{j=1}^{n_\delta}\sum_{k=1}^{n_\delta} \mathcal{L}_{jk}^{(2)}(p,q,\tilde{\delta})d_j d_k \\ &+ \frac{1}{3!}\sum_{j=1}^{n_\delta}\sum_{k=1}^{n_\delta}\sum_{l=1}^{n_\delta} \mathcal{L}_{jkl}^{(3)}(p,q,\tilde{\delta})d_j d_k d_l \\ &+ \frac{1}{4!}\sum_{j=1}^{n_\delta}\sum_{k=1}^{n_\delta}\sum_{l=1}^{n_\delta}\sum_{m=1}^{n_\delta} \mathcal{L}_{jklm}^{(4)}(p,q,\bar{\delta})d_j d_k d_l d_m,\end{aligned} \tag{4.5}$$

where in the last term $\bar{\delta}$ is a value that lies between δ_2 and $\tilde{\delta}$. Two lemmas will be provided to analyze this expansion. Here, two more assumptions, which are similar to Assumptions 4 and 5 in Qu and Zhuo (2015), are needed.

Assumption 5. *There exists an open neighborhood of (β_*, δ_*), denoted by $B(\beta_*, \delta_*)$, and a sequence of positive, strictly stationary and ergodic random variables $\{\upsilon_t\}$ satisfying $E\upsilon_t^{1+c} < L < \infty$ for some $c > 0$, such that*

$$\sup_{(\beta,\delta_1)\in B(\beta_*,\delta_*)} \left| \frac{\nabla_{\theta_{i_1}} ... \nabla_{\theta_{i_k}} f_t(\beta,\delta_1)}{f_t(\beta,\delta_1)} \right|^{\frac{\alpha(k)}{k}} < \upsilon_t$$

for all $i_1, ..., i_k \in \{1, ..., n_\beta + n_\delta\}$, where $1 \le k \le 5$; $\alpha(k) = 6$ if $k = 1,2,3$ and $\alpha(k) = 5$ if $k = 4,5$.

Assumption 6. *There exists $\eta > 0$, such that $\sup_{p,q\in[\epsilon,1-\epsilon]} \sup_{|\delta-\tilde{\delta}|<\eta} T^{-1}|\mathcal{L}_{jklmn}^{(5)}(p,q,\delta)| = O_p(1)$ for all $j,k,l,m,n \in \{1,...,n_\delta\}$, where ϵ is an arbitrary small constant satisfying $0 < \epsilon < 1/2$.*

The next lemma characterizes the derivatives of $\hat{\beta}(\delta_2)$ and $\hat{\delta}_1(\delta_2)$ with respect to δ_2 evaluated at $\delta_2 = \tilde{\delta}$. To shorten the expressions, let $\tilde{\xi}_{t+1|t}$ and $\tilde{f}_t$ denote $\xi_{t+1|t}(p,q,\beta,\delta_1,\delta_2)$ and $f_t(\beta,\delta_1)$ evaluated at $(\beta,\delta_1,\delta_2) = (\tilde{\beta},\tilde{\delta},\tilde{\delta})$. Also, let $\nabla_{\delta_{1i_1}} ... \nabla_{\delta_{1i_k}} \tilde{\xi}_{t|t-1}$ and $\nabla_{\delta_{1i_1}} ... \nabla_{\delta_{1i_k}} \tilde{f}_{1t}$ denote the k-th order derivative of $\xi_{t+1|t}(p,q,\beta,\delta_1,\delta_2)$ and $f_t(\beta,\delta_1)$ with respect to the $(i_1,...,i_k)$-th elements of δ_1, evaluated at $(\beta,\delta_1,\delta_2) = (\tilde{\beta},\tilde{\delta},\tilde{\delta})$. Finally, define

$$\begin{aligned}\tilde{U}_{jk,t} &= \left(\frac{1-\xi_*}{\xi_*}\right)^2 \left[\xi_* \frac{\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\tilde{f}_{1t}}{\tilde{f}_t} + (1-\xi_*)\frac{\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\tilde{f}_{2t}}{\tilde{f}_t}\right] + \left[\xi_* \frac{\nabla_{\delta_{2j}}\nabla_{\delta_{2k}}\tilde{f}_{1t}}{\tilde{f}_t} + (1-\xi_*)\frac{\nabla_{\delta_{2j}}\nabla_{\delta_{2k}}\tilde{f}_{2t}}{\tilde{f}_t}\right] \\ &- \left(\frac{1-\xi_*}{\xi_*}\right)\left[\xi_* \frac{\nabla_{\delta_{2j}}\nabla_{\delta_{1k}}\tilde{f}_{1t}}{\tilde{f}_t} + (1-\xi_*)\frac{\nabla_{\delta_{2j}}\nabla_{\delta_{1k}}\tilde{f}_{2t}}{\tilde{f}_t} + \xi_*\frac{\nabla_{\delta_{1j}}\nabla_{\delta_{2k}}\tilde{f}_{1t}}{\tilde{f}_t} + (1-\xi_*)\frac{\nabla_{\delta_{1j}}\nabla_{\delta_{2k}}\tilde{f}_{2t}}{\tilde{f}_t}\right] \\ &+ \frac{1}{\xi_*^2}\left[\nabla_{\delta_{1j}}\tilde{\xi}_{t|t-1}\left(\frac{\nabla_{\theta_k}\bar{f}_{1t}}{\bar{f}_t} - \frac{\nabla_{\theta_k}\bar{f}_{2t}}{\bar{f}_t}\right) + \left(\frac{\nabla_{\theta_j}\bar{f}_{1t}}{\bar{f}_t} - \frac{\nabla_{\theta_j}\bar{f}_{2t}}{\bar{f}_t}\right)\nabla_{\delta_{1k}}\tilde{\xi}_{t|t-1}\right],\end{aligned} \tag{4.6}$$

and

$$\tilde{D}_{jk,t} = \frac{\xi_* \nabla_{(\beta',\delta_1')'} \tilde{f}_{1t} + (1-\xi_*) \nabla_{(\beta',\delta_2')'} \tilde{f}_{2t}}{\tilde{f}_t} \tilde{U}_{jk,t},$$

$$\tilde{I}_t = \left(\frac{\xi_* \nabla_{(\beta',\delta_1')'} \tilde{f}_{1t} + (1-\xi_*) \nabla_{(\beta',\delta_2')'} \tilde{f}_{2t}}{\tilde{f}_t} \right) \left(\frac{\xi_* \nabla_{(\beta',\delta_1')} \tilde{f}_{1t} + (1-\xi_*) \nabla_{(\beta',\delta_2')} \tilde{f}_{2t}}{\tilde{f}_t} \right),$$

$$\tilde{V}_{jklm} = T^{-1} \sum_{t=1}^{T} \tilde{U}_{jk,t} \tilde{U}_{lm,t}, \quad \tilde{D}_{lm} = T^{-1} \sum_{t=1}^{T} \tilde{D}_{lm,t}, \quad \tilde{I} = T^{-1} \sum_{t=1}^{T} \tilde{I}_t. \tag{4.7}$$

As will be seen, $\tilde{U}_{jk,t}$ is the leading term in $\mathcal{L}^{(2)}_{jk}(p,q,\tilde{\delta})$, while $\tilde{D}_{jk,t}$ and $\tilde{I}_t$ appear when constructing the leading term of $\mathcal{L}^{(4)}_{jklm}(p,q,\tilde{\delta})$. The next lemma, analogous to Lemma 3 in Qu and Zhuo (2015), presents the properties of $\mathcal{L}^{(k)}_{i_1 \ldots i_k}(p,q,\delta_2)$ when δ_2 is evaluated at the null estimate, i.e. $\tilde{\delta}$.

Lemma 4. *Under the null hypothesis and Assumptions 1-6, for all $j,k,l,m \in \{1,...,n_\delta\}$, we have*

1. $\mathcal{L}^{(1)}_{j}(p,q,\tilde{\delta}) = 0$.
2. $T^{-1/2}\mathcal{L}^{(2)}_{jk}(p,q,\tilde{\delta}) = T^{-1/2}\sum_{t=1}^{T} \tilde{U}_{jk,t} + o_p(1)$.
3. $T^{-3/4}\mathcal{L}^{(3)}_{jkl}(p,q,\tilde{\delta}) = O_p\left(T^{-1/4}\right)$.
4. $T^{-1}\mathcal{L}^{(4)}_{jklm}(p,q,\bar{\delta}) = -\{\tilde{V}_{jklm} - \tilde{D}'_{jk}\tilde{I}^{-1}\tilde{D}_{lm} + \tilde{V}_{jmkl} - \tilde{D}'_{jm}\tilde{I}^{-1}\tilde{D}_{kl} + \tilde{V}_{jlkm} - \tilde{D}'_{jl}\tilde{I}^{-1}\tilde{D}_{km}\} + o_p(1)$.

The illustrative model (cont'd.) We use the $ARMA(1,1)$ model to illustrate the leading terms of $T^{-1/2}\mathcal{L}^{(2)}_{jk}(p,q,\tilde{\delta})$ and $T^{-1}\mathcal{L}^{(4)}_{jklm}(p,q,\tilde{\delta})$ in Lemma 4. Suppose only α_{s_t} switches. Then, $\tilde{U}_{jk,t}$ and $\tilde{D}_{jk,t}$ equal, respectively,

$$\left(\frac{1-\xi_*}{\xi_*}\right)\left\{\left(\frac{1+\tilde{\phi}_1^2+2\tilde{\phi}_1\tilde{\theta}_1}{1-\tilde{\theta}_1^2}\right)\left(\frac{1}{\tilde{\sigma}^2}\right)\left(\frac{\tilde{\mu}_t^2}{\tilde{\sigma}^2}-1\right)+2\left(\frac{\tilde{\mu}_t}{\tilde{\sigma}^2}\right)\left[\sum_{s=1}^{t-1}\rho^s\left(\frac{\tilde{\mu}_{t-s}}{\tilde{\sigma}^2}\right)\right]\right.$$
$$\left.-2(\tilde{\phi}_1+\tilde{\theta}_1)\sum_{j=1}^{t-1}(-\tilde{\theta}_1)^{j-1}\left[\sum_{s=1}^{t-j}\rho^{s-1}\left(\frac{\tilde{\mu}_{t-s}}{\tilde{\sigma}^2}\right)+\tilde{\phi}_1\tilde{\theta}_1\left(\frac{\tilde{\mu}_{t-j}}{\tilde{\sigma}^2}\right)\right]\right\}$$

and

$$\left[\begin{array}{cccc} \frac{(y_{t-1}-\tilde{\alpha})\tilde{\mu}_t}{\tilde{\sigma}^2} & \frac{\tilde{\mu}_t\tilde{\mu}_{t-1}}{\tilde{\sigma}^2} & \frac{1}{2\tilde{\sigma}^2}\left(\frac{\tilde{\mu}_t^2}{\tilde{\sigma}^2}-1\right) & -\frac{\tilde{\mu}_t}{\tilde{\sigma}^2}\left(\frac{\xi_*(1-\tilde{\phi}_1)}{1+\tilde{\theta}_1}\right) \end{array}\right]\tilde{U}_{jk,t},$$

where $\tilde{\mu}_t$ denotes the residuals under the null hypothesis and $\tilde{\sigma}^2 = T^{-1}\sum_{t=1}^{T}\tilde{\mu}_t^2$. This makes the variance function of $T^{-1/2}\sum_{t=1}^{T}\tilde{U}_{jk,t}$, and therefore of $T^{-1/2}\mathcal{L}^{(2)}_{jk}(p,q,\tilde{\delta})$, consistently estimable.

5 Asymptotic approximations

Let $\mathcal{L}^{(2)}(p,q,\tilde{\delta})$ be a square matrix with its (j,k)-th element given by $\mathcal{L}^{(2)}_{jk}(p,q,\tilde{\delta})$ for $j,k \in \{1,2,...,n_\delta\}$. This section includes three sets of results. (1) The weak convergence of $T^{-1/2}\mathcal{L}^{(2)}(p,q,\tilde{\delta})$ over $\epsilon \leq p,q \leq 1-\epsilon$. (2) The limiting distribution of $SupMLR(\Lambda_\epsilon)$. (3) A finite sample refinement that improves the asymptotic approximation.

5.1 Weak convergence of $\mathcal{L}^{(2)}(p,q,\tilde{\delta})$

For $0 < p_r, q_r, p_s, q_s < 1$ and $j,k,l,m \in \{1,2,...,n_\delta\}$, define

$$\omega_{jklm}(p_r,q_r;p_s,q_s) = V_{jklm}(p_r,q_r;p_s,q_s) - D'_{jk}(p_r,q_r)I^{-1}D_{lm}(p_s,q_s), \quad (5.1)$$

where $V_{jklm}(p_r,q_r;p_s,q_s) = E\left[U_{jk,t}(p_r,q_r)U_{lm,t}(p_s,q_s)\right], D_{jk}(p_r,q_r) = ED_{jk,t}(p_r,q_r)$, and $I = EI_t$. Here, $U_{jk,t}(p_r,q_r)$, $D_{jk,t}(p_r,q_r)$ and I_t have the same definitions as $\tilde{U}_{jk,t}$, $\tilde{D}_{jk,t}$ and $\tilde{I}_t$ in (4.6) and (4.7) but evaluated at $(p_r,q_r,\beta_*,\delta_*)$ instead of $(p_r,q_r,\tilde{\beta},\tilde{\delta})$. The following lemma is parallel to Lemma 4 in Qu and Zhuo (2015).

Lemma 5. *Under the null hypothesis and Assumptions 1-6, we have, over* $\epsilon \leq p,q \leq 1-\epsilon$:

$$T^{-1/2}\mathcal{L}^{(2)}(p,q,\tilde{\delta}) \Rightarrow G(p,q),$$

where the elements of $G(p,q)$ *are mean zero continuous Gaussian processes satisfying*

$$Cov[G_{jk}(p_r,q_r), G_{lm}(p_s,q_s)] = \omega_{jklm}(p_r,q_r;p_s,q_s)$$

for $j,k,l,m \in \{1,2,...,n_\delta\}$, *where* $\omega_{jklm}(p_r,q_r;p_s,q_s)$ *is given by (5.1).*

In the appendix, this lemma is proved by first showing the finite dimensional convergence and then the stochastic equicontinuity.

5.2 Limiting distribution of $SupMLR(\Lambda_\varepsilon)$

Let E be a set of open balls that includes all possible values of (p,q) such that $\mathcal{L}^{(2)}_{jk}(p,q,\tilde{\delta}) \equiv 0$ for any $j,k \in \{1,2,...,n_\delta\}$. For example, if for some specific j_1 and k_1, $\mathcal{L}^{(2)}_{j_1k_1}(p_1,q_1,\tilde{\delta}) \equiv 0$, then $(p,q) \in E$ if

$p \in (p_1 - \epsilon_1, p_1 + \epsilon_1)$ and $q \in (q_1 - \epsilon_1, q_1 + \epsilon_1)$ for any small ϵ_1, say $\epsilon_1 = 0.01$. Define

$$\Lambda_\epsilon = \{(p, q) : \epsilon \leq p, q \leq 1 - \epsilon, \text{ and } (p, q) \notin E\}. \tag{5.2}$$

Let $\Omega(p, q)$ be an n_δ^2-dimensional square matrix whose $(j + (k-1)n_\delta, l + (m-1)n_\delta)$-th element is given by $\omega_{jklm}(p, q; p, q)$. Then, Lemma 5 implies $E[\text{vec}G(p,q) \ \text{vec}G(p,q)'] = \Omega(p, q)$. The next result, which is analogous to Proposition 2 in Qu and Zhuo (2015), gives the asymptotic distribution of $SupMLR(\Lambda_\epsilon)$.

Proposition 1. *Suppose the null hypothesis and Assumptions 1-6 hold. Then*

$$SupMLR(\Lambda_\epsilon) \Rightarrow \sup_{(p,q)\in\Lambda_\epsilon} \sup_{\eta \in R^{n_\delta}} \mathcal{W}^{(2)}(p, q, \eta), \tag{5.3}$$

where Λ_ϵ *is given by (5.2) and*

$$\mathcal{W}^{(2)}(p, q, \eta) = \left(\eta^{\otimes 2}\right)' vecG(p, q) - \frac{1}{4}\left(\eta^{\otimes 2}\right)' \Omega(p, q) \left(\eta^{\otimes 2}\right).$$

Some important features of $\omega_{jklm}(p, q; p, q)$ have been shown in section 5.1 of Qu and Zhuo (2015) by simple examples. In the current context, I also observe the similar features of $\omega_{jklm}(p, q; p, q)$ that this function depends on: (1) the model's dynamic properties (e.g., whether the regressors are strictly exogenous or predetermined), (2) which parameters are allowed to switch (e.g., regressions coefficients or the variance of the errors), and (3) whether nuisance parameters are present.

5.3 A refinement

Qu and Zhuo (2015) provide a refinement to the asymptotic distribution when $\mathcal{L}^{(2)}(p, q, \tilde{\delta}) \equiv 0$. In such situations, the magnitude of $T^{-1/2}\sum_{t=1}^{T} \tilde{U}_{jk,t}$ can be too small to dominate the higher order terms in the likelihood expansion when $p + q$ is close to 1. This indicates that an asymptotic distribution that relies entirely on $T^{-1/2}\sum_{t=1}^{T} \tilde{U}_{jk,t}$ can be inadequate. Motivated by this observation, I consider a refinement to the asymptotic approximation under Markov switching state space models. The following assumption is parallel to Assumption 6 in Qu and Zhuo (2015).

Assumption 7. There exists an open neighborhood of (β_*, δ_*), $B(\beta_*, \delta_*)$, and a sequence of positive, strictly stationary and ergodic random variables $\{\upsilon_t\}$ satisfying $E\upsilon_t^{1+c} < \infty$ for some $c > 0$, such that the supremums of the following quantities over $B(\beta_*, \delta_*)$ are bounded from above by υ_t: $\left|\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}} f_t\left(\beta,\delta_1\right)/f_t\left(\beta,\delta_1\right)\right|^4$, $\left|\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_m}} f_t\left(\beta,\delta_1\right)/f_t\left(\beta,\delta_1\right)\right|^2$, $\left|\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_8}} f_t\left(\beta,\delta_1\right)/f_t\left(\beta,\delta_1\right)\right|$, $\left|\nabla_{\theta_{j_1}}\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_7}} f_t\left(\beta,\delta_1\right)/f_t\left(\beta,\delta_1\right)\right|$, $\left|\nabla_{\theta_{j_1}}\nabla_{\theta_{j_2}}\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_6}} f_t\left(\beta,\delta_1\right)/f_t\left(\beta,\delta_1\right)\right|$, where $k = 1, 2, 3, 4$, $m = 5, 6, 7$, $i_1, ..., i_8 \in \{1, ..., n_\beta + n_\delta\}$ and $j_1, j_2 \in \{1, ..., n_\beta\}$.

Obtaining all the leading terms in an even higher order expansion of the modified likelihood function is very difficult in the current context. This paper considers incorporating some specific terms for the refinement. Define

$$\tilde{s}_{jkl,t}(p,q) = -\frac{(1-\xi_*)(1-2\xi_*)}{\xi_*^2}\frac{\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\tilde{f}_{1t}}{\tilde{f}_t}, \tag{5.4}$$

where $\tilde{x}_{t|t}$ and $\tilde{P}_{t|t}$ are treated as constant when calculating $\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\tilde{f}_{1t}$ here. For $j, k, l, m, n, u \in \{1, ..., n_\delta\}$, let $G^{(3)}_{jkl}(p,q)$ be a continuous Gaussian process with mean zero and satisfy

$$\begin{aligned}
&\omega^{(3)}_{jklmnu}(p_r, q_r; p_s, q_s)\\
&= Cov(G^{(3)}_{jkl}\left(p_r, q_r\right), G^{(3)}_{mnu}\left(p_s, q_s\right))\\
&= E\left[s_{jkl,t}(p_r, q_r)s_{mnu,t}(p_s, q_s)\right]\\
&- E\left[\frac{\xi_*\nabla_{(\beta',\delta_1')}f_{1t} + (1-\xi_*)\nabla_{(\beta',\delta_2')}f_{2t}}{\tilde{f}_t}s_{jkl,t}(p_r, q_r)\right] I^{-1}\left[\frac{\xi_*\nabla_{(\beta',\delta_1')'}f_{1t} + (1-\xi_*)\nabla_{(\beta',\delta_2')'}f_{2t}}{\tilde{f}_t}s_{mnu,t}(p_s, q_s)\right],
\end{aligned}$$

where $s_{jkl,t}(p,q)$ is the same as $\tilde{s}_{jkl,t}(p,q)$ but is evaluated at true parameter values. The other quantities on the right hand side are also evaluated at the true parameter values. For the fourth and eighth order derivatives, define

$$\tilde{k}_{jklm,t}(p,q) = (1-\xi_*)\left(1 + \left(\frac{1-\xi_*}{\xi_*}\right)^3\right)\frac{\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\nabla_{\delta_{1m}}\tilde{f}_{1t}}{\tilde{f}_t}, \tag{5.5}$$

where $\tilde{x}_{t|t}$ and $\tilde{P}_{t|t}$ are treated as constant when calculating $\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\nabla_{\delta_{1m}}\tilde{f}_{1t}$ here. For $i_1, ..., i_8 \in \{1, ..., n_\delta\}$, let $G^{(4)}_{i_1i_2i_3i_4}(p,q)$ denote a continuous Gaussian process with mean zero and satisfy

$$\begin{aligned}
&\omega^{(4)}_{i_1i_2...i_8}(p_r, q_r; p_s, q_s)\\
&= Cov\left(G^{(4)}_{i_1i_2i_3i_4}\left(p_r, q_r\right), G^{(4)}_{i_5i_6i_7i_8}\left(p_s, q_s\right)\right)\\
&= E\left[k_{i_1i_2i_3i_4,t}\left(p_r, q_r\right)k_{i_5i_6i_7i_8,t}\left(p_s, q_s\right)\right]\\
&- E\left[\frac{\xi_*\nabla_{(\beta',\delta_1')}f_{1t} + (1-\xi_*)\nabla_{(\beta',\delta_2')}f_{2t}}{\tilde{f}_t}k_{i_1i_2i_3i_4,t}\left(p_r, q_r\right)\right] I^{-1}\left[\frac{\xi_*\nabla_{(\beta',\delta_1')}f_{1t} + (1-\xi_*)\nabla_{(\beta',\delta_2')}f_{2t}}{\tilde{f}_t}k_{i_5i_6i_7i_8,t}\left(p_s, q_s\right)\right],
\end{aligned}$$

where $k_{i_1i_2i_3i_4,t}(p,q)$ equals $\tilde{k}_{i_1i_2i_3i_4,t}(p,q)$ but evaluated at the true parameter values. The remaining quantities on the right hand side are also evaluated at the true parameter values. The next lemma characterizes the asymptotic properties of $\tilde{s}_{jkl,t}(p,q)$ and $\tilde{k}_{jklm,t}(p,q)$ when $j,k,l,m \in \{1,...,n_\delta\}$.

Lemma 6. *Under the null hypothesis and Assumptions 1-7, we have*

$$T^{-1/2}\sum_{t=1}^{T}\tilde{s}_{jkl,t}(p,q) \Rightarrow G^{(3)}_{jkl}(p,q)$$

and

$$T^{-1/2}\sum_{t=1}^{T}\tilde{k}_{jklm,t}(p,q) \Rightarrow G^{(4)}_{jklm}(p,q).$$

We now incorporate the corresponding terms to obtain a refined approximation. Let $G^{(3)}(p,q)$ be a n_δ^3- dimensional vector whose $(j+(k-1)n_\delta+(l-1)n_\delta^2)$-th element is given by $G^{(3)}_{jkl}(p,q)$. Let $\Omega^{(3)}(p,q)$ denote an $n_\delta^3 - by - n_\delta^3$ matrix whose $(j+(k-1)n_\delta+(l-1)n_\delta^2, m+(n-1)n_\delta+(r-1)n_\delta^2)$-th element is given by $\omega^{(3)}_{jklmnr}(p,q;p,q)$. Define

$$\mathcal{W}^{(3)}(p,q,\eta) = T^{-1/4}\frac{1}{3}\left(\eta^{\otimes 3}\right)' \text{vec}G^{(3)}(p,q) - T^{-1/2}\frac{1}{36}\left(\eta^{\otimes 3}\right)'\Omega^{(3)}(p,q)\left(\eta^{\otimes 3}\right).$$

Let $G^{(4)}(p,q)$ be an n_δ^4- dimensional vector whose $(j+(k-1)n_\delta+(l-1)n_\delta^2+(m-1)n_\delta^3)$-th element is given by $G^{(4)}_{jklm}(p,q)$. Let $\Omega^{(4)}(p,q)$ be an $n_\delta^4 - by - n_\delta^4$ matrix whose $(j+(k-1)n_\delta+(l-1)n_\delta^2+(m-1)n_\delta^3, n+(r-1)n_\delta+(s-1)n_\delta^2+(u-1)n_\delta^3)$-th element is given by $\omega^{(4)}_{jklmnrsu}(p,q;p,q)$. Define

$$\mathcal{W}^{(4)}(p,q,\eta) = T^{-1/2}\frac{1}{12}\left(\eta^{\otimes 4}\right)' \text{vec}G^{(4)}(p,q) - T^{-1}\frac{1}{576}\left(\eta^{\otimes 4}\right)'\Omega^{(4)}(p,q)\left(\eta^{\otimes 4}\right).$$

Then, the distribution of the $SupMLR(\Lambda_\epsilon)$ test can be approximated by:

$$S_\infty(\Lambda_\epsilon) \equiv \sup_{(p,q)\in\Lambda_\epsilon}\sup_{\eta\in R^{n_\delta}}\left\{\mathcal{W}^{(2)}(p,q,\eta)+\mathcal{W}^{(3)}(p,q,\eta)+\mathcal{W}^{(4)}(p,q,\eta)\right\}, \tag{5.6}$$

where Λ_ϵ is specified in (5.2). The following corollary is analogous to Corollary 1 in Qu and Zhuo (2015).

Corollary 1. *Under Assumptions 1-7 and the null hypothesis, we have:*

$$\Pr\left(SupMLR(\Lambda_\epsilon)\le s\right) - \Pr\left(S_\infty(\Lambda_\epsilon)\le s\right) \to 0,$$

over Λ_ϵ in (5.2).

Note that the above result holds irrespective of the model. This follows because the additional terms $\mathcal{W}^{(3)}(p, q, \eta)$ and $\mathcal{W}^{(4)}(p, q, \eta)$ both converge to zero as $T \to \infty$. These terms provide refinements in finite samples, having no effect asymptotically. The critical values can be obtained by following the simulation procedures described in section 5.4 of Qu and Zhuo (2015).

The illustrative model (cont'd). Consider the $ARMA(1, 1)$ model in (2.18)-(2.19) and assume α_{s_t} switches. This model can be written as:

$$m_t = \phi m_{t-1} + e_t + \theta e_{t-1}, \quad e_t \sim i.i.d.\ N(0, \sigma^2)$$
$$y_t = \alpha_{s_t} + m_t.$$

To illustrate the effects of the refined approximation, I simulate data using the $ARMA(1, 1)$ model with $T = 2000$, $\alpha_1 = \alpha_2 = 0$, $\phi = 0.9$, $\theta = -0.70$ and $\sigma = 0.2$. Then, for each simulated sample and fixed (p, q), I calculate $MLR(p, q)$, the approximation to $MLR(p, q)$ using only the second and fourth order terms in the Taylor expansion, and the approximation to $MLR(p, q)$ using the second order, fourth order and refinement terms in the Taylor expansion. After simulating 500 samples, I calculate both the correlation between $MLR(p, q)$ and its approximation using only the second and fourth order terms, and the correlation between $MLR(p, q)$ and its approximation using the second order, fourth order and refinement terms.

Table 1: Comparison of correlations between $MLR(p, q)$ statistic and original and refined approximations

	Between $MLR(p, q)$ and	
(p, q)	original approximation	refined approximation
$(0.90, 0.90)$	0.989	0.994
$(0.70, 0.90)$	0.217	0.843
$(0.50, 0.80)$	0.239	0.822

I also check these correlations for different p and q. Results are summarized in Table 1. The results show that including the refinement terms brings the approximation closer to the $MLR(p, q)$ statistic.

6 Monte Carlo

This section examines the size and power properties of the $SupMLR$ test statistics. The DGP is

$$m_t = \phi m_{t-1} + e_t + \theta e_{t-1}, \quad e_t \sim i.i.d.\ N(0, \sigma^2)$$
$$y_t = \alpha_{s_t} + m_t,$$

where y_t is observable and α_{s_t} switches with $p(s_t = 1|s_{t-1} = 1) = p$ and $p(s_t = 2|s_{t-1} = 2) = q$. Assign $\phi = 0.9$, $\theta = -0.70$ and $\sigma = 0.2$. The choice of this DGP is motivated by both the model studied in Perron (1993) and the empirical application in the next section. In this section, Λ_ϵ is specified as in (5.2) with $\epsilon = 0.01$. The critical values and rejection frequencies are all based on 3000 replications.

Table 2: Rejection frequencies under the null hypothesis

	Level	1.00	2.50	5.00	7.50	10.00
$T = 200$	$SupMLR(\Lambda_{0.01})$	1.20	3.67	7.40	10.20	14.23
$T = 500$	$SupMLR(\Lambda_{0.01})$	1.17	2.63	6.33	9.60	12.70

Table 2 reports the sizes of the $SupMLR(\Lambda_\epsilon)$ test statistic at five different nominal levels. Under the null hypothesis, we set $\alpha_1 = \alpha_2 = 0$. The rejection frequencies overall are close to the nominal levels with mild over-rejections in some cases. For example, the rejection rates at the 5% and 10% levels are 7.40% and 14.23% respectively, for $\epsilon = 0.01$ and sample size $T = 200$. Similar rejection rates are observed when $T = 500$.

For power properties, I set $\alpha_1 = -\tau$ and $\alpha_2 = \tau$, with $\tau = 0.05$, 0.10, 0.15, 0.20 and 0.25. The sample size $T = 500$ and the number of replications is 3000. Three pairs of values for (p, q) are considered: $(0.70, 0.70)$, $(0.70, 0.90)$ and $(0.90, 0.95)$.

Table 3: Rejection frequencies under the alternative hypotheses

(p, q)	τ	0.05	0.10	0.15	0.20	0.25
$(0.70, 0.70)$	$SupMLR(\Lambda_{0.01})$	6.33	6.67	21.00	74.33	99.67
$(0.70, 0.90)$	$SupMLR(\Lambda_{0.01})$	7.67	8.67	26.33	85.67	100
$(0.90, 0.95)$	$SupMLR(\Lambda_{0.01})$	5.47	6.13	17.70	69.57	97.13

Nominal level, 5%.

The rejection frequencies at 5% nominal levels are reported in Table 3. The power of the $SupMLR$ statistic increases consistently as the magnitude of $|\alpha_1 - \alpha_2|$ increases.

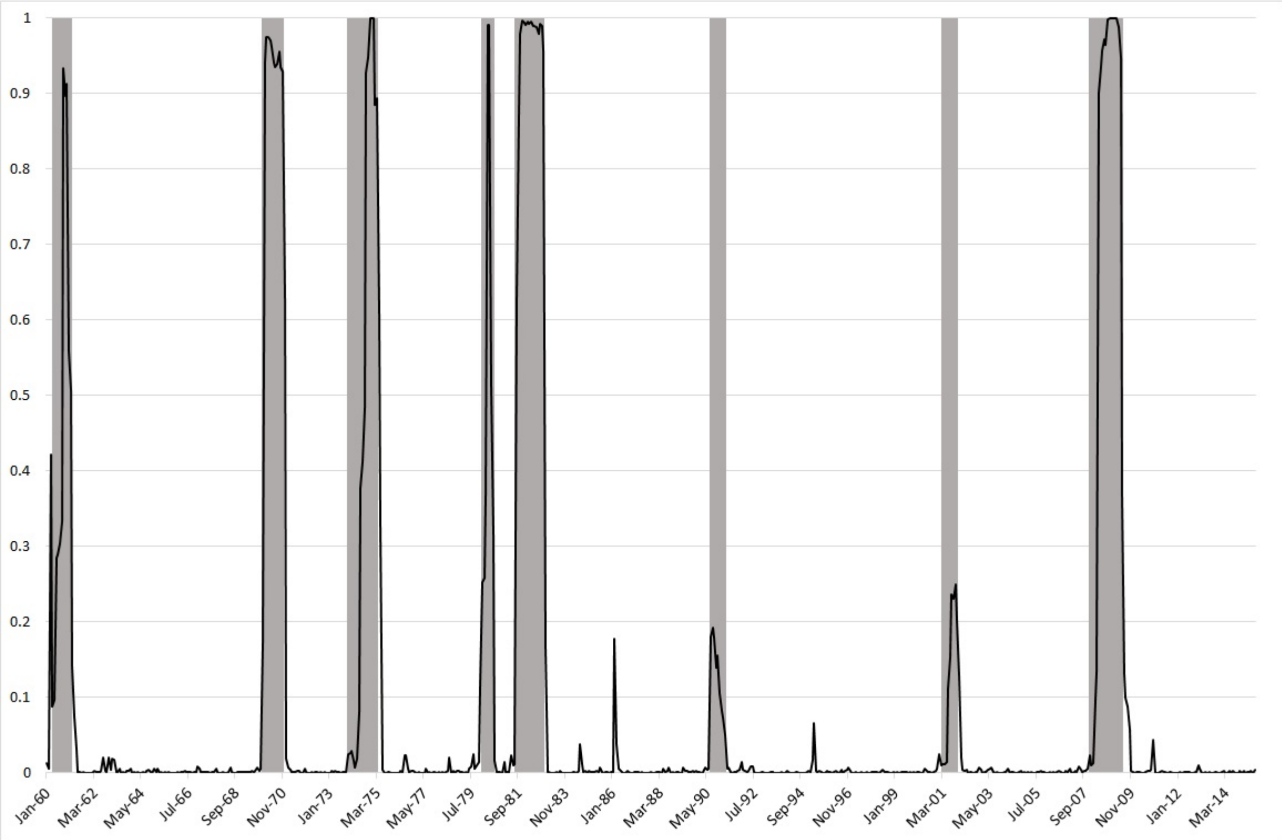
7 Application

In this section, I apply the MLR test developed in the preceding sections to study the changes in monthly U.S. unemployment rates. The data are from the labor force statistics reported in the Current Population Survey. The full sample is from January 1960 to July 2015. A simple $ARMA(1,1)$ model as in (2.18) and (2.19) is considered, i.e.

$$
\begin{aligned}
m_t &= \phi m_{t-1} + e_t + \theta e_{t-1}, \quad e_t \sim i.i.d.\ N(0, \sigma^2) \\
y_t &= \alpha_{s_t} + m_t.
\end{aligned}
$$

Here, α_{s_t} switches and indicates the mean level of change in the unemployment rate at time t. $SupMLR(\Lambda_{0.01})$ equals 36.99 for the full sample with the critical value being 8.83 at the 5% level. The test statistic therefore provides strong evidence favoring the regime switching specification. To provide some further evidence for the relevance of the regime switching specification, I estimate the regimes implied by the model. Let $s_t = 1$ and $s_t = 2$ represent the tight and slack labor market regimes, respectively. Here the tight and slack labor market regimes are defined from the perspective of the job-seekers. Estimation results show that the regime shifts closely follow the recession and expansion periods dated by the National Bureau of Economic Research (NBER) and shown in Figure 2. Comparing the smoothed regime probabilities and the dates of the recession and expansion periods, I find that the labor market normally takes time to react at the beginning of a recession. Estimation under two regimes specification also provides detailed information on changes in the labor market and the durations of each regime. In the tight labor market regime, the unemployment rate increases by 0.26% per month on average and this regime lasts about 8.8 months. In the slack labor market regime, the unemployment rate decreases by 0.03% per month and the regime lasts about 79.4 months. The model assigns low probabilities, around $20 - 30\%$, to the tight regime during the relatively shallow recessions of July 1990 to March 1991 and March 2001 to November 2001. This makes sense, because the increases in the unemployment rate during these two recessions are relatively moderate when compared with other recessions, such as the recent Great Recession from December 2007 to June 2009.

Figure 2. Smoothed probabilities of in tight labor market regime



Note. The shaded areas correspond to the NBER defined recessions. The solid line indicates the smoothed probabilities of being in the tight labor market regime.

8 Conclusion

This paper develops a modified likelihood ratio (MLR) based test for detecting regime switching in state space models. The asymptotic distribution of this test statistics is also established. When applied to changes in U.S. monthly unemployment rates, the test finds strong evidence favoring the regime switching specification. This paper is the first to develop a test that is based on the likelihood ratio principle for detecting regime switching in state space models. The techniques developed in this paper can have implications for hypothesis testing in more general contexts, such as testing for regime switching in state space models with multiple observables.

References

Billingsley, P. (1999): *Convergence of Probability Measures*, John Willey & Sons.

Carrasco, M., L. Hu, and W. Ploberger (2014): "Optimal Test for Markov Switching Parameters," *Econometrica*, 82, 765–784.

Carter, A. V. and D. G. Steigerwald (2012): "Testing for Regime Switching: A Comment," *Econometrica*, 80, 1809–1812.

Chesher, A. D. (1984): "Testing for Neglected Heterogeneity," *Econometrica*, 52, 865–872.

Cho, J. S. and H. White (2007): "Testing for Regime Switching," *Econometrica*, 75, 1671–1720.

Davidson, R. and J. MacKinnon (1991): "Une nouvelle forme du test de la matrice d'information," *Annales d'Economie et de Statistique*, 171–192.

Davies, R. B. (1977): "Hypothesis Testing When a Nuisance Parameter Is Present Only under the Alternative," *Biometrika*, 64, 247–254.

Garcia, R. (1998): "Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models," *International Economic Review*, 39, 763–788.

Gordon, K. and A. Smith (1988): "Modeling and Monitoring Discontinuous Changes in Time Series," in *Bayesian Analysis of Time Series and Dynamic Linear models*, ed. by J. Spall, New York: Marcel Dekker, 359–392.

Hamilton, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

——— (1990): "Analysis of Time Series Subject to Changes in Regime," *Journal of Econometrics*, 45, 39–70.

——— (1994): *Time Series Analysis*, Princeton, NJ: Princeton University Press.

——— (1996): "Specification Testing in Markov-Switching Time-Series Models," *Journal of Econometrics*, 70, 127–157.

——— (2008): "Regime-Switching Models," in *New Palgrave Dictionary of Economics*, ed. by S. Durlauf and L. Blume, London: Palgrave Macmillan.

Hansen, B. E. (1992): "The Likelihood Ratio Test under Non-Standard Conditions: Testing the Markov Switching Model of GNP," *Journal of Applied Econometrics*, 7, S61–S82.

Harrison, P. J. and C. F. Stevens (1976): "Bayesian Forecasting," *Journal of the Royal Statistical Society. Series B (Methodological)*, 38, 205–247.

Harvey, A. C. (1981): *Time Series Models*, Oxford, UK: Philip Allan and Humanities Press.

Highfield, R. A. (1990): "Bayesian Approaches to Turning Point Prediction," in *Proceedings of the Business and Economics Section*, Washington, DC: American Statistical Association, 89–98.

Kalman, R. E. (1960): "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, 82, 35–45.

Kim, C.-J. and C. R. Nelson (1999): *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, Cambridge, MA: The MIT Press.

Lancaster, T. (1984): "The Covariance Matrix of the Information Matrix Test," *Econometrica*, 52, 1051–1053.

Meinhold, R. J. and N. D. Singpurwalla (1983): "Understanding the Kalman Filter," *The American Statistician*, 37, 123–127.

Neyman, J. and E. Scott (1966): "On the Use of C(a) Optimal Tests of Composite Hypotheses," *Bull. Inst. Int. Statist.*, 41, 477–497.

Perron, P. (1993): "Non-stationarities and Non-linearities in Canadian Inflation," in *Economic Behaviour and Policy Choice under Price Stability : Proceedings of a Conference Held at the Bank of Canada.*

Qu, Z. and F. Zhuo (2015): "Likelihood Ratio Based Tests for Markov Regime Switching," *Working Paper.*

Appendix

A Derivatives of the density function

The derivatives of f_{it} in (2.16) are calculated here. By (2.5), (2.6), (2.7), (2.8), (2.14) and (4.2), we have:

$$f_{it} = \left[2\pi C^{(i)}_{t|t-1}\right]^{-1/2} \exp\left\{-\frac{\left[\mu^{(i)}_{t|t-1}\right]^2}{2C^{(i)}_{t|t-1}}\right\},$$

where

$$\mu^{(i)}_{t|t-1} = y_t - H_i'\left(G_i + F_i x_{t-1|t-1}\right) - A_i' z_t$$

and

$$C^{(i)}_{t|t-1} = H_i'\left(F_i P_{t-1|t-1} F_i' + Q_i\right) H_i.$$

Then the first order derivative of $\mu^{(i)}_{t|t-1}$ with respect to the j-th component in θ is

$$\nabla_{\theta_j}\mu^{(i)}_{t|t-1} = -\nabla_{\theta_j} H_i'\left(G_i + F_i x_{t-1|t-1}\right) - H_i'\left(\nabla_{\theta_j} G_i + \nabla_{\theta_j} F_i x_{t-1|t-1} + F_i \nabla_{\theta_j} x_{t-1|t-1}\right) - \nabla_{\theta_j} A_i' z_t. \tag{A.1}$$

and the first order derivative of $C^{(i)}_{t|t-1}$ with respect to the j-th component in θ is

$$\begin{aligned}\nabla_{\theta_j} C^{(i)}_{t|t-1} &= \nabla_{\theta_j} H_i'\left(F_i P_{t-1|t-1} F_i' + Q_i\right) H_i + H_i'\left(F_i P_{t-1|t-1} F_i' + Q_i\right)\nabla_{\theta_j} H_i \\ &\quad + H_i'\left(\nabla_{\theta_j} F_i P_{t-1|t-1} F_i' + F_i \nabla_{\theta_j} P_{t-1|t-1} F_i' + F_i P_{t-1|t-1} \nabla_{\theta_j} F_i' + \nabla_{\theta_j} Q_i\right) H_i.\end{aligned} \tag{A.2}$$

Then the first order derivative of f_{it} with respect to the j-th component in θ is

$$\nabla_{\theta_j} f_{it} = -f_{it}\left(\frac{\mu^{(i)}_{t|t-1}}{C^{(i)}_{t|t-1}}\right)\left(\nabla_{\theta_j}\mu^{(i)}_{t|t-1}\right) + f_{it}\left(\frac{1}{2C^{(i)}_{t|t-1}}\right)\left(\frac{\left[\mu^{(i)}_{t|t-1}\right]^2}{C^{(i)}_{t|t-1}} - 1\right)\left(\nabla_{\theta_j} C^{(i)}_{t|t-1}\right).$$

Similarly, the derivative of $\nabla_{\theta_j}\mu^{(i)}_{t|t-1}$ respect to the k-th component in θ is

$$\begin{aligned}\nabla_{\theta_j}\nabla_{\theta_k}\mu^{(i)}_{t|t-1} &= -\nabla_{\theta_j}\nabla_{\theta_k}H_i'\left(G_i + F_i x_{t-1|t-1}\right) - \nabla_{\theta_j}H_i'\left(\nabla_{\theta_k}G_i + \nabla_{\theta_k}F_i x_{t-1|t-1} + F_i\nabla_{\theta_k}x_{t-1|t-1}\right)\\ &- \nabla_{\theta_k}H_i'\left(\nabla_{\theta_j}G_i + \nabla_{\theta_j}F_i x_{t-1|t-1} + F_i\nabla_{\theta_j}x_{t-1|t-1}\right)\\ &- H_i'\left(\nabla_{\theta_j}\nabla_{\theta_k}G_i + \nabla_{\theta_j}\nabla_{\theta_k}F_i x_{t-1|t-1} + \nabla_{\theta_j}F_i\nabla_{\theta_k}x_{t-1|t-1} + F_i\nabla_{\theta_j}\nabla_{\theta_k}x_{t-1|t-1}\right)\\ &- \nabla_{\theta_j}\nabla_{\theta_k}A_i' z_t. \end{aligned} \tag{A.3}$$

The derivative of $\nabla_{\theta_j}C^{(i)}_{t|t-1}$ respect to the k-th component in θ is

$$\begin{aligned}\nabla_{\theta_j}\nabla_{\theta_k}C^{(i)}_{t|t-1} &= \nabla_{\theta_j}\nabla_{\theta_k}H_i'\left(F_iP_{t-1|t-1}F_i' + Q_i\right)H_i + \nabla_{\theta_j}H_i'\left(F_iP_{t-1|t-1}F_i' + Q_i\right)\nabla_{\theta_k}H_i\\ &\nabla_{\theta_j}H_i'\left(\nabla_{\theta_k}F_iP_{t-1|t-1}F_i' + F_i\nabla_{\theta_k}P_{t-1|t-1}F_i' + F_iP_{t-1|t-1}\nabla_{\theta_k}F_i' + \nabla_{\theta_k}Q_i\right)H_i\\ &+ \nabla_{\theta_k}H_i'\left(F_iP_{t-1|t-1}F_i' + Q_i\right)\nabla_{\theta_j}H_i + H_i'\left(F_iP_{t-1|t-1}F_i' + Q_i\right)\nabla_{\theta_j}\nabla_{\theta_k}H_i\\ &+ H_i'\left(\nabla_{\theta_k}F_iP_{t-1|t-1}F_i' + F_i\nabla_{\theta_k}P_{t-1|t-1}F_i' + F_iP_{t-1|t-1}\nabla_{\theta_k}F_i' + \nabla_{\theta_k}Q_i\right)\nabla_{\theta_j}H_i\\ &+ \nabla_{\theta_k}H_i'\left(\nabla_{\theta_j}F_iP_{t-1|t-1}F_i' + F_i\nabla_{\theta_j}P_{t-1|t-1}F_i' + F_iP_{t-1|t-1}\nabla_{\theta_j}F_i' + \nabla_{\theta_j}Q_i\right)H_i\\ &+ H_i'\left(\nabla_{\theta_j}F_iP_{t-1|t-1}F_i' + F_i\nabla_{\theta_j}P_{t-1|t-1}F_i' + F_iP_{t-1|t-1}\nabla_{\theta_j}F_i' + \nabla_{\theta_j}Q_i\right)\nabla_{\theta_k}H_i\\ &+ H_i'\left(\nabla_{\theta_j}\nabla_{\theta_k}F_iP_{t-1|t-1}F_i' + \nabla_{\theta_j}F_i\nabla_{\theta_k}P_{t-1|t-1}F_i' + \nabla_{\theta_j}F_iP_{t-1|t-1}\nabla_{\theta_k}F_i'\right)H_i\\ &+ H_i'\left(\nabla_{\theta_k}F_i\nabla_{\theta_j}P_{t-1|t-1}F_i' + F_i\nabla_{\theta_j}\nabla_{\theta_k}P_{t-1|t-1}F_i' + F_i\nabla_{\theta_j}P_{t-1|t-1}\nabla_{\theta_k}F_i'\right)H_i\\ &+ H_i'\left(\nabla_{\theta_k}F_iP_{t-1|t-1}\nabla_{\theta_j}F_i' + F_i\nabla_{\theta_k}P_{t-1|t-1}\nabla_{\theta_j}F_i' + F_iP_{t-1|t-1}\nabla_{\theta_j}\nabla_{\theta_k}F_i'\right)H_i\\ &+ H_i'\left(\nabla_{\theta_j}\nabla_{\theta_k}Q_i\right)H_i. \end{aligned} \tag{A.4}$$

Then the derivative of $\nabla_{\theta_j} f_{it}$ respect to the k-th component in θ is

$$\begin{aligned}\nabla_{\theta_j}\nabla_{\theta_k} f_{it} =& -\left(\nabla_{\theta_k} f_{it}\right)\left(\frac{\mu^{(i)}_{t|t-1}}{C^{(i)}_{t|t-1}}\right)\left(\nabla_{\theta_j}\mu^{(i)}_{t|t-1}\right) - f_{it}\left(\frac{\mu^{(i)}_{t|t-1}}{C^{(i)}_{t|t-1}}\right)\left(\nabla_{\theta_j}\nabla_{\theta_k}\mu^{(i)}_{t|t-1}\right)\\
&- f_{it}\left[\frac{\nabla_{\theta_k}\mu^{(i)}_{t|t-1}}{C^{(i)}_{t|t-1}} - \frac{\mu^{(i)}_{t|t-1}\left(\nabla_{\theta_k} C^{(i)}_{t|t-1}\right)}{\left[C^{(i)}_{t|t-1}\right]^2}\right]\left(\nabla_{\theta_j}\mu^{(i)}_{t|t-1}\right)\\
&+ \left(\nabla_{\theta_k} f_{it}\right)\left(\frac{1}{2C^{(i)}_{t|t-1}}\right)\left(\frac{\left[\mu^{(i)}_{t|t-1}\right]^2}{C^{(i)}_{t|t-1}} - 1\right)\left(\nabla_{\theta_j} C^{(i)}_{t|t-1}\right)\\
&- f_{it}\left(\frac{\nabla_{\theta_k} C^{(i)}_{t|t-1}}{2\left[C^{(i)}_{t|t-1}\right]^2}\right)\left(\frac{\left[\mu^{(i)}_{t|t-1}\right]^2}{C^{(i)}_{t|t-1}} - 1\right)\left(\nabla_{\theta_j} C^{(i)}_{t|t-1}\right)\\
&+ f_{it}\left(\frac{1}{2C^{(i)}_{t|t-1}}\right)\left(\frac{2\left[\mu^{(i)}_{t|t-1}\right]\left(\nabla_{\theta_k}\mu^{(i)}_{t|t-1}\right)}{C^{(i)}_{t|t-1}} - \frac{\left[\mu^{(i)}_{t|t-1}\right]^2\left(\nabla_{\theta_k} C^{(i)}_{t|t-1}\right)}{\left[C^{(i)}_{t|t-1}\right]^2}\right)\left(\nabla_{\theta_j} C^{(i)}_{t|t-1}\right)\\
&+ f_{it}\left(\frac{1}{2C^{(i)}_{t|t-1}}\right)\left(\frac{\left[\mu^{(i)}_{t|t-1}\right]^2}{C^{(i)}_{t|t-1}} - 1\right)\left(\nabla_{\theta_j}\nabla_{\theta_k} C^{(i)}_{t|t-1}\right).\end{aligned}$$

B Proofs

Proof of Lemma 1. The equation (4.1) can be written as

$$\xi_{t+1|t} = p + \rho\frac{A_t}{B_t}, \tag{B.1}$$

where $A_t = f_{2t}(\xi_{t|t-1} - 1)$ and $B_t = (f_{1t} - f_{2t})\xi_{t|t-1} + f_{2t}$. Let "-" (e.g. $\bar{\xi}_{t|t-1}$) denote that the quantity is evaluated at $(\beta', \delta', \delta')$.

Consider Lemma 1.1. Because $\bar{f}_{1t} = \bar{f}_{2t} = \bar{f}_t$, it follows that

$$\bar{A}_t = \bar{f}_t(\bar{\xi}_{t|t-1} - 1) \qquad \text{and} \qquad \bar{B}_t = \bar{f}_t. \tag{B.2}$$

Plugging this into (B.1), we have

$$\bar{\xi}_{t+1|t} = p + \rho(\bar{\xi}_{t|t-1} - 1).$$

This, together with (2.13), implies $\bar{\xi}_{2|1} = p + \rho(\bar{\xi}_{1|0} - 1) = p + \rho(\xi_* - 1) = \xi_*$, where the last equality follows from the definition of ρ and ξ_*. This can be iterated forward, leading to $\bar{\xi}_{t+1|t} = \xi_*$ for all $t \geq 1$.

Consider Lemma 1.2. Differentiate (B.1) with respect to the j-th component in θ, we have

$$\nabla_{\theta_j}\xi_{t+1|t} = \rho\left\{\frac{\nabla_{\theta_j}A_t}{B_t} - \frac{A_t\nabla_{\theta_j}B_t}{B_t^2}\right\}, \tag{B.3}$$

where

$$\nabla_{\theta_j}A_t = \nabla_{\theta_j}f_{2t}(\xi_{t|t-1}-1) + f_{2t}\nabla_{\theta_j}\xi_{t|t-1}$$

and

$$\nabla_{\theta_j}B_t = (\nabla_{\theta_j}f_{1t} - \nabla_{\theta_j}f_{2t})\xi_{t|t-1} + (f_{1t}-f_{2t})\nabla_{\theta_j}\xi_{t|t-1} + \nabla_{\theta_j}f_{2t}.$$

Below, we evaluate the right hand side of (B.3) at (β',δ',δ') for two possible situations:

(1). If $j\in I_0$, then $\nabla_{\theta_j}\bar{f}_{1t} = \nabla_{\theta_j}\bar{f}_{2t}$ and $\bar{f}_{1t}=\bar{f}_{2t}=\bar{f}_t$. Consequently

$$\begin{aligned}\nabla_{\theta_j}\bar{A}_t &= \nabla_{\theta_j}\bar{f}_{2t}(\xi_*-1) + \bar{f}_t\nabla_{\theta_j}\bar{\xi}_{t|t-1},\\ \nabla_{\theta_j}\bar{B}_t &= \nabla_{\theta_j}\bar{f}_{2t}.\end{aligned} \tag{B.4}$$

Combining (B.4) with (B.2), we have $\nabla_{\theta_j}\bar{\xi}_{t+1|t} = \rho\nabla_{\theta_j}\bar{\xi}_{t|t-1}$. This implies, at $t=1$, we have $\nabla_{\theta_j}\bar{\xi}_{2|1} = \rho\nabla_{\theta_j}\bar{\xi}_{1|0} = \rho\nabla_{\theta_j}\xi_* = 0$. This can be iterated forward leading to $\nabla_{\theta_j}\bar{\xi}_{t|t-1}=0$.

(2). If $j\in I_1$ or $j\in I_2$, then $\bar{f}_{1t}=\bar{f}_{2t}=\bar{f}_t$ and

$$\begin{aligned}\nabla_{\theta_j}\bar{A}_t &= \nabla_{\theta_j}\bar{f}_{2t}(\xi_*-1) + \bar{f}_t\nabla_{\theta_j}\bar{\xi}_{t|t-1},\\ \nabla_{\theta_j}\bar{B}_t &= \xi_*\nabla_{\theta_j}\bar{f}_{1t} + (1-\xi_*)\nabla_{\theta_j}\bar{f}_{2t}.\end{aligned} \tag{B.5}$$

Combining this with (B.2), we have

$$\begin{aligned}\nabla_{\theta_j}\bar{\xi}_{t+1|t} &= \rho\left\{\nabla_{\theta_j}\bar{\xi}_{t|t-1} - (\xi_*-1)\xi_*\left(\frac{\nabla_{\theta_j}\bar{f}_{1t}}{\bar{f}_t} - \frac{\nabla_{\theta_j}\bar{f}_{2t}}{\bar{f}_t}\right)\right\}\\ &= \rho\nabla_{\theta_j}\bar{\xi}_{t|t-1} + r\left(\frac{\nabla_{\theta_j}\bar{f}_{1t}}{\bar{f}_t} - \frac{\nabla_{\theta_j}\bar{f}_{2t}}{\bar{f}_t}\right),\end{aligned}$$

where $r=\rho(1-\xi_*)\xi_*$. Note that $\nabla_{\theta_j}\bar{\xi}_{t+1|t}$ can also be written as

$$\nabla_{\theta_j}\bar{\xi}_{t+1|t} = r\sum_{s=0}^{t-1}\rho^s\left(\frac{\nabla_{\theta_j}\bar{f}_{1t-s}}{\bar{f}_t} - \frac{\nabla_{\theta_j}\bar{f}_{2t-s}}{\bar{f}_t}\right). \tag{B.6}$$

Because

$$\left(\frac{\nabla_{\theta_j}\bar{f}_{1t}}{\bar{f}_t} - \frac{\nabla_{\theta_j}\bar{f}_{2t}}{\bar{f}_t}\right) = -\left(\frac{\nabla_{\theta_{j-n_\delta}}\bar{f}_{1t}}{\bar{f}_t} - \frac{\nabla_{\theta_{j-n_\delta}}\bar{f}_{2t}}{\bar{f}_t}\right), \tag{B.7}$$

when $j \in I_2$, we have

$$\nabla_{\theta_j}\bar{\xi}_{t+1|t} = -\nabla_{\theta_{j-n_\delta}}\bar{\xi}_{t+1|t}.$$

In addition, from (2.13) and (B.6), we have

$$\begin{aligned}\nabla_{\theta_j}\bar{\xi}_{t|t} &= (1-\xi_*)\xi_*\sum_{s=0}^{t-1}\rho^s\left(\frac{\nabla_{\theta_j}\bar{f}_{1t-s}}{\bar{f}_t} - \frac{\nabla_{\theta_j}\bar{f}_{2t-s}}{\bar{f}_t}\right) \\ &= \rho\nabla_{\theta_j}\bar{\xi}_{t-1|t-1} + (1-\xi_*)\xi_*\left(\frac{\nabla_{\theta_j}\bar{f}_{1t}}{\bar{f}_t} - \frac{\nabla_{\theta_j}\bar{f}_{2t}}{\bar{f}_t}\right),\end{aligned} \tag{B.8}$$

when $j \in I_a,\ a = 1, 2$, and

$$\nabla_{\theta_j}\bar{\xi}_{t|t} = -\nabla_{\theta_{j-n_\delta}}\bar{\xi}_{t|t}.$$

Consider Lemma 1.3. Differentiating (B.3) with respect to θ_k:

$$\nabla_{\theta_j}\nabla_{\theta_k}\xi_{t+1|t} = \rho\left\{\frac{\nabla_{\theta_j}\nabla_{\theta_k}A_t}{B_t} - \frac{\nabla_{\theta_j}A_t\nabla_{\theta_k}B_t}{B_t^2} - \frac{\nabla_{\theta_k}A_t\nabla_{\theta_j}B_t}{B_t^2} - \frac{A_t\nabla_{\theta_j}\nabla_{\theta_k}B_t}{B_t^2} + 2\frac{A_t\nabla_{\theta_j}B_t\nabla_{\theta_k}B_t}{B_t^3}\right\}, \tag{B.9}$$

where

$$\begin{aligned}\nabla_{\theta_j}\nabla_{\theta_k}A_t &= \nabla_{\theta_j}\nabla_{\theta_k}f_{2t}(\xi_{t|t-1}-1) + \nabla_{\theta_j}f_{2t}\nabla_{\theta_k}\xi_{t|t-1} + \nabla_{\theta_k}f_{2t}\nabla_{\theta_j}\xi_{t|t-1} + f_{2t}\nabla_{\theta_j}\nabla_{\theta_k}\xi_{t|t-1}, \\ \nabla_{\theta_j}\nabla_{\theta_k}B_t &= (\nabla_{\theta_j}\nabla_{\theta_k}f_{1t} - \nabla_{\theta_j}\nabla_{\theta_k}f_{2t})\xi_{t|t-1} + (\nabla_{\theta_j}f_{1t} - \nabla_{\theta_j}f_{2t})\nabla_{\theta_k}\xi_{t|t-1} \\ &\quad + (\nabla_{\theta_k}f_{1t} - \nabla_{\theta_k}f_{2t})\nabla_{\theta_j}\xi_{t|t-1} + (f_{1t}-f_{2t})\nabla_{\theta_j}\nabla_{\theta_k}\xi_{t|t-1} + \nabla_{\theta_j}\nabla_{\theta_k}f_{2t}.\end{aligned}$$

We now evaluate the right hand side of (B.9) at $\delta_1 = \delta_2 = \delta$ under three possible situations:

(1) If $j \in I_0$ and $k \in I_0$, then $\bar{f}_{1t} = \bar{f}_{2t} = \bar{f}_t$, $\nabla_{\theta_j}\bar{f}_{1t} = \nabla_{\theta_j}\bar{f}_{2t}$, $\nabla_{\theta_k}\bar{f}_{1t} = \nabla_{\theta_k}\bar{f}_{2t}$, $\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{1t} = \nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}$ and $\nabla_{\theta_j}\bar{\xi}_{t+1|t} = \nabla_{\theta_k}\bar{\xi}_{t+1|t} = 0$, implying $\nabla_{\theta_j}\nabla_{\theta_k}\bar{A}_t = \nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}(\bar{\xi}_{t|t-1}-1) + \bar{f}_t\nabla_{\theta_j}\nabla_{\theta_k}\bar{\xi}_{t|t-1}$ and $\nabla_{\theta_j}\nabla_{\theta_k}\bar{B}_t = \nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}$. Combining them with (B.4) and (B.2), $\nabla_{\theta_j}\nabla_{\theta_k}\bar{\xi}_{t+1|t}$ equals

$$\begin{aligned}&\rho\left\{\frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}(\bar{\xi}_{t|t-1}-1)+\bar{f}_t\nabla_{\theta_j}\nabla_{\theta_k}\bar{\xi}_{t|t-1}}{\bar{f}_t} - \frac{(\bar{\xi}_{t|t-1}-1)\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{f}_{2t}}{\bar{f}_t^2}\right. \\ &\left.-\frac{(\bar{\xi}_{t|t-1}-1)\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_j}\bar{f}_{2t}}{\bar{f}_t^2} - \frac{(\bar{\xi}_{t|t-1}-1)\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}}{\bar{f}_t} + 2\frac{(\bar{\xi}_{t|t-1}-1)\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{f}_{2t}}{\bar{f}_t^2}\right\} \\ &= \rho\nabla_{\theta_j}\nabla_{\theta_k}\bar{\xi}_{t|t-1}.\end{aligned}$$

Starting at $t = 1$ and iterating forward, we have $\nabla_{\theta_j}\nabla_{\theta_k}\bar{\xi}_{t+1|t} = 0$ for all $t \geq 1$.

(2) If $j \in I_0$ and $k \in I_1$, then $\nabla_{\theta_j}\bar{f}_{1t} = \nabla_{\theta_j}\bar{f}_{2t}$ and $\nabla_{\theta_j}\bar{\xi}_{t+1|t} = 0$, which imply that

$$\nabla_{\theta_j}\nabla_{\theta_k}\bar{A}_t = \nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}(\xi_* - 1) + \nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{\xi}_{t|t-1} + \bar{f}_t\nabla_{\theta_j}\nabla_{\theta_k}\bar{\xi}_{t|t-1}$$

and

$$\nabla_{\theta_j}\nabla_{\theta_k}\bar{B}_t = \xi_*\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{1t} + (1-\xi_*)\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}.$$

Combing these two equations with (B.2), (B.4) and (B.5), $\nabla_{\theta_j}\nabla_{\theta_k}\bar{\xi}_{t+1|t}$ equals

$$\begin{aligned}&\rho\Bigg\{\frac{1}{\bar{f}_t}\left[\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}(\xi_*-1) + \nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{\xi}_{t|t-1} + \bar{f}_t\nabla_{\theta_j}\nabla_{\theta_k}\bar{\xi}_{t|t-1}\right] - \frac{1}{\bar{f}_t^2}\left[\left(\nabla_{\theta_j}\bar{f}_{2t}(\xi_*-1)\right)\left(\xi_*\nabla_{\theta_k}\bar{f}_{1t} + (1-\xi_*)\nabla_{\theta_k}\bar{f}_{2t}\right)\right]\\
&-\frac{1}{\bar{f}_t^2}\left[\nabla_{\theta_j}\bar{f}_{2t}\left(\nabla_{\theta_k}\bar{f}_{2t}(\xi_*-1) + \bar{f}_t\nabla_{\theta_k}\bar{\xi}_{t|t-1}\right)\right] - (\xi_*-1)\frac{1}{\bar{f}_t}\left(\xi_*\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{1t} + (1-\xi_*)\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}\right)\\
&+2(\xi_*-1)\frac{1}{\bar{f}_t^2}\left[\nabla_{\theta_j}\bar{f}_{2t}\left(\xi_*\nabla_{\theta_k}\bar{f}_{1t} + (1-\xi_*)\nabla_{\theta_k}\bar{f}_{2t}\right)\right]\Bigg\}.\end{aligned}$$

The result follows from rearranging the terms. For the case $j \in I_0$ and $k \in I_2$, we have the same result.

(3) If $j \in I_1$ and $k \in I_1$, then

$$\nabla_{\theta_j}\nabla_{\theta_k}\bar{A}_t = \nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}(\xi_*-1) + \nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{\xi}_{t|t-1} + \nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_j}\bar{\xi}_{t|t-1} + \bar{f}_t\nabla_{\theta_j}\nabla_{\theta_k}\bar{\xi}_{t|t-1}$$

and

$$\nabla_{\theta_j}\nabla_{\theta_k}\bar{B}_t = \xi_*\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{1t} + (1-\xi_*)\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t} + (\nabla_{\theta_j}\bar{f}_{1t} - \nabla_{\theta_j}\bar{f}_{2t})\nabla_{\theta_k}\bar{\xi}_{t|t-1} + (\nabla_{\theta_k}\bar{f}_{1t} - \nabla_{\theta_k}\bar{f}_{2t})\nabla_{\theta_j}\bar{\xi}_{t|t-1}.$$

Applying the similar derivative above, we have $\nabla_{\theta_j}\nabla_{\theta_k}\bar{\xi}_{t+1|t}$ equals

$$\begin{aligned}&\rho\Bigg\{\frac{1}{\bar{f}_t}\left[\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}(\xi_*-1) + \nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{\xi}_{t|t-1} + \nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_j}\bar{\xi}_{t|t-1} + \nabla_{\theta_j}\nabla_{\theta_k}\bar{\xi}_{t|t-1}\right]\\
&-\frac{1}{\bar{f}_t^2}\left[\left(\nabla_{\theta_j}\bar{f}_{2t}(\xi_*-1) + \bar{f}_t\nabla_{\theta_j}\bar{\xi}_{t|t-1}\right)\left(\xi_*\nabla_{\theta_k}\bar{f}_{1t} + (1-\xi_*)\nabla_{\theta_k}\bar{f}_{2t}\right)\right]\\
&-\frac{1}{\bar{f}_t^2}\left[\left(\xi_*\nabla_{\theta_j}\bar{f}_{1t} + (1-\xi_*)\nabla_{\theta_j}\bar{f}_{2t}\right)\left(\nabla_{\theta_k}\bar{f}_{2t}(\xi_*-1) + \bar{f}_t\nabla_{\theta_k}\bar{\xi}_{t|t-1}\right)\right]\\
&-(\xi_*-1)\frac{1}{\bar{f}_t}\left(\xi_*\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{1t} + (1-\xi_*)\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t} + (\nabla_{\theta_j}\bar{f}_{1t} - \nabla_{\theta_j}\bar{f}_{2t})\nabla_{\theta_k}\bar{\xi}_{t|t-1} + (\nabla_{\theta_k}\bar{f}_{1t} - \nabla_{\theta_k}\bar{f}_{2t})\nabla_{\theta_j}\bar{\xi}_{t|t-1}\right)\\
&+2(\xi_*-1)\frac{1}{\bar{f}_t^2}\left[\left(\xi_*\nabla_{\theta_j}\bar{f}_{1t} + (1-\xi_*)\nabla_{\theta_j}\bar{f}_{2t}\right)\left(\xi_*\nabla_{\theta_k}\bar{f}_{1t} + (1-\xi_*)\nabla_{\theta_k}\bar{f}_{2t}\right)\right]\Bigg\}.\end{aligned}$$

The result follows from rearranging the terms. For the cases $j \in I_1$ and $k \in I_2$, and $j \in I_2$ and $k \in I_2$, we have the same result.

Consider Lemma 1.4. Differentiating (B.9) with respect to θ_l:

$$
\begin{aligned}
&\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}\xi_{t+1|t} \\
&= \rho\Bigg\{\frac{\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}A_t}{B_t} - \frac{\nabla_{\theta_j}\nabla_{\theta_k}A_t\nabla_{\theta_l}B_t}{B_t^2} - \frac{\nabla_{\theta_j}\nabla_{\theta_l}A_t\nabla_{\theta_k}B_t}{B_t^2} - \frac{\nabla_{\theta_j}A_t\nabla_{\theta_k}\nabla_{\theta_l}B_t}{B_t^2} + \frac{2\nabla_{\theta_j}A_t\nabla_{\theta_k}B_t\nabla_{\theta_l}B_t}{B_t^3} \\
&- \frac{\nabla_{\theta_k}\nabla_{\theta_l}A_t\nabla_{\theta_j}B_t}{B_t^2} - \frac{\nabla_{\theta_k}A_t\nabla_{\theta_j}\nabla_{\theta_l}B_t}{B_t^2} + \frac{2\nabla_{\theta_k}A_t\nabla_{\theta_j}B_t\nabla_{\theta_l}B_t}{B_t^3} \\
&- \frac{\nabla_{\theta_l}A_t\nabla_{\theta_j}\nabla_{\theta_k}B_t}{B_t^2} - \frac{A_t\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}B_t}{B_t^2} + \frac{2A_t\nabla_{\theta_j}\nabla_{\theta_k}B_t\nabla_{\theta_l}B_t}{B_t^3} \\
&+ \frac{2\nabla_{\theta_l}A_t\nabla_{\theta_j}B_t\nabla_{\theta_k}B_t}{B_t^3} + \frac{2A_t\nabla_{\theta_j}\nabla_{\theta_l}B_t\nabla_{\theta_k}B_t}{B_t^3} + \frac{2A_t\nabla_{\theta_j}B_t\nabla_{\theta_k}\nabla_{\theta_l}B_t}{B_t^3} - \frac{6A_t\nabla_{\theta_j}B_t\nabla_{\theta_k}B_t\nabla_{\theta_l}B_t}{B_t^4}\Bigg\},
\end{aligned}
$$

where

$$
\begin{aligned}
\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}A_t &= \nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}f_{2t}(\xi_{t|t-1}-1) + \nabla_{\theta_j}\nabla_{\theta_l}f_{2t}\nabla_{\theta_k}\xi_{t|t-1} + \nabla_{\theta_k}\nabla_{\theta_l}f_{2t}\nabla_{\theta_j}\xi_{t|t-1} \\
&+ \nabla_{\theta_l}f_{2t}\nabla_{\theta_j}\nabla_{\theta_k}\xi_{t|t-1} + \nabla_{\theta_j}\nabla_{\theta_k}f_{2t}\nabla_{\theta_l}\xi_{t|t-1} + \nabla_{\theta_j}f_{2t}\nabla_{\theta_k}\nabla_{\theta_l}\xi_{t|t-1} \\
&+ \nabla_{\theta_k}f_{2t}\nabla_{\theta_j}\nabla_{\theta_l}\xi_{t|t-1} + f_{2t}\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}\xi_{t|t-1}
\end{aligned}
$$

and

$$
\begin{aligned}
\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}B_t &= (\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}f_{1t} - \nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}f_{2t})\xi_{t|t-1} + (\nabla_{\theta_j}\nabla_{\theta_l}f_{1t} - \nabla_{\theta_j}\nabla_{\theta_l}f_{2t})\nabla_{\theta_k}\xi_{t|t-1} \\
&+ (\nabla_{\theta_k}\nabla_{\theta_l}f_{1t} - \nabla_{\theta_k}\nabla_{\theta_l}f_{2t})\nabla_{\theta_j}\xi_{t|t-1} + (\nabla_{\theta_l}f_{1t} - \nabla_{\theta_l}f_{2t})\nabla_{\theta_j}\nabla_{\theta_k}\xi_{t|t-1} \\
&+ \nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}f_{2t} + (\nabla_{\theta_j}\nabla_{\theta_k}f_{1t} - \nabla_{\theta_j}\nabla_{\theta_k}f_{2t})\nabla_{\theta_l}\xi_{t|t-1} \\
&+ (\nabla_{\theta_j}f_{1t} - \nabla_{\theta_j}f_{2t})\nabla_{\theta_k}\nabla_{\theta_l}\xi_{t|t-1} + (\nabla_{\theta_k}f_{1t} - \nabla_{\theta_k}f_{2t})\nabla_{\theta_j}\nabla_{\theta_l}\xi_{t|t-1} \\
&+ (f_{1t} - f_{2t})\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}\xi_{t|t-1}.
\end{aligned}
$$

We now evaluate the above terms at $\delta_1 = \delta_2 = \delta$ for 4 possible cases. We only report the values of $\bar{\mathcal{E}}_{jkl,t}$ but omit the derivation details.

(1) If $j \in I_0$, $k \in I_0$ and $l \in I_0$, then $\bar{\mathcal{E}}_{jkl,t} = 0$.

(2) If $j \in I_0$, $k \in I_0$ and $l \notin I_0$, then $\bar{\mathcal{E}}_{jkl,t}$ equals

$$
\begin{aligned}
&r\Bigg\{\frac{1}{\bar{f}_t}\left(\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{1t} - \nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{2t}\right) - \frac{1}{\bar{f}_t^{\,2}}\left[\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}(\nabla_{\theta_l}\bar{f}_{1t} - \nabla_{\theta_l}\bar{f}_{2t})\right] - \frac{1}{\bar{f}_t^{\,2}}\left[\nabla_{\theta_k}\bar{f}_{2t}(\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{1t} - \nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{2t})\right] \\
&-\frac{1}{\bar{f}_t^{\,2}}\left[\nabla_{\theta_j}\bar{f}_{2t}(\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{1t} - \nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{2t})\right] + 2\frac{1}{\bar{f}_t^{\,3}}\left[\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{f}_{2t}(\nabla_{\theta_l}\bar{f}_{1t} - \nabla_{\theta_l}\bar{f}_{2t})\right]\Bigg\}.
\end{aligned}
$$

(3) If $j \in I_0$, $k \notin I_0$ and $l \notin I_0$, then $\bar{\mathcal{E}}_{jkl,t}$ equals

$$
\begin{aligned}
& r\left[\frac{\nabla_{\theta_j}\nabla_{\theta_l}\nabla_{\theta_k}\bar{f}_{1t}}{\bar{f}_t}-\frac{\nabla_{\theta_j}\nabla_{\theta_l}\nabla_{\theta_k}\bar{f}_{2t}}{\bar{f}_t}\right] \\
& +\rho(1-2\xi_*)\left[\frac{\left(\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{1t}-\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{2t}\right)}{\bar{f}_t}\nabla_{\theta_k}\bar{\xi}_{t|t-1}+\frac{\left(\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{1t}-\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}\right)}{\bar{f}_t}\nabla_{\theta_l}\bar{\xi}_{t|t-1}+\frac{\left(\nabla_{\theta_l}\bar{f}_{1t}-\nabla_{\theta_l}\bar{f}_{2t}\right)}{\bar{f}_t}\nabla_{\theta_j}\nabla_{\theta_k}\bar{\xi}_{t|t-1}\right. \\
& \left.+\frac{\left(\nabla_{\theta_k}\bar{f}_{1t}-\nabla_{\theta_k}\bar{f}_{2t}\right)}{\bar{f}_t}\nabla_{\theta_j}\nabla_{\theta_l}\bar{\xi}_{t|t-1}-\frac{\nabla_{\theta_j}\bar{f}_{2t}\left(\nabla_{\theta_l}\bar{f}_{1t}-\nabla_{\theta_l}\bar{f}_{2t}\right)}{\bar{f}_t^2}\nabla_{\theta_k}\bar{\xi}_{t|t-1}-\frac{\nabla_{\theta_j}\bar{f}_{2t}\left(\nabla_{\theta_k}\bar{f}_{1t}-\nabla_{\theta_k}\bar{f}_{2t}\right)}{\bar{f}_t^2}\nabla_{\theta_l}\bar{\xi}_{t|t-1}\right] \\
& -r\left[\frac{\nabla_{\theta_j}\bar{f}_{2t}\left(\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{1t}-\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{2t}\right)}{\bar{f}_t}+\frac{\nabla_{\theta_k}\bar{f}_{2t}\left(\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{1t}-\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{2t}\right)}{\bar{f}_t}+\frac{\nabla_{\theta_l}\bar{f}_{2t}\left(\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{1t}-\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}\right)}{\bar{f}_t}\right. \\
& \left.+\frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}\left(\nabla_{\theta_l}\bar{f}_{1t}-\nabla_{\theta_l}\bar{f}_{2t}\right)}{\bar{f}_t}+\frac{\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{2t}\left(\nabla_{\theta_k}\bar{f}_{1t}-\nabla_{\theta_k}\bar{f}_{2t}\right)}{\bar{f}_t}\right] \\
& -2r\xi_*\left[\frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{1t}\left(\nabla_{\theta_l}\bar{f}_{1t}-\nabla_{\theta_l}\bar{f}_{2t}\right)}{\bar{f}_t^2}+\frac{\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{1t}\left(\nabla_{\theta_k}\bar{f}_{1t}-\nabla_{\theta_k}\bar{f}_{2t}\right)}{\bar{f}_t^2}\right. \\
& \left.-\frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}\left(\nabla_{\theta_l}\bar{f}_{1t}-\nabla_{\theta_l}\bar{f}_{2t}\right)}{\bar{f}_t^2}-\frac{\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{2t}\left(\nabla_{\theta_k}\bar{f}_{1t}-\nabla_{\theta_k}\bar{f}_{2t}\right)}{\bar{f}_t^2}\right] \\
& +2r(1-2\xi_*)\left(\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_l}\bar{f}_{1t}}{\bar{f}_t^3}+\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t^3}\right)+4r\left(\frac{\xi_*\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_l}\bar{f}_{1t}\nabla_{\theta_k}\bar{f}_{1t}}{\bar{f}_t^3}-\frac{(1-\xi_*)\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_k}\bar{f}_{2t}}{\bar{f}_t^3}\right)
\end{aligned}
$$

(4) If $j \notin I_0$, $k \notin I_0$ and $l \notin I_0$, then $\bar{\mathcal{E}}_{jkl,t}$ equals

$$
\begin{aligned}
& r\left[\frac{\nabla_{\theta_j}\nabla_{\theta_l}\nabla_{\theta_k}\bar{f}_{1t}}{\bar{f}_t} - \frac{\nabla_{\theta_j}\nabla_{\theta_l}\nabla_{\theta_k}\bar{f}_{2t}}{\bar{f}_t}\right] \\
& + \rho(1-2\xi_*)\left[\frac{\left(\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{1t} - \nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{2t}\right)}{\bar{f}_t}\nabla_{\theta_k}\bar{\xi}_{t|t-1} + \frac{\left(\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{1t} - \nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}\right)}{\bar{f}_t}\nabla_{\theta_l}\bar{\xi}_{t|t-1} + \frac{\left(\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{1t} - \nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{2t}\right)}{\bar{f}_t}\nabla_{\theta_j}\bar{\xi}_{t|t-1}\right. \\
& \left. + \frac{\left(\nabla_{\theta_l}\bar{f}_{1t} - \nabla_{\theta_l}\bar{f}_{2t}\right)}{\bar{f}_t}\nabla_{\theta_j}\nabla_{\theta_k}\bar{\xi}_{t|t-1} + \frac{\left(\nabla_{\theta_k}\bar{f}_{1t} - \nabla_{\theta_k}\bar{f}_{2t}\right)}{\bar{f}_t}\nabla_{\theta_j}\nabla_{\theta_l}\bar{\xi}_{t|t-1} + \frac{\left(\nabla_{\theta_j}\bar{f}_{1t} - \nabla_{\theta_j}\bar{f}_{2t}\right)}{\bar{f}_t}\nabla_{\theta_k}\nabla_{\theta_l}\bar{\xi}_{t|t-1}\right] \\
& - r\left[\frac{\nabla_{\theta_j}\bar{f}_{2t}\left(\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{1t} - \nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{2t}\right)}{\bar{f}_t} + \frac{\nabla_{\theta_k}\bar{f}_{2t}\left(\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{1t} - \nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{2t}\right)}{\bar{f}_t} + \frac{\nabla_{\theta_l}\bar{f}_{2t}\left(\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{1t} - \nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}\right)}{\bar{f}_t}\right. \\
& \left. + \frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}\left(\nabla_{\theta_l}\bar{f}_{1t} - \nabla_{\theta_l}\bar{f}_{2t}\right)}{\bar{f}_t} + \frac{\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{2t}\left(\nabla_{\theta_k}\bar{f}_{1t} - \nabla_{\theta_k}\bar{f}_{2t}\right)}{\bar{f}_t} + \frac{\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{2t}\left(\nabla_{\theta_j}\bar{f}_{1t} - \nabla_{\theta_j}\bar{f}_{2t}\right)}{\bar{f}_t}\right] \\
& - 2r\xi_*\left[\frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{1t}\left(\nabla_{\theta_l}\bar{f}_{1t} - \nabla_{\theta_l}\bar{f}_{2t}\right)}{\bar{f}_t^2} + \frac{\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{1t}\left(\nabla_{\theta_k}\bar{f}_{1t} - \nabla_{\theta_k}\bar{f}_{2t}\right)}{\bar{f}_t^2} + \frac{\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{1t}\left(\nabla_{\theta_j}\bar{f}_{1t} - \nabla_{\theta_j}\bar{f}_{2t}\right)}{\bar{f}_t^2}\right. \\
& \left. - \frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}\left(\nabla_{\theta_l}\bar{f}_{1t} - \nabla_{\theta_l}\bar{f}_{2t}\right)}{\bar{f}_t^2} - \frac{\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{2t}\left(\nabla_{\theta_k}\bar{f}_{1t} - \nabla_{\theta_k}\bar{f}_{2t}\right)}{\bar{f}_t^2} - \frac{\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{2t}\left(\nabla_{\theta_j}\bar{f}_{1t} - \nabla_{\theta_j}\bar{f}_{2t}\right)}{\bar{f}_t^2}\right] \\
& - 2\rho\left[\frac{\left(\nabla_{\theta_j}\bar{f}_{1t} - \nabla_{\theta_j}\bar{f}_{2t}\right)}{\bar{f}_t}\nabla_{\theta_k}\bar{\xi}_{t|t-1}\nabla_{\theta_l}\bar{\xi}_{t|t-1} + \frac{\left(\nabla_{\theta_k}\bar{f}_{1t} - \nabla_{\theta_k}\bar{f}_{2t}\right)}{\bar{f}_t}\nabla_{\theta_j}\bar{\xi}_{t|t-1}\nabla_{\theta_l}\bar{\xi}_{t|t-1} + \frac{\left(\nabla_{\theta_l}\bar{f}_{1t} - \nabla_{\theta_l}\bar{f}_{2t}\right)}{\bar{f}_t}\nabla_{\theta_j}\bar{\xi}_{t|t-1}\nabla_{\theta_k}\bar{\xi}_{t|t-1}\right] \\
& + \rho(6\xi_*^2 - 4\xi_*)\left[\frac{\left(\nabla_{\theta_j}\bar{f}_{1t}\nabla_{\theta_k}\bar{f}_{1t}\right)}{\bar{f}_t^2}\nabla_{\theta_l}\bar{\xi}_{t|t-1} + \frac{\left(\nabla_{\theta_j}\bar{f}_{1t}\nabla_{\theta_l}\bar{f}_{1t}\right)}{\bar{f}_t^2}\nabla_{\theta_k}\bar{\xi}_{t|t-1} + \frac{\left(\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\bar{f}_{1t}\right)}{\bar{f}_t^2}\nabla_{\theta_j}\bar{\xi}_{t|t-1}\right] \\
& - \rho(6\xi_*^2 - 6\xi_* + 1)\left[\frac{\left(\nabla_{\theta_j}\bar{f}_{1t}\nabla_{\theta_k}\bar{f}_{2t}\right)}{\bar{f}_t^2}\nabla_{\theta_l}\bar{\xi}_{t|t-1} + \frac{\left(\nabla_{\theta_j}\bar{f}_{1t}\nabla_{\theta_l}\bar{f}_{2t}\right)}{\bar{f}_t^2}\nabla_{\theta_k}\bar{\xi}_{t|t-1} + \frac{\left(\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_j}\bar{f}_{2t}\right)}{\bar{f}_t^2}\nabla_{\theta_l}\bar{\xi}_{t|t-1}\right. \\
& \left. + \frac{\left(\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\bar{f}_{2t}\right)}{\bar{f}_t^2}\nabla_{\theta_j}\bar{\xi}_{t|t-1} + \frac{\left(\nabla_{\theta_l}\bar{f}_{1t}\nabla_{\theta_j}\bar{f}_{2t}\right)}{\bar{f}_t^2}\nabla_{\theta_k}\bar{\xi}_{t|t-1} + \frac{\left(\nabla_{\theta_l}\bar{f}_{1t}\nabla_{\theta_k}\bar{f}_{2t}\right)}{\bar{f}_t^2}\nabla_{\theta_j}\bar{\xi}_{t|t-1}\right] \\
& + \rho[6(1-\xi_*)^2 - 4(1-\xi_*)]\left[\frac{\left(\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{f}_{2t}\right)}{\bar{f}_t^2}\nabla_{\theta_l}\bar{\xi}_{t|t-1} + \frac{\left(\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_l}\bar{f}_{2t}\right)}{\bar{f}_t^2}\nabla_{\theta_k}\bar{\xi}_{t|t-1} + \frac{\left(\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_l}\bar{f}_{2t}\right)}{\bar{f}_t^2}\nabla_{\theta_j}\bar{\xi}_{t|t-1}\right] \\
& + 6r\xi_*^2\frac{\nabla_{\theta_j}\bar{f}_{1t}\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\bar{f}_{1t}}{\bar{f}_t^3} - r(6\xi_*^2 - 4\xi_*)\left(\frac{\nabla_{\theta_j}\bar{f}_{1t}\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t^3} + \frac{\nabla_{\theta_j}\bar{f}_{1t}\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_l}\bar{f}_{1t}}{\bar{f}_t^3} + \frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\bar{f}_{1t}}{\bar{f}_t^3}\right) \\
& + r[6(1-\xi_*)^2 - 4(1-\xi_*)]\left(\frac{\nabla_{\theta_j}\bar{f}_{1t}\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t^3} + \frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_l}\bar{f}_{1t}}{\bar{f}_t^3} + \frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t^3}\right) \\
& - 6r(1-\xi_*)^2\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t^3}.
\end{aligned}
$$

□

Proof of Lemma 2. When $\delta_1 = \delta_2$, we can directly apply the proposition 13.1 in Hamilton (1994) to show that the sequence of predicted mean squared error $\bar{P}_{t|t-1}$ convergences. Then by (2.10) and (2.12), $\bar{P}_{t|t}$ also convergences to a positive semidefinite matrix $\bar{P}_*$. Let $P_{0|0} = \bar{P}_*$, then the results in this lemma hold and (4.4) can be achieved by combining (2.6), (2.10) with (2.12) when $\delta_1 = \delta_2$. In addition, by (2.12) and (2.6), we also have both $\bar{P}^{(i)}_{t|t}$ and $\bar{P}_{t+1|t}$ equal to $\bar{P}_*$ and $\bar{F}\bar{P}_*\bar{F}' + \bar{Q}$ for

$t = 1, ..., T$. □

Proof of Lemma 3. To show Lemma 3.1, combining (2.8) and (2.10), then

$$P^{(i)}_{t|t} = (I - P^{(i)}_{t|t-1} H_i [H'_i P^{(i)}_{t|t-1} H_i]^{-1} H'_i) P^{(i)}_{t|t-1}.$$

The first order derivative of $P^{(i)}_{t|t}$ with respect to θ_j gives

$$\begin{aligned}
\nabla_{\theta_j} P^{(i)}_{t|t} &= \left(I - \frac{P^{(i)}_{t|t-1} H_i H'_i}{H'_i P^{(i)}_{t|t-1} H_i} \right) \nabla_{\theta_j} P^{(i)}_{t|t-1} \left(I - \frac{H_i H'_i P^{(i)}_{t|t-1}}{H'_i P^{(i)}_{t|t-1} H_i} \right) \\
&- \left(\frac{P^{(i)}_{t|t-1} \left(\nabla_{\theta_j} H_i H'_i + H_i \nabla_{\theta_j} H'_i \right)}{H'_i P^{(i)}_{t|t-1} H_i} - \frac{\left(P^{(i)}_{t|t-1} H_i H'_i \right) \left(\nabla_{\theta_j} H'_i P^{(i)}_{t|t-1} H_i + H'_i P^{(i)}_{t|t-1} \nabla_{\theta_j} H_i \right)}{\left(H'_i P^{(i)}_{t|t-1} H_i \right)^2} \right) P^{(i)}_{t|t-1}.
\end{aligned} \tag{B.10}$$

The first order derivative of $P^{(i)}_{t|t-1}$ in (2.6) with respect to θ_j gives

$$\nabla_{\theta_j} P^{(i)}_{t|t-1} = F_i \nabla_{\theta_j} P_{t-1|t-1} F'_i + \nabla_{\theta_j} F_i P_{t-1|t-1} F'_i + F_i P_{t-1|t-1} \nabla_{\theta_j} F'_i + \nabla_{\theta_j} Q_i.$$

By the last two equations, we have

$$\begin{aligned}
\nabla_{\theta_j} P^{(i)}_{t|t} &= \left[\left(I - \frac{P^{(i)}_{t|t-1} H_i H'_i}{H'_i P^{(i)}_{t|t-1} H_i} \right) F_i \right] \nabla_{\theta_j} P_{t-1|t-1} \left[F'_i \left(I - \frac{H_i H'_i P^{(i)}_{t|t-1}}{H'_i P^{(i)}_{t|t-1} H_i} \right) \right] \\
&+ \left(I - \frac{P^{(i)}_{t|t-1} H_i H'_i}{H'_i P^{(i)}_{t|t-1} H_i} \right) \left(\nabla_{\theta_j} F_i P_{t-1|t-1} F'_i + F_i P_{t-1|t-1} \nabla_{\theta_j} F'_i + \nabla_{\theta_j} Q_i \right) \left(I - \frac{H_i H'_i P^{(i)}_{t|t-1}}{H'_i P^{(i)}_{t|t-1} H_i} \right) \\
&- \left(\frac{P^{(i)}_{t|t-1} \left(\nabla_{\theta_j} H_i H'_i + H_i \nabla_{\theta_j} H'_i \right)}{H'_i P^{(i)}_{t|t-1} H_i} - \frac{\left(P^{(i)}_{t|t-1} H_i H'_i \right) \left(\nabla_{\theta_j} H'_i P^{(i)}_{t|t-1} H_i + H'_i P^{(i)}_{t|t-1} \nabla_{\theta_j} H_i \right)}{\left(H'_i P^{(i)}_{t|t-1} H_i \right)^2} \right) P^{(i)}_{t|t-1}.
\end{aligned} \tag{B.11}$$

Let $\delta_1 = \delta_2$, then

$$\begin{aligned}
\nabla_{\theta_j} \bar{P}^{(i)}_{t|t} &= \left[\left(I - \frac{\bar{P} \bar{H} \bar{H}'}{\bar{H}' \bar{P} \bar{H}} \right) \bar{F} \right] \nabla_{\theta_j} \bar{P}_{t-1|t-1} \left[\bar{F}' \left(I - \frac{\bar{H} \bar{H}' \bar{P}}{\bar{H}' \bar{P} \bar{H}} \right) \right] \\
&+ \left(I - \frac{\bar{P} \bar{H} \bar{H}'}{\bar{H}' \bar{P} \bar{H}} \right) \left[\nabla_{\theta_j} \bar{F}_i \bar{P}_* \bar{F}' + \bar{F} \bar{P}_* \nabla_{\theta_j} \bar{F}'_i + \nabla_{\theta_j} Q_i \right] \left(I - \frac{\bar{H} \bar{H}' \bar{P}}{\bar{H}' \bar{P} \bar{H}} \right) \\
&- \left(\frac{\bar{P} \left(\nabla_{\theta_j} \bar{H}_i \bar{H}' + \bar{H} \nabla_{\theta_j} \bar{H}'_i \right)}{\bar{H}' \bar{P} \bar{H}} - \frac{\left(\bar{P} \bar{H} \bar{H}' \right) \left(\nabla_{\theta_j} \bar{H}'_i \bar{P} \bar{H} + \bar{H}' \bar{P} \nabla_{\theta_j} \bar{H}_i \right)}{\left(\bar{H}' \bar{P} \bar{H} \right)^2} \right) \left[\bar{F} \bar{P}_* \bar{F}' + \bar{Q} \right],
\end{aligned} \tag{B.12}$$

where

$$\left(I - \frac{\bar{P}\bar{H}\bar{H}'}{\bar{H}'\bar{P}\bar{H}}\right)\bar{F} \text{ and } \bar{F}'\left(I - \frac{\bar{H}\bar{H}'\bar{P}}{\bar{H}'\bar{P}\bar{H}}\right)$$

have all their eigenvalues inside the unit circle by Assumption 4.

Taking first order derivative with respect to θ_j in (2.12) and let $\delta_1 = \delta_2$, we have

$$\nabla_{\theta_j}\bar{\hat{P}}_{t|t} = \xi_* \nabla_{\theta_j}\bar{P}^{(1)}_{t|t} + (1-\xi_*)\nabla_{\theta_j}\bar{P}^{(2)}_{t|t}.$$

Together with the result in (B.12),

$$\begin{aligned}
\nabla_{\theta_j}\bar{P}_{t|t} =& \left[\left(I - \frac{\bar{P}\bar{H}\bar{H}'}{\bar{H}'\bar{P}\bar{H}}\right)\bar{F}\right]\nabla_{\theta_j}\bar{P}_{t-1|t-1}\left[\bar{F}'\left(I - \frac{\bar{H}\bar{H}'\bar{P}}{\bar{H}'\bar{P}\bar{H}}\right)\right] \qquad \text{(B.13)}\\
&+ \xi_*\left(I - \frac{\bar{P}\bar{H}\bar{H}'}{\bar{H}'\bar{P}\bar{H}}\right)\left[\nabla_{\theta_j}\bar{F}_1\bar{P}_*\bar{F}' + \bar{F}\bar{P}_*\nabla_{\theta_j}\bar{F}_1' + \nabla_{\theta_j}Q_1\right]\left(I - \frac{\bar{H}\bar{H}'\bar{P}}{\bar{H}'\bar{P}\bar{H}}\right)\\
&+ (1-\xi_*)\left(I - \frac{\bar{P}\bar{H}\bar{H}'}{\bar{H}'\bar{P}\bar{H}}\right)\left[\nabla_{\theta_j}\bar{F}_2\bar{P}_*\bar{F}' + \bar{F}\bar{P}_*\nabla_{\theta_j}\bar{F}_2' + \nabla_{\theta_j}Q_2\right]\left(I - \frac{\bar{H}\bar{H}'\bar{P}}{\bar{H}'\bar{P}\bar{H}}\right)\\
&- \xi_*\left(\frac{\bar{P}\left(\nabla_{\theta_j}\bar{H}_1\bar{H}' + \bar{H}\nabla_{\theta_j}\bar{H}_1'\right)}{\bar{H}'\bar{P}\bar{H}} - \frac{\left(\bar{P}\bar{H}\bar{H}'\right)\left(\nabla_{\theta_j}\bar{H}_1'\bar{P}\bar{H} + \bar{H}'\bar{P}\nabla_{\theta_j}\bar{H}_1\right)}{\left(\bar{H}'\bar{P}\bar{H}\right)^2}\right)\left[\bar{F}\bar{P}_*\bar{F}' + \bar{Q}\right]\\
&- (1-\xi_*)\left(\frac{\bar{P}\left(\nabla_{\theta_j}\bar{H}_2\bar{H}' + \bar{H}\nabla_{\theta_j}\bar{H}_2'\right)}{\bar{H}'\bar{P}\bar{H}} - \frac{\left(\bar{P}\bar{H}\bar{H}'\right)\left(\nabla_{\theta_j}\bar{H}_2'\bar{P}\bar{H} + \bar{H}'\bar{P}\nabla_{\theta_j}\bar{H}_2\right)}{\left(\bar{H}'\bar{P}\bar{H}\right)^2}\right)\left[\bar{F}\bar{P}_*\bar{F}' + \bar{Q}\right],
\end{aligned}$$

which gives the result in the lemma.

To show Lemma 3.2, by (2.5), (2.7), (2.8) and (2.9), we have

$$x^{(i)}_{t|t} = \left(I - \frac{P^{(i)}_{t|t-1}H_iH_i'}{H_i'P^{(i)}_{t|t-1}H_i}\right)\left(F_ix_{t-1|t-1} + G_i\right) + \left(\frac{P^{(i)}_{t|t-1}H_i}{H_i'P^{(i)}_{t|t-1}H_i}\right)\left(y_t - A_i'z_t\right).$$

Then the first order derivative of $x^{(i)}_{t|t}$ with respect to θ_j gives

$$\begin{aligned}
\nabla_{\theta_j} x_{t|t}^{(i)} = & \left(I - \frac{P_{t|t-1}^{(i)} H_i H_i'}{H_i' P_{t|t-1}^{(i)} H_i} \right) F_i \nabla_{\theta_j} x_{t-1|t-1} \\
& + \left(I - \frac{P_{t|t-1}^{(i)} H_i H_i'}{H_i' P_{t|t-1}^{(i)} H_i} \right) \left(\nabla_{\theta_j} F_i x_{t-1|t-1} + \nabla_{\theta_j} G_i \right) \\
& - \left(\frac{\nabla_{\theta_j} P_{t|t-1}^{(i)} H_i H_i' + P_{t|t-1}^{(i)} \nabla_{\theta_j} H_i H_i' + P_{t|t-1}^{(i)} H_i \nabla_{\theta_j} H_i'}{H_i' P_{t|t-1}^{(i)} H_i} \right) \left(F_i x_{t-1|t-1} + G_i \right) \\
& + \left(\frac{\left(P_{t|t-1}^{(i)} H_i H_i' \right) \left(\nabla_{\theta_j} H_i' P_{t|t-1}^{(i)} H_i + H_i' \nabla_{\theta_j} P_{t|t-1}^{(i)} H_i + H_i' P_{t|t-1}^{(i)} \nabla_{\theta_j} H_i \right)}{\left(H_i' P_{t|t-1}^{(i)} H_i \right)^2} \right) \left(F_i x_{t-1|t-1} + G_i \right) \\
& + \left(\frac{\nabla_{\theta_j} P_{t|t-1}^{(i)} H_i + P_{t|t-1}^{(i)} \nabla_{\theta_j} H_i}{H_i' P_{t|t-1}^{(i)} H_i} \right) \left(y_t - A_i' z_t \right) \\
& - \left(\frac{\left(P_{t|t-1}^{(i)} H_i \right) \left(\nabla_{\theta_j} H_i' P_{t|t-1}^{(i)} H_i + H_i' \nabla_{\theta_j} P_{t|t-1}^{(i)} H_i + H_i' P_{t|t-1}^{(i)} \nabla_{\theta_j} H_i \right)}{\left(H_i' P_{t|t-1}^{(i)} H_i \right)^2} \right) \left(y_t - A_i' z_t \right) \\
& - \left(\frac{P_{t|t-1}^{(i)} H_i}{H_i' P_{t|t-1}^{(i)} H_i} \right) \nabla_{\theta_j} A_i' z_t.
\end{aligned} \tag{B.14}$$

When $\delta_1 = \delta_2$, we have

$$\begin{aligned}
\nabla_{\theta_j} \bar{x}_{t|t}^{(i)} = & \left(I - \frac{\bar{P} \bar{H} \bar{H}'}{\bar{H}' \bar{P} \bar{H}} \right) \bar{F} \nabla_{\theta_j} \bar{x}_{t-1|t-1} \\
& + \left(I - \frac{\bar{P} \bar{H} \bar{H}'}{\bar{H}' \bar{P} \bar{H}} \right) \left(\nabla_{\theta_j} \bar{F}_i \bar{x}_{t-1|t-1} + \nabla_{\theta_j} \bar{G}_i \right) \\
& - \left(\frac{\nabla_{\theta_j} \bar{P}_{t|t-1}^{(i)} \bar{H} \bar{H}' + \bar{P} \nabla_{\theta_j} \bar{H}_i \bar{H}' + \bar{P} \bar{H} \nabla_{\theta_j} \bar{H}_i'}{\bar{H}' \bar{P} \bar{H}} \right) \left(\bar{F} \bar{x}_{t-1|t-1} + \bar{G} \right) \\
& + \left(\frac{\left(\bar{P} \bar{H} \bar{H}' \right) \left(\nabla_{\theta_j} \bar{H}_i' \bar{P} \bar{H} + \bar{H}' \nabla_{\theta_j} \bar{P}_{t|t-1}^{(i)} \bar{H} + \bar{H}' \bar{P} \nabla_{\theta_j} \bar{H}_i \right)}{\left(\bar{H}' \bar{P} \bar{H} \right)^2} \right) \left(\bar{F} \bar{x}_{t-1|t-1} + \bar{G} \right) \\
& + \left(\frac{\nabla_{\theta_j} \bar{P}_{t|t-1}^{(i)} \bar{H} + \bar{P} \nabla_{\theta_j} \bar{H}_i}{\bar{H}' \bar{P} \bar{H}} - \frac{\left(\bar{P} \bar{H} \right) \left(\nabla_{\theta_j} \bar{H}_i' \bar{P} \bar{H} + \bar{H}' \nabla_{\theta_j} \bar{P}_{t|t-1}^{(i)} \bar{H} + \bar{H}' \bar{P} \nabla_{\theta_j} \bar{H}_i \right)}{\left(\bar{H}' \bar{P} \bar{H} \right)^2} \right) \left(y_t - \bar{A}' z_t \right) \\
& - \left(\frac{\bar{P} \bar{H}}{\bar{H}' \bar{P} \bar{H}} \right) \nabla_{\theta_j} \bar{A}_i' z_t,
\end{aligned} \tag{B.15}$$

where the expressions for $\nabla_{\theta_j} \bar{P}_{t|t-1}^{(1)}$ and $\nabla_{\theta_j} \bar{P}_{t|t-1}^{(1)}$, based on (2.6), are given by:

$$\nabla_{\theta_j} \bar{P}_{t|t-1}^{(1)} = \nabla_{\theta_j} \bar{F}_1 \bar{P} \bar{F}' + \bar{F} \nabla_{\theta_j} \bar{P}_{t-1|t-1} \bar{F}' + \bar{F} \bar{P} \nabla_{\theta_j} \bar{F}_1' + \nabla_{\theta_j} \bar{Q}_1, \tag{B.16}$$

and

$$\nabla_{\theta_j}\bar{P}^{(2)}_{t|t-1} = \nabla_{\theta_j}\bar{F}_2\bar{P}\bar{F}' + \bar{F}\nabla_{\theta_j}\bar{P}_{t-1|t-1}\bar{F}' + \bar{F}\bar{P}\nabla_{\theta_j}\bar{F}_2' + \nabla_{\theta_j}\bar{Q}_2. \tag{B.17}$$

By (2.11), when $\delta_1 = \delta_2$,

$$\nabla_{\theta_j}\bar{x}_{t|t} = \xi_*\nabla_{\theta_j}\bar{x}^{(1)}_{t|t} + (1-\xi_*)\nabla_{\theta_j}\bar{x}^{(2)}_{t|t}.$$

Therefore,

$$\nabla_{\theta_j}\bar{x}_{t|t} = \left(I - \frac{\bar{P}\bar{H}\bar{H}'}{\bar{H}'\bar{P}\bar{H}}\right)\bar{F}\nabla_{\theta_j}\bar{x}_{t-1|t-1} + \bar{\mathcal{X}}_{j,t}, \tag{B.18}$$

where

$$\begin{aligned}
\bar{\mathcal{X}}_{j,t} = & \left(I - \frac{\bar{P}\bar{H}\bar{H}'}{\bar{H}'\bar{P}\bar{H}}\right)\left[\xi_*\left(\nabla_{\theta_j}\bar{F}_1\bar{x}_{t-1|t-1} + \nabla_{\theta_j}\bar{G}_1\right) + (1-\xi_*)\left(\nabla_{\theta_j}\bar{F}_2\bar{x}_{t-1|t-1} + \nabla_{\theta_j}\bar{G}_2\right)\right] \\
& - \xi_*\left(\frac{\nabla_{\theta_j}\bar{P}^{(1)}_{t|t-1}\bar{H}\bar{H}' + \bar{P}\nabla_{\theta_j}\bar{H}_1\bar{H}' + \bar{P}\bar{H}\nabla_{\theta_j}\bar{H}_1'}{\bar{H}'\bar{P}\bar{H}}\right)\left(\bar{F}\bar{x}_{t-1|t-1} + \bar{G}\right) \\
& + \xi_*\left(\frac{\left(\bar{P}\bar{H}\bar{H}'\right)\left(\nabla_{\theta_j}\bar{H}_1'\bar{P}\bar{H} + \bar{H}'\nabla_{\theta_j}\bar{P}^{(1)}_{t|t-1}\bar{H} + \bar{H}'\bar{P}\nabla_{\theta_j}\bar{H}_1\right)}{\left(\bar{H}'\bar{P}\bar{H}\right)^2}\right)\left(\bar{F}\bar{x}_{t-1|t-1} + \bar{G}\right) \\
& - (1-\xi_*)\left(\frac{\nabla_{\theta_j}\bar{P}^{(2)}_{t|t-1}\bar{H}\bar{H}' + \bar{P}\nabla_{\theta_j}\bar{H}_2\bar{H}' + \bar{P}\bar{H}\nabla_{\theta_j}\bar{H}_2'}{\bar{H}'\bar{P}\bar{H}}\right)\left(\bar{F}\bar{x}_{t-1|t-1} + \bar{G}\right) \\
& + (1-\xi_*)\left(\frac{\left(\bar{P}\bar{H}\bar{H}'\right)\left(\nabla_{\theta_j}\bar{H}_2'\bar{P}\bar{H} + \bar{H}'\nabla_{\theta_j}\bar{P}^{(2)}_{t|t-1}\bar{H} + \bar{H}'\bar{P}\nabla_{\theta_j}\bar{H}_2\right)}{\left(\bar{H}'\bar{P}\bar{H}\right)^2}\right)\left(\bar{F}\bar{x}_{t-1|t-1} + \bar{G}\right) \\
& + \xi_*\left(\frac{\nabla_{\theta_j}\bar{P}^{(1)}_{t|t-1}\bar{H} + \bar{P}\nabla_{\theta_j}\bar{H}_1}{\bar{H}'\bar{P}\bar{H}} - \frac{\left(\bar{P}\bar{H}\right)\left(\nabla_{\theta_j}\bar{H}_1'\bar{P}\bar{H} + \bar{H}'\nabla_{\theta_j}\bar{P}^{(1)}_{t|t-1}\bar{H} + \bar{H}'\bar{P}\nabla_{\theta_j}\bar{H}_1\right)}{\left(\bar{H}'\bar{P}\bar{H}\right)^2}\right)\left(y_t - \bar{A}_1'z_t\right) \\
& + (1-\xi_*)\left(\frac{\nabla_{\theta_j}\bar{P}^{(2)}_{t|t-1}\bar{H} + \bar{P}\nabla_{\theta_j}\bar{H}_2}{\bar{H}'\bar{P}\bar{H}} - \frac{\left(\bar{P}\bar{H}\right)\left(\nabla_{\theta_j}\bar{H}_2'\bar{P}\bar{H} + \bar{H}'\nabla_{\theta_j}\bar{P}^{(2)}_{t|t-1}\bar{H} + \bar{H}'\bar{P}\nabla_{\theta_j}\bar{H}_2\right)}{\left(\bar{H}'\bar{P}\bar{H}\right)^2}\right)\left(y_t - \bar{A}_2'z_t\right) \\
& - \left(\frac{\bar{P}\bar{H}}{\bar{H}'\bar{P}\bar{H}}\right)\left[\xi_*\nabla_{\theta_j}\bar{A}_1' + (1-\xi_*)\nabla_{\theta_j}\bar{A}_2'\right]z_t.
\end{aligned}$$

To show Lemma 3.3. By (2.12), we have the second order derivatives of $P_{t|t}$ with respect to θ_j and

θ_k, when evaluated at some $\delta_1 = \delta_2$, is given by

$$\begin{aligned}
\nabla_{\theta_j}\nabla_{\theta_k}\bar{P}_{t|t} &= \xi_*(1-\xi_*)\left(\nabla_{\theta_j}\bar{x}^{(1)}_{t|t} - \nabla_{\theta_j}\bar{x}^{(2)}_{t|t}\right)\left(\nabla_{\theta_k}\bar{x}^{(1)}_{t|t} - \nabla_{\theta_k}\bar{x}^{(2)}_{t|t}\right)' \\
&\quad + \xi_*(1-\xi_*)\left(\nabla_{\theta_k}\bar{x}^{(1)}_{t|t} - \nabla_{\theta_k}\bar{x}^{(2)}_{t|t}\right)\left(\nabla_{\theta_j}\bar{x}^{(1)}_{t|t} - \nabla_{\theta_j}\bar{x}^{(2)}_{t|t}\right)' \\
&\quad + \nabla_{\theta_j}\bar{\xi}_{t|t}\left(\nabla_{\theta_k}\bar{P}^{(1)}_{t|t} - \nabla_{\theta_k}\bar{P}^{(2)}_{t|t}\right) + \nabla_{\theta_k}\bar{\xi}_{t|t}\left(\nabla_{\theta_j}\bar{P}^{(1)}_{t|t} - \nabla_{\theta_j}\bar{P}^{(2)}_{t|t}\right) \\
&\quad + \xi_*\nabla_{\theta_j}\nabla_{\theta_k}\bar{P}^{(1)}_{t|t} + (1-\xi_*)\nabla_{\theta_j}\nabla_{\theta_k}\bar{P}^{(2)}_{t|t}.
\end{aligned}$$

There are three main components in this expression, $\nabla_{\theta_j}\bar{x}^{(1)}_{t|t} - \nabla_{\theta_j}\bar{x}^{(2)}_{t|t}$, $\nabla_{\theta_j}\bar{P}^{(1)}_{t|t} - \nabla_{\theta_j}\bar{P}^{(2)}_{t|t}$ and $\xi_*\nabla_{\theta_j}\nabla_{\theta_k}\bar{P}^{(1)}_{t|t} + (1-\xi_*)\nabla_{\theta_j}\nabla_{\theta_k}\bar{P}^{(2)}_{t|t}$. $\nabla_{\theta_j}\bar{x}^{(i)}_{t|t}$ and $\nabla_{\theta_j}\bar{P}^{(1)}_{t|t}$ are given in (B.15) and (B.12). To study $\nabla_{\theta_j}\nabla_{\theta_k}\bar{P}^{(i)}_{t|t}$, by (B.11), we have

$$\nabla_{\theta_j}\nabla_{\theta_k}P^{(i)}_{t|t} = \left[\left(I - \frac{P^{(i)}_{t|t-1}H_iH_i'}{H_i'P^{(i)}_{t|t-1}H_i}\right)F_i\right]\nabla_{\theta_j}\nabla_{\theta_k}P_{t-1|t-1}\left[F_i'\left(I - \frac{H_iH_i'P^{(i)}_{t|t-1}}{H_i'P^{(i)}_{t|t-1}H_i}\right)\right] + \mathcal{P}^{(i)}_{jk,t},$$

where

$$\begin{aligned}
\mathcal{P}^{(i)}_{jk,t} &= \nabla_{\theta_k}\left[\left(I - \frac{P^{(i)}_{t|t-1}H_iH_i'}{H_i'P^{(i)}_{t|t-1}H_i}\right)F_i\right]\nabla_{\theta_j}P_{t-1|t-1}\left[F_i'\left(I - \frac{H_iH_i'P^{(i)}_{t|t-1}}{H_i'P^{(i)}_{t|t-1}H_i}\right)\right] \\
&\quad + \left[\left(I - \frac{P^{(i)}_{t|t-1}H_iH_i'}{H_i'P^{(i)}_{t|t-1}H_i}\right)F_i\right]\nabla_{\theta_j}P_{t-1|t-1}\nabla_{\theta_k}\left[F_i'\left(I - \frac{H_iH_i'P^{(i)}_{t|t-1}}{H_i'P^{(i)}_{t|t-1}H_i}\right)\right] \\
&\quad + \nabla_{\theta_k}\left(I - \frac{P^{(i)}_{t|t-1}H_iH_i'}{H_i'P^{(i)}_{t|t-1}H_i}\right)\left(\nabla_{\theta_j}F_iP_{t-1|t-1}F_i' + F_iP_{t-1|t-1}\nabla_{\theta_j}F_i' + \nabla_{\theta_j}Q_i\right)\left(I - \frac{H_iH_i'P^{(i)}_{t|t-1}}{H_i'P^{(i)}_{t|t-1}H_i}\right) \\
&\quad + \left(I - \frac{P^{(i)}_{t|t-1}H_iH_i'}{H_i'P^{(i)}_{t|t-1}H_i}\right)\nabla_{\theta_k}\left(\nabla_{\theta_j}F_iP_{t-1|t-1}F_i' + F_iP_{t-1|t-1}\nabla_{\theta_j}F_i' + \nabla_{\theta_j}Q_i\right)\left(I - \frac{H_iH_i'P^{(i)}_{t|t-1}}{H_i'P^{(i)}_{t|t-1}H_i}\right) \\
&\quad + \left(I - \frac{P^{(i)}_{t|t-1}H_iH_i'}{H_i'P^{(i)}_{t|t-1}H_i}\right)\left(\nabla_{\theta_j}F_iP_{t-1|t-1}F_i' + F_iP_{t-1|t-1}\nabla_{\theta_j}F_i' + \nabla_{\theta_j}Q_i\right)\nabla_{\theta_k}\left(I - \frac{H_iH_i'P^{(i)}_{t|t-1}}{H_i'P^{(i)}_{t|t-1}H_i}\right) \\
&\quad - \nabla_{\theta_k}\left(\frac{P^{(i)}_{t|t-1}\left(\nabla_{\theta_j}H_iH_i' + H_i\nabla_{\theta_j}H_i'\right)}{H_i'P^{(i)}_{t|t-1}H_i} - \frac{\left(P^{(i)}_{t|t-1}H_iH_i'\right)\left(\nabla_{\theta_j}H_i'P^{(i)}_{t|t-1}H_i + H_i'P^{(i)}_{t|t-1}\nabla_{\theta_j}H_i\right)}{\left(H_i'P^{(i)}_{t|t-1}H_i\right)^2}\right)P^{(i)}_{t|t-1} \\
&\quad - \left(\frac{P^{(i)}_{t|t-1}\left(\nabla_{\theta_j}H_iH_i' + H_i\nabla_{\theta_j}H_i'\right)}{H_i'P^{(i)}_{t|t-1}H_i} - \frac{\left(P^{(i)}_{t|t-1}H_iH_i'\right)\left(\nabla_{\theta_j}H_i'P^{(i)}_{t|t-1}H_i + H_i'P^{(i)}_{t|t-1}\nabla_{\theta_j}H_i\right)}{\left(H_i'P^{(i)}_{t|t-1}H_i\right)^2}\right)\nabla_{\theta_k}P^{(i)}_{t|t-1}.
\end{aligned}$$

Therefore, we have

$$
\begin{aligned}
&\xi_* \nabla_{\theta_j} \nabla_{\theta_k} \bar{P}_{t|t}^{(1)} + (1-\xi_*) \nabla_{\theta_j} \nabla_{\theta_k} \bar{P}_{t|t}^{(2)} \\
&= \left[\left(I - \frac{\bar{P}\bar{H}\bar{H}'}{\bar{H}'\bar{P}\bar{H}} \right) \bar{F} \right] \nabla_{\theta_j} \nabla_{\theta_k} \bar{P}_{t-1|t-1} \left[\bar{F}' \left(I - \frac{\bar{H}\bar{H}'\bar{P}}{\bar{H}'\bar{P}\bar{H}} \right) \right] + \xi_* \bar{\mathcal{P}}_{jk,t}^{(1)} + (1-\xi_*) \bar{\mathcal{P}}_{jk,t}^{(2)}
\end{aligned}
$$

and

$$
\nabla_{\theta_j} \nabla_{\theta_k} \bar{P}_{t|t} = \left[\left(I - \frac{\bar{P}\bar{H}\bar{H}'}{\bar{H}'\bar{P}\bar{H}} \right) \bar{F} \right] \nabla_{\theta_j} \nabla_{\theta_k} \bar{P}_{t-1|t-1} \left[\bar{F}' \left(I - \frac{\bar{H}\bar{H}'\bar{P}}{\bar{H}'\bar{P}\bar{H}} \right) \right] + \bar{\mathcal{P}}_{jk,t},
$$

where

$$
\begin{aligned}
\bar{\mathcal{P}}_{jk,t} &= \xi_* \bar{\mathcal{P}}_{jk,t}^{(1)} + (1-\xi_*) \bar{\mathcal{P}}_{jk,t}^{(2)} \\
&\quad + \xi_*(1-\xi_*) \left(\nabla_{\theta_j} \bar{x}_{t|t}^{(1)} - \nabla_{\theta_j} \bar{x}_{t|t}^{(2)} \right) \left(\nabla_{\theta_k} \bar{x}_{t|t}^{(1)} - \nabla_{\theta_k} \bar{x}_{t|t}^{(2)} \right)' \\
&\quad + \xi_*(1-\xi_*) \left(\nabla_{\theta_k} \bar{x}_{t|t}^{(1)} - \nabla_{\theta_k} \bar{x}_{t|t}^{(2)} \right) \left(\nabla_{\theta_j} \bar{x}_{t|t}^{(1)} - \nabla_{\theta_j} \bar{x}_{t|t}^{(2)} \right)' \\
&\quad + \nabla_{\theta_j} \bar{\xi}_{t|t} \left(\nabla_{\theta_k} \bar{P}_{t|t}^{(1)} - \nabla_{\theta_k} \bar{P}_{t|t}^{(2)} \right) + \nabla_{\theta_k} \bar{\xi}_{t|t} \left(\nabla_{\theta_j} \bar{P}_{t|t}^{(1)} - \nabla_{\theta_j} \bar{P}_{t|t}^{(2)} \right).
\end{aligned}
$$

This gives the result in Lemma 3.3.

To show Lemma 3.4. By (2.11), we have

$$
\begin{aligned}
\nabla_{\theta_j} \nabla_{\theta_k} x_{t|t} &= \nabla_{\theta_j} \nabla_{\theta_k} \xi_{t|t} \left(x_{t|t}^{(1)} - x_{t|t}^{(2)} \right) + \nabla_{\theta_j} \xi_{t|t} \left(\nabla_{\theta_k} x_{t|t}^{(1)} - \nabla_{\theta_k} x_{t|t}^{(2)} \right) \\
&\quad + \nabla_{\theta_k} \xi_{t|t} \left(\nabla_{\theta_j} x_{t|t}^{(1)} - \nabla_{\theta_j} x_{t|t}^{(2)} \right) + \xi_{t|t} \nabla_{\theta_j} \nabla_{\theta_k} x_{t|t}^{(1)} + (1-\xi_{t|t}) \nabla_{\theta_j} \nabla_{\theta_k} x_{t|t}^{(2)}
\end{aligned}
$$

and

$$
\begin{aligned}
\nabla_{\theta_j} \nabla_{\theta_k} \bar{x}_{t|t} &= \nabla_{\theta_j} \bar{\xi}_{t|t} \left(\nabla_{\theta_k} \bar{x}_{t|t}^{(1)} - \nabla_{\theta_k} \bar{x}_{t|t}^{(2)} \right) + \nabla_{\theta_k} \bar{\xi}_{t|t} \left(\nabla_{\theta_j} \bar{x}_{t|t}^{(1)} - \nabla_{\theta_j} \bar{x}_{t|t}^{(2)} \right) \\
&\quad + \xi_* \nabla_{\theta_j} \nabla_{\theta_k} \bar{x}_{t|t}^{(1)} + (1-\xi_*) \nabla_{\theta_j} \nabla_{\theta_k} \bar{x}_{t|t}^{(2)},
\end{aligned}
$$

when $\delta_1 = \delta_2$. By the previous results, we just need to derive $\nabla_{\theta_j} \nabla_{\theta_k} \bar{x}_{t|t}^{(i)}$. Differentiating $\nabla_{\theta_j} x_{t|t}^{(i)}$ in (B.14) with respect to θ_k, we have

$$
\nabla_{\theta_j} \nabla_{\theta_k} x_{t|t}^{(i)} = \left[\left(I - \frac{P_{t|t-1}^{(i)} H_i H_i'}{H_i' P_{t|t-1}^{(i)} H_i} \right) F_i \right] \nabla_{\theta_j} \nabla_{\theta_k} x_{t-1|t-1} + \mathcal{X}_{jk,t}^{(i)},
$$

where

$$
\begin{aligned}
\mathcal{X}_{jk,t}^{(i)} = {} & \nabla_{\theta_k}\left[\left(I - \frac{P_{t|t-1}^{(i)} H_i H_i'}{H_i' P_{t|t-1}^{(i)} H_i}\right) F_i\right] \nabla_{\theta_j} x_{t-1|t-1} \\
& + \nabla_{\theta_k}\left(I - \frac{P_{t|t-1}^{(i)} H_i H_i'}{H_i' P_{t|t-1}^{(i)} H_i}\right) \left(\nabla_{\theta_j} F_i x_{t-1|t-1} + \nabla_{\theta_j} G_i\right) \\
& \left(I - \frac{P_{t|t-1}^{(i)} H_i H_i'}{H_i' P_{t|t-1}^{(i)} H_i}\right) \nabla_{\theta_k}\left(\nabla_{\theta_j} F_i x_{t-1|t-1} + \nabla_{\theta_j} G_i\right) \\
& - \nabla_{\theta_k}\left(\frac{\nabla_{\theta_j} P_{t|t-1}^{(i)} H_i H_i' + P_{t|t-1}^{(i)} \nabla_{\theta_j} H_i H_i' + P_{t|t-1}^{(i)} H_i \nabla_{\theta_j} H_i'}{H_i' P_{t|t-1}^{(i)} H_i}\right) \left(F_i x_{t-1|t-1} + G_i\right) \\
& - \left(\frac{\nabla_{\theta_j} P_{t|t-1}^{(i)} H_i H_i' + P_{t|t-1}^{(i)} \nabla_{\theta_j} H_i H_i' + P_{t|t-1}^{(i)} H_i \nabla_{\theta_j} H_i'}{H_i' P_{t|t-1}^{(i)} H_i}\right) \nabla_{\theta_k}\left(F_i x_{t-1|t-1} + G_i\right) \\
& + \nabla_{\theta_k}\left(\frac{\left(P_{t|t-1}^{(i)} H_i H_i'\right)\left(\nabla_{\theta_j} H_i' P_{t|t-1}^{(i)} H_i + H_i' \nabla_{\theta_j} P_{t|t-1}^{(i)} H_i + H_i' P_{t|t-1}^{(i)} \nabla_{\theta_j} H_i\right)}{\left(H_i' P_{t|t-1}^{(i)} H_i\right)^2}\right) \left(F_i x_{t-1|t-1} + G_i\right) \\
& + \left(\frac{\left(P_{t|t-1}^{(i)} H_i H_i'\right)\left(\nabla_{\theta_j} H_i' P_{t|t-1}^{(i)} H_i + H_i' \nabla_{\theta_j} P_{t|t-1}^{(i)} H_i + H_i' P_{t|t-1}^{(i)} \nabla_{\theta_j} H_i\right)}{\left(H_i' P_{t|t-1}^{(i)} H_i\right)^2}\right) \nabla_{\theta_k}\left(F_i x_{t-1|t-1} + G_i\right) \\
& + \nabla_{\theta_k}\left(\frac{\nabla_{\theta_j} P_{t|t-1}^{(i)} H_i + P_{t|t-1}^{(i)} \nabla_{\theta_j} H_i}{H_i' P_{t|t-1}^{(i)} H_i}\right) \left(y_t - A_i' z_t\right) \\
& + \left(\frac{\nabla_{\theta_j} P_{t|t-1}^{(i)} H_i + P_{t|t-1}^{(i)} \nabla_{\theta_j} H_i}{H_i' P_{t|t-1}^{(i)} H_i}\right) \nabla_{\theta_k}\left(y_t - A_i' z_t\right) \\
& - \nabla_{\theta_k}\left(\frac{\left(P_{t|t-1}^{(i)} H_i\right)\left(\nabla_{\theta_j} H_i' P_{t|t-1}^{(i)} H_i + H_i' \nabla_{\theta_j} P_{t|t-1}^{(i)} H_i + H_i' P_{t|t-1}^{(i)} \nabla_{\theta_j} H_i\right)}{\left(H_i' P_{t|t-1}^{(i)} H_i\right)^2}\right) \left(y_t - A_i' z_t\right) \\
& - \left(\frac{\left(P_{t|t-1}^{(i)} H_i\right)\left(\nabla_{\theta_j} H_i' P_{t|t-1}^{(i)} H_i + H_i' \nabla_{\theta_j} P_{t|t-1}^{(i)} H_i + H_i' P_{t|t-1}^{(i)} \nabla_{\theta_j} H_i\right)}{\left(H_i' P_{t|t-1}^{(i)} H_i\right)^2}\right) \nabla_{\theta_k}\left(y_t - A_i' z_t\right) \\
& - \nabla_{\theta_k}\left(\frac{P_{t|t-1}^{(i)} H_i}{H_i' P_{t|t-1}^{(i)} H_i}\right) \nabla_{\theta_j} A_i' z_t - \left(\frac{P_{t|t-1}^{(i)} H_i}{H_i' P_{t|t-1}^{(i)} H_i}\right) \nabla_{\theta_j} \nabla_{\theta_k} A_i' z_t.
\end{aligned}
$$

Then, when $\delta_1 = \delta_2$, $\bar{\mathcal{X}}_{jk,t}^{(i)}$ can be achieved and

$$
\nabla_{\theta_j} \nabla_{\theta_k} \bar{x}_{t|t} = \left[\left(I - \frac{\bar{P}_{t|t-1} \bar{H} \bar{H}'}{\bar{H}' \bar{P}_{t|t-1} \bar{H}}\right) \bar{F}\right] \nabla_{\theta_j} \nabla_{\theta_k} \bar{x}_{t-1|t-1} + \bar{\mathcal{X}}_{jk,t},
$$

where

$$\bar{\mathcal{X}}_{jk,t} = \nabla_{\theta_j}\bar{\xi}_{t|t}\left(\nabla_{\theta_k}\bar{x}^{(1)}_{t|t} - \nabla_{\theta_k}\bar{x}^{(2)}_{t|t}\right) + \nabla_{\theta_k}\bar{\xi}_{t|t}\left(\nabla_{\theta_j}\bar{x}^{(1)}_{t|t} - \nabla_{\theta_j}\bar{x}^{(2)}_{t|t}\right) + \xi_*\bar{\mathcal{X}}^{(1)}_{jk,t} + (1-\xi_*)\bar{\mathcal{X}}^{(2)}_{jk,t}.$$

□

The next lemma provides stochastic bounds for $\bar{\xi}_{t+1|t}$ and its derivatives.

Lemma A.1,1 *Suppose Assumption 5 hold. Then, there exists an open neighborhood of* (β_*, δ_*), *denoted by* $B(\beta_*, \delta_*)$, *and a sequence of strictly stationary and ergodic random variables* $\{\lambda_t\}$ *satisfying* $E\lambda_t^{1+c} < M < \infty$ *for some* $c, M > 0$, *such that:*

$$\sup_{(\beta,\delta_1)\in B(\beta_*,\delta_*)} \left|\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\bar{\xi}_{t+1|t}\right|^{\frac{\alpha(k)}{k}} < \lambda_t \quad (t = 1, ..., T)$$

for all $i_1, ..., i_k \in \{1, .., 2n_\delta + n_\beta\}$ *and* $k = 1, 2, 3$ *and* 4, *where* $\alpha(k) = 6$ *if* $k = 1, 2, 3$ *and* $\alpha(k) = 5$ *if* $k = 4$. *The above inequalities hold uniformly over* $\epsilon \le p, q \le 1-\epsilon$ *with* ϵ *being an arbitrary number satisfying* $0 < \epsilon < 1/2$.

Proof of Lemma A.1.1 We use the difference equations in Lemma 1 to relate $\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\bar{\xi}_{t+1|t}$ to the density functions $\bar{f}_{1t}$ and $\bar{f}_{2t}$ and their derivatives. Because the higher order derivatives depend successively on the lower orders, we start with $k = 1$. Without loss of generality, suppose $j \in I_1$. Then, apply (B.6):

$$\left|\nabla_{\theta_j}\bar{\xi}_{t+1|t}\right|^6 \le \left(\sum_{s=0}^{t-1}\left|r\rho^s\left(\frac{\nabla_{\theta_j}\bar{f}_{1(t-s)}}{\bar{f}_{t-s}} - \frac{\nabla_{\theta_j}\bar{f}_{2(t-s)}}{\bar{f}_{t-s}}\right)\right|\right)^6 \le 2\left(\sum_{s=0}^{\infty}|r\rho^s|\,\upsilon_{t-s}^{1/6}\right)^6 \le 2\left(\sum_{s=0}^{\infty}(1-\epsilon)^s\upsilon_{t-s}^{1/6}\right)^6,$$

where the second inequality follows from Assumption 5 and the last inequality uses $\rho = p + q - 1$. Because $\{\upsilon_t\}$ is stationary and ergodic, the right hand side is also stationary and ergodic (White, 2001, Theorem 3.35). Denote it by λ_t and apply Minkowski's inequality for an infinite sum:

$$\begin{aligned} E\lambda_t^{1+c} &= 2E\left[\sum_{s=0}^{\infty}(1-\epsilon)^s\upsilon_{t-s}^{1/6}\right]^{6(1+c)} \le 2\left\{\sum_{s=0}^{\infty}\left[E((1-\epsilon)^s\upsilon_{t-s}^{1/6})^{6(1+c)}\right]^{\frac{1}{6(1+c)}}\right\}^{6(1+c)} \\ &= 2\left\{\sum_{s=0}^{\infty}(1-\epsilon)^s\left[E\upsilon_{t-s}^{1+c}\right]^{\frac{1}{6(1+c)}}\right\}^{6(1+c)} \le 2L\left\{\sum_{s=0}^{\infty}(1-\epsilon)^s\right\}^{6(1+c)}, \end{aligned}$$

where the last inequality holds because $E\upsilon_{t-s}^{1+c}$ is finite by Assumption 5. Because $\sum_{s=0}^{\infty}(1-\epsilon)^s = 1/\epsilon < \infty$, we have $E\lambda_t^{1+c} \le 2L/\epsilon^{6(1+c)} < \infty$. This establishes the result for $k = 1$ by setting $M = 2L/\epsilon^{6(1+c)}$.

The proof for $k>1$ is similar. For $k=2$, we have $|\nabla_{\theta_j}\nabla_{\theta_i}\bar{\xi}_{t+1|t}|^3 \leq (\sum_{s=0}^{\infty}\left|\rho^s\bar{\mathcal{E}}_{ji,t-s}\right|)^3$. I provide upper bounds for $|\bar{\mathcal{E}}_{ji,t}|$ for five possible cases. Specifically, if $j\in I_0$ and $i\in I_1$, then

$$\begin{aligned}\left|\bar{\mathcal{E}}_{ji,t}\right| &= r\left|\frac{\nabla_{\theta_j}\nabla_{\theta_i}\bar{f}_{1t}}{\bar{f}_t} - \frac{\nabla_{\theta_j}\nabla_{\theta_i}\bar{f}_{2t}}{\bar{f}_t} + \frac{\nabla_{\theta_j}\bar{f}_{2t}}{\bar{f}_t}\frac{\nabla_{\theta_i}\bar{f}_{2t}}{\bar{f}_t} - \frac{\nabla_{\theta_j}\bar{f}_{2t}}{\bar{f}_t}\frac{\nabla_{\theta_i}\bar{f}_{1t}}{\bar{f}_t}\right| \\ &\leq r\left(\left|\frac{\nabla_{\theta_j}\nabla_{\theta_i}\bar{f}_{1t}}{\bar{f}_t}\right| + \left|\frac{\nabla_{\theta_j}\nabla_{\theta_i}\bar{f}_{2t}}{\bar{f}_t}\right| + \left|\frac{\nabla_{\theta_j}\bar{f}_{2t}}{\bar{f}_t}\frac{\nabla_{\theta_i}\bar{f}_{2t}}{\bar{f}_t}\right| + \left|\frac{\nabla_{\theta_j}\bar{f}_{2t}}{\bar{f}_t}\frac{\nabla_{\theta_i}\bar{f}_{1t}}{\bar{f}_t}\right|\right) \\ &\leq 4\left|r\right|\upsilon_t^{1/3}.\end{aligned}$$

The same bound holds if $j\in I_0$ and $i\in I_2$. If $j\in I_1$ and $i\in I_1$, then $|\bar{\mathcal{E}}_{ji,t}|\leq 4\left|\rho\left(1-2\xi_*\right)\right|\lambda_{t-1}^{1/6}\upsilon_t^{1/6} + 2\left(2\left|r\right| + \left|r(2\xi_*-1)\right|\right)\upsilon_t^{1/3}$. The same bound holds if $j\in I_1$ and $i\in I_2$, and $j\in I_2$ and $i\in I_2$. Consequently, there exists a finite constant C_1, such that for all the three cases we have $|\bar{\mathcal{E}}_{ji,t}|\leq C_1(\lambda_{t-1}^{1/6}\upsilon_t^{1/6}+\upsilon_t^{1/3})$. This implies $\left|\nabla_{\theta_j}\nabla_{\theta_i}\bar{\xi}_{t+1|t}\right|^3 \leq \left(\sum_{s=0}^{\infty}C_1(1-\epsilon)^s(\lambda_{t-1}^{1/6}\upsilon_t^{1/6}+\upsilon_t^{1/3})\right)^3$. The right side is stationary and ergodic; we continue to denote it by λ_t. By Minkowski's inequality:

$$E\lambda_t^{1+c} \leq \left\{\sum_{s=0}^{\infty}\left[E\left(C_1(1-\epsilon)^s(\lambda_{t-1}^{1/6}\upsilon_t^{1/6}+\upsilon_t^{1/3})\right)^{3(1+c)}\right]^{\frac{1}{3(1+c)}}\right\}^{3(1+c)}. \tag{B.19}$$

Apply Minkowski's inequality followed by the Cauchy–Schwarz inequality to the summands:

$$\begin{aligned}&E\left(C_1(1-\epsilon)^s(\lambda_{t-1}^{1/6}\upsilon_t^{1/6}+\upsilon_t^{1/3})\right)^{3(1+c)} \\ &\leq (C_1(1-\epsilon)^s)^{3(1+c)}\left[\left(E\lambda_{t-1}^{(1+c)/2}\upsilon_t^{(1+c)/2}\right)^{\frac{1}{3(1+c)}} + \left(E\upsilon_t^{(1+c)}\right)^{\frac{1}{3(1+c)}}\right]^{3(1+c)} \\ &\leq (C_1(1-\epsilon)^s)^{3(1+c)}\left[\left(E\lambda_{t-1}^{(1+c)}E\upsilon_t^{(1+c)}\right)^{\frac{1}{6(1+c)}} + \left(E\upsilon_t^{(1+c)}\right)^{\frac{1}{3(1+c)}}\right]^{3(1+c)}.\end{aligned}$$

Because $E\lambda_{t-1}^{(1+c)}<M$ and $E\upsilon_t^{(1+c)}<L$, the last term in the preceding display is no greater than

$$(1-\epsilon)^{3(1+c)s}C_1^{3(1+c)}\left[(ML)^{\frac{1}{6(1+c)}} + L^{\frac{1}{3(1+c)}}\right]^{3(1+c)} \leq C_2(1-\epsilon)^{3(1+c)s}, \tag{B.20}$$

where C_2 is a finite constant independent of p and q. Plug this into (B.19), we have $E\lambda_t^{1+c}\leq C_2(\sum_{s=0}^{\infty}(1-\epsilon)^s)^{3(1+c)} = C_2/\epsilon^{3(1+c)}<\infty$. This proves the result for $k=2$.

Now, consider $k=3$. Inspecting the expressions of $\bar{\mathcal{E}}_{jil,t}$ reported in the proof of Lemma 1 shows that they comprise the following terms $(a, b, c = 1, 2)$:

$$\begin{aligned} &\frac{\nabla_{\theta_j}\nabla_{\theta_i}\bar{f}_{at}\nabla_{\theta_l}\bar{f}_{bt}}{\bar{f}_t^2}, \frac{\nabla_{\theta_j}\nabla_{\theta_i}\nabla_{\theta_l}\bar{f}_{at}}{\bar{f}_t}, \frac{\nabla_{\theta_j}\bar{f}_{at}\nabla_{\theta_i}\bar{f}_{bt}\nabla_{\theta_l}\bar{f}_{ct}}{\bar{f}_t^3}, \frac{\nabla_{\theta_l}\bar{f}_{at}\nabla_{\theta_j}\nabla_{\theta_i}\bar{\xi}_{t|t-1}}{\bar{f}_t}, \\ &\frac{\nabla_{\theta_j}\bar{f}_{at}\nabla_{\theta_l}\bar{f}_{bt}\nabla_{\theta_i}\bar{\xi}_{t|t-1}}{\bar{f}_t^2}, \frac{\nabla_{\theta_j}\nabla_{\theta_i}\bar{f}_{at}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t}, \frac{\nabla_{\theta_j}\bar{f}_{at}\nabla_{\theta_i}\bar{\xi}_{t|t-1}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t}. \end{aligned} \tag{B.21}$$

By Assumption 4 and the above results for $k = 1$ and 2, the quantities in (B.21) are bounded, respectively, by $\upsilon_t^{1/2}, \upsilon_t^{1/2}, \upsilon_t^{1/2}, \upsilon_t^{1/6}\lambda_{t-1}^{1/3}, \upsilon_t^{1/3}\lambda_{t-1}^{1/6}, \upsilon_t^{1/3}\lambda_{t-1}^{1/6}$ and $\upsilon_t^{1/6}\lambda_{t-1}^{1/3}$. Therefore, the ten cases specified in Lemma 1 all satisfy $\left|\bar{\mathcal{E}}_{jil,t}\right| \leq C_3(\upsilon_t^{1/2} + \upsilon_t^{1/6}\lambda_{t-1}^{1/3} + \upsilon_t^{1/3}\lambda_{t-1}^{1/6})$, where C_3 is a finite constant independent of p and q. This implies $\left|\nabla_{\theta_j}\nabla_{\theta_i}\nabla_{\theta_l}\bar{\xi}_{t+1|t}\right|^2 \leq \left|\sum_{s=0}^{\infty}(1-\epsilon)^s C_3(\upsilon_t^{1/2} + \upsilon_t^{1/6}\lambda_{t-1}^{1/3} + \upsilon_t^{1/3}\lambda_{t-1}^{1/6})\right|^2$. Denote the right hand side by λ_t and proceed along the same lines as between (B.19) and (B.20). It then follows that $E\lambda_t^{1+c} < \infty$. For $k = 4$, the expressions of $\bar{\mathcal{E}}_{jilm,t}$, although omitted here, include terms as in (B.21) but with the orders of derivatives sum to 4 instead of 3. Using the same arguments as between (B.19) and (B.20), it can be shown that $E\lambda_t^{1+c} < \infty$ holds. □

The follow two lemmas provides stochastic bounds for $\bar{x}_{t|t}$, $\bar{P}_{t|t}$ and their derivatives.

Lemma A.1.2 *Suppose Assumption 5 hold. Then, there exists an open neighborhood of* (β_*, δ_*) *, denoted by* $B(\beta_*, \delta_*)$ *, and a sequence of strictly stationary and ergodic random variables* $\{\lambda_{i,t}\}$ *satisfying* $E\lambda_{i,t}^{1+c_i} < M_i < \infty$ *for some* $c_i, M_i > 0$*, such that for the i-th entry of* $\bar{x}_{t|t}$ *and its derivatives:*

$$\sup_{(\beta,\delta_1)\in B(\beta_*,\delta_*)} \left|\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\bar{x}_{t|t,i}\right|^{\frac{\alpha(k)}{k}} < \lambda_{i,t} \quad (t = 1, ..., T)$$

for all $i_1, ..., i_k \in \{1, .., 2n_\delta + n_\beta\}$ *and* $k = 1, 2, 3$ *and* 4*, where* $\alpha(k) = 6$ *if* $k = 1, 2, 3$ *and* $\alpha(k) = 5$ *if* $k = 4$*. The above inequalities hold uniformly over* $\epsilon \leq p, q \leq 1 - \epsilon$ *with* ϵ *being an arbitrary number satisfying* $0 < \epsilon < 1/2$.

Lemma A.1.3 *Suppose Assumption 5 hold. Then, there exists an open neighborhood of* (β_*, δ_*) *, denoted by* $B(\beta_*, \delta_*)$ *, and a sequence of strictly stationary and ergodic random variables* $\{\lambda_{ij,t}\}$ *satisfying* $E\lambda_{ij,t}^{1+c_{ij}} < M_{ij} < \infty$ *for some* $c_{ij}, M_{ij} > 0$*, such that for the (i,j)-th entry of* $\bar{P}_{t|t}$ *and its derivatives:*

$$\sup_{(\beta,\delta_1)\in B(\beta_*,\delta_*)} \left|\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\bar{P}_{t|t,ij}\right|^{\frac{\alpha(k)}{k}} < \lambda_{ij,t} \quad (t = 1, ..., T)$$

for all $i_1, ..., i_k \in \{1, .., 2n_\delta + n_\beta\}$ *and* $k = 1, 2, 3$ *and* 4*, where* $\alpha(k) = 6$ *if* $k = 1, 2, 3$ *and* $\alpha(k) = 5$ *if* $k = 4$*. The above inequalities hold uniformly over* $\epsilon \leq p, q \leq 1 - \epsilon$ *with* ϵ *being an arbitrary number satisfying* $0 < \epsilon < 1/2$.

Proof of Lemma A.1.2-A.1.3 : The proofs for Lemma A.1(2)-(3) are omitted here since they follow the similar proofs in Lemma A.1.1.

The next lemma establishes stochastic orders of some quantities related to $\xi_{t|t-1}$, f_{1t} and f_{2t}. The quantities are all evaluated at $(\tilde{\beta}, \tilde{\delta}, \tilde{\delta})$.

Lemma A.2 *Let i_s, j_s, l_s, m_s, n_s be arbitrary integers satisfying $1 \leq i_s, j_s, l_s, m_s, n_s \leq 2n_\delta + n_\beta$ for $s \in \{1, 2, 3, 4\}$. The following results hold uniformly over $\epsilon \leq p, q \leq 1 - \epsilon$ with ϵ being an arbitrary number satisfying $0 < \epsilon < 1/2$:*

1. *For any $a \in \{1, 2\}, u \in \{1, 2, 3, 4\}$ and $v \in \{0, 1, 2, 3\}$ satisfying $u + v \leq 4$, we have (interpret $\nabla_{\theta_{j_1}} ... \nabla_{\theta_{j_v}} \tilde{\xi}_{t|t-1}$ as 1 when $v = 0$)*

$$\frac{1}{T} \sum_{t=1}^{T} \frac{\nabla_{\theta_{i_1}} ... \nabla_{\theta_{i_u}} \tilde{f}_{at}}{\tilde{f}_t} \nabla_{\theta_{j_1}} ... \nabla_{\theta_{j_v}} \tilde{\xi}_{t|t-1} = o_p(1), \tag{B.22}$$

Further, if $u + v \leq 3$, then the result holds with $o_p(1)$ replaced by $O_p(T^{-1/2})$.

2. *For any $(a, b, c) \in \{1, 2\}, (u, w) \in \{1, 2, 3\}$ and $v \in \{0, 1, 2\}$ satisfying $u + v + w \leq 4$:*

$$\frac{1}{T} \sum_{t=1}^{T} \frac{\nabla_{\theta_{i_1}} ... \nabla_{\theta_{i_u}} \tilde{f}_{at}}{\tilde{f}_t} \frac{\nabla_{\theta_{j_1}} ... \nabla_{\theta_{j_v}} \tilde{f}_{bt}}{\tilde{f}_t} \frac{\nabla_{\theta_{l_1}} ... \nabla_{\theta_{l_w}} \tilde{f}_{ct}}{\tilde{f}_t} = O_p(1).$$

3. *For any $(a, b, c) \in \{1, 2\}, (u, w) \in \{1, 2, 3\}$ and $(v, z) \in \{0, 1\}$ satisfying $u + v + w + z \leq 3$:*

$$\frac{1}{T} \sum_{t=1}^{T} \frac{\nabla_{\theta_{i_1}} \tilde{f}_{at}}{\tilde{f}_t} \frac{\nabla_{\theta_{j_1}} ... \nabla_{\theta_{j_u}} \tilde{f}_{bt}}{\tilde{f}_t} \frac{\nabla_{\theta_{l_1}} ... \nabla_{\theta_{l_v}} \tilde{f}_{ct}}{\tilde{f}_t} \nabla_{\theta_{m_1}} ... \nabla_{\theta_{m_w}} \tilde{\xi}_{t|t-1} \nabla_{\theta_{n_1}} ... \nabla_{\theta_{n_z}} \tilde{\xi}_{t|t-1} = O_p(1).$$

Proof of Lemma A.2. By the mean value theorem, the left hand side of (B.22) equals

$$\begin{aligned} & T^{-1} \sum_{t=1}^{T} \frac{\nabla_{\theta_{i_1}} ... \nabla_{\theta_{i_u}} f_{at}^*}{f_t^*} \nabla_{\theta_{j_1}} ... \nabla_{\theta_{j_v}} \xi_{t|t-1}^* \\ & + \left\{ T^{-3/2} \sum_{t=1}^{T} \nabla_{\theta'} \left(\frac{\nabla_{\theta_{i_1}} ... \nabla_{\theta_{i_u}} \bar{f}_{at}}{\bar{f}_{at}} \nabla_{\theta_{j_1}} ... \nabla_{\theta_{j_v}} \bar{\xi}_{t|t-1} \right) \right\} T^{1/2} \left(\tilde{\theta} - \theta_* \right), \end{aligned} \tag{B.23}$$

where "$*$" and "$-$" denote that the relevant quantities are evaluated at the true values $\theta'_* = (\beta'_*, \delta'_*, \delta'_*)$ and $\bar{\theta}' = \left(\bar{\beta}', \bar{\delta}', \bar{\delta}' \right)$, where $\bar{\theta}$ lies between $\tilde{\theta}' = (\tilde{\beta}', \tilde{\delta}', \tilde{\delta}')$ and θ_*. The first summation is over terms that are stationary and ergodic, which are bounded by $\lambda_t^{v/\alpha(k)} \upsilon_t^{u/\alpha(k)}$ by Assumption 5 and Lemma

A.1. Apply Hölder's inequality:

$$\begin{aligned}E(\lambda_t^{v/\alpha(k)}\upsilon_t^{u/\alpha(k)})^{1+c} &\leq \left(E\left(\lambda_t^{v(1+c)/\alpha(k)}\right)^{\alpha(k)/v}\right)^{\frac{v}{\alpha(k)}}\left(E\left(\upsilon_t^{u(1+c)/\alpha(k)}\right)^{\alpha(k)/(\alpha(k)-v)}\right)^{\frac{\alpha(k)-v}{\alpha(k)}} \\ &\leq \left(E\lambda_t^{1+c}\right)^{\frac{v}{\alpha(k)}}\left(E\upsilon_t^{1+c}\right)^{\frac{\alpha(k)-v}{\alpha(k)}}\end{aligned}$$

where the last inequality follows because $u+v<\alpha(k)$. Both terms on the right hand side are finite by Assumption 5 and Lemma A.1. Therefore, the first term in the display (B.23) is $o_p(1)$ by Theorem 3.34 in White (2001). Now turn to the second term in the display (B.23). We have, for any $k\in\{1,...,2n_\delta+n_\beta\}$:

$$\begin{aligned}&\left|T^{-3/2}\sum_{t=1}^{T}\nabla_{\theta_k}\left(\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_u}}\bar{f}_{at}}{\bar{f}_{at}}\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_v}}\bar{\xi}_{t|t-1}\right)\right| \\ &\leq\left|T^{-3/2}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_u}}\nabla_{\theta_k}\bar{f}_{at}}{\bar{f}_{at}}\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_v}}\bar{\xi}_{t|t-1}\right|+\left|T^{-3/2}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_u}}\bar{f}_{at}}{\bar{f}_{at}}\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_v}}\nabla_{\theta_k}\bar{\xi}_{t|t-1}\right| \\ &+\left|T^{-3/2}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_u}}\bar{f}_{at}}{\bar{f}_{at}}\frac{\nabla_{\theta_k}\bar{f}_{at}}{\bar{f}_{at}}\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_v}}\bar{\xi}_{t|t-1}\right| \\ &\leq T^{-3/2}\sum_{t=1}^{T}\left\{2\upsilon_t^{(u+1)/\alpha(k)}\lambda_t^{v/\alpha(k)}+\upsilon_t^{u/\alpha(k)}\lambda_t^{(v+1)/\alpha(k)}\right\}=O_p\left(T^{-1/2}\right),\end{aligned}$$

where the equality follows from Assumption 4, Lemma A.1 and $u+v+1\leq 5$. Therefore, the display (B.23) is $o_p(1)$.

Now we consider the cases with $u+v\leq 3$. If $u+v<3$, then the terms inside the first summation of (B.23) are bounded by $\lambda_t^{v/6}\upsilon_t^{u/6}$. We have

$$E(\lambda_t^{v/6}\upsilon_t^{u/6})^{2(1+c)}\leq\left(E(\lambda_t^{v(1+c)/3})^{\frac{3}{v}}\right)^{v/3}\left(E(\upsilon_t^{u(1+c)/3})^{\frac{3}{(3-v)}}\right)^{(3-v)/3}\leq\left(E\lambda_t^{1+c}\right)^{v/3}\left(E\upsilon_t^{1+c}\right)^{(3-v)/3}.$$

The right hand side is finite. If $u+v=3$, i.e., $u=3$ and $v=0$, then $E(\lambda_t^{v/6}\upsilon_t^{u/6})^{2(1+c)}=E\upsilon_t{}^{(1+c)}<\infty$. Apply the central limit theorem; it follows that the left hand side of (B.22) is $O_p\left(T^{-1/2}\right)$.

Lemma A2.2 and A2.3 can be proved using the same arguments, i.e., first applying the mean value theorem and then obtaining bounds for the two resulting terms separately. It follows that the left hand side quantity in Lemma A2.2 is bounded by $T^{-1}\sum_{t=1}^{T}\upsilon_t^{(u+v+w)/\alpha(k)}+O_p(T^{-1/2})$, while that in Lemma A2.3 is bounded by $T^{-1}\sum_{t=1}^{T}\upsilon_t^{(1+u+v)/\alpha(k)}\lambda_t^{(w+z)/\alpha(k)}+O_p(T^{-1/2})$. The two leading terms both satisfy a law of large numbers, therefore are $O_p(1)$. □

Lemma A.3 *Under the null hypothesis and Assumptions 1-5, for all $k, l, m \in \{1, ..., n_\delta\}$, we have:*

1. Let e_k be an n_δ-dimensional unit vector whose k-th element equals 1, then

$$\begin{bmatrix} \nabla_{\delta_{2k}}\hat{\beta}(\tilde{\delta}) \\ \xi_* \nabla_{\delta_{2k}}\hat{\delta}_1(\tilde{\delta}) \end{bmatrix} = (\xi_* - 1) \begin{bmatrix} 0 \\ e_k \end{bmatrix} + O_p(T^{-1/2}).$$

2. The second order derivatives satisfy

$$\begin{bmatrix} \nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\tilde{\delta}) \\ \xi_* \nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\tilde{\delta}) \end{bmatrix} = -\tilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\tilde{D}_{kl,t} + O_p(T^{-1/2}).$$

3. The third order derivatives satisfy

$$\begin{bmatrix} \nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\beta}(\tilde{\delta}) \\ \xi_* \nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\delta}_1(\tilde{\delta}) \end{bmatrix} = O_p(1).$$

Proof of Lemma A.3 As the proof is long, we organize it into three parts, corresponding to Lemma A.3.1, A.3.2 and A.3.3 respectively.

Proof of Lemma A.3.1 By construction, $\hat{\theta}(\delta_2)$ satisfies the first order conditions:

$$\mathcal{M}_j^{(1)}(p, q, \delta_2) = \frac{1}{T}\sum_{t=1}^{T}\frac{\hat{M}_t}{\hat{B}_t} = 0, \qquad (j \in \{1, ..., n_\beta + n_\delta\}), \tag{B.24}$$

where

$$\begin{aligned} \hat{B}_t &= \hat{f}_{1t}\hat{\xi}_{t|t-1} + \hat{f}_{2t}(1 - \hat{\xi}_{t|t-1}), \\ \hat{M}_{jt} &= (\nabla_{\theta_j}\hat{f}_{1t} - \nabla_{\theta_j}\hat{f}_{2t})\hat{\xi}_{t|t-1} + (\hat{f}_{1t} - \hat{f}_{2t})\nabla_{\theta_j}\hat{\xi}_{t|t-1} + \nabla_{\theta_j}\hat{f}_{2t}. \end{aligned} \tag{B.25}$$

Because (B.24) holds for any $\delta_2 \in \Delta$, its derivative with respect to δ_2 must equal zero. The proof exploits this fact. It consists of three steps. The first step takes first order derivatives of the $n_\beta + n_\delta$ restrictions in (B.24) with respect to δ_{2k}, where $k \in \{1, ..., n_\delta\}$, obtain a system of $n_\beta + n_\delta$ linear equations with $\nabla_{\delta_{2k}}\hat{\beta}(\delta_2)$ and $\nabla_{\delta_{2k}}\hat{\delta}_1(\delta_2)$ being the unknowns. The second step evaluates these equations at $\delta_2 = \tilde{\delta}$ and obtains approximation to them. The third step solves these approximating equations to obtain explicit expressions for $\nabla_{\delta_{2k}}\hat{\beta}(\delta_2)$ and $\nabla_{\delta_{2k}}\hat{\delta}_1(\delta_2)$. These three steps are then repeated for every $k \in \{1, ..., n_\delta\}$ to obtain the lemma.

Step 1 for proving Lemma. Pick an arbitrary $k \in \{1, ..., n_\delta\}$ and an arbitrary $j \in \{1, ..., n_\beta + n_\delta\}$. Differentiate the j-th equation from (B.24) with respect to the k-th element of δ_2 to obtain

$$\mathcal{M}_{jk}^{(2)}(p, q, \delta_2) = \frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\hat{M}_{jt}}{\hat{B}_t} - \frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\hat{B}_t}{\hat{B}_t^2}\hat{M}_{jt} = 0, \tag{B.26}$$

where

$$\begin{aligned}\nabla_{\delta_{2k}}\hat{M}_{jt} \;&=\; \Big\{\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{1t} + \left(1-\hat{\xi}_{t|t-1}\right)\nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{2t} + \left(\nabla_{\theta_j}\hat{f}_{1t} - \nabla_{\theta_j}\hat{f}_{2t}\right)\nabla_{\theta'}\hat{\xi}_{t|t-1} \\ &+\nabla_{\theta_j}\hat{\xi}_{t|t-1}\left(\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta'}\hat{f}_{2t}\right) + (\hat{f}_{1t} - \hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\hat{\theta}(\delta_2)\end{aligned} \tag{B.27}$$

and

$$\nabla_{\delta_{2k}}\hat{B}_t = \left\{\hat{\xi}_{t|t-1}\nabla_{\theta'}\hat{f}_{1t} + \left(1-\hat{\xi}_{t|t-1}\right)\nabla_{\theta'}\hat{f}_{2t} + \left(\hat{f}_{1t} - \hat{f}_{2t}\right)\nabla_{\theta'}\hat{\xi}_{t|t-1}\right\}\nabla_{\delta_{2k}}\hat{\theta}(\delta_2) \tag{B.28}$$

with

$$\nabla_{\delta_{2k}}\hat{\theta}(\delta_2) = \begin{bmatrix} \nabla_{\delta_{2k}}\hat{\beta}(\delta_2) \\ \nabla_{\delta_{2k}}\hat{\delta}_1(\delta_2) \\ e_k \end{bmatrix}, \tag{B.29}$$

where e_k is an n_δ dimensional vector, whose k-th element equals 1 and the others zero. We view (B.26) as a linear equation with the first $n_\beta + n_\delta$ elements of $\nabla_{\delta_{2k}}\hat{\theta}(\delta_2)$ being the unknowns.

The above differentiation can be carried for all $j = 1, ..., n_\beta + n_\delta$ while keeping k fixed at the same value. This delivers $n_\beta + n_\delta$ equations with the same number of unknown contained in (B.29).

Step 2 for proving Lemma. We evaluate the term $T^{-1}\sum_{t=1}^{T}(\nabla_{\delta_{2k}}\hat{B}_t/\hat{B}_t^2)\hat{M}_{jt}$ in (B.26) at $\delta_2 = \tilde{\delta}$ for an arbitrary $j \in \{1, ..., n_\beta + n_\delta\}$. It equals

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\tilde{f}_{1t} + (1-\xi_*)\nabla_{\theta_j}\tilde{f}_{2t}}{\tilde{f}_t^2}\left[\xi_*\nabla_{\theta'}\tilde{f}_{1t} + (1-\xi_*)\nabla_{\theta'}\tilde{f}_{2t}\right]\nabla_{\delta_{2k}}\hat{\theta}(\tilde{\delta}).$$

Using (B.29), this can be rewritten as

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_j}\tilde{f}_{2t}}{\tilde{f}_t^2}\left[\xi_*\nabla_{\beta'}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\beta'}\tilde{f}_{2t}\right]\nabla_{\delta_{2k}}\hat{\beta}(\tilde{\delta}) \\
+ \ & \frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_j}\tilde{f}_{2t}}{\tilde{f}_t^2}\left[\xi_*\nabla_{\delta_1'}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\delta_1'}\tilde{f}_{2t}\right]\nabla_{\delta_{2k}}\hat{\delta}_1(\tilde{\delta}) \\
+ \ & \frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_j}\tilde{f}_{2t}}{\tilde{f}_t^2}\left[\xi_*\nabla_{\delta_{2k}}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\delta_{2k}}\tilde{f}_{2t}\right]. \qquad \text{(B.30)}
\end{aligned}
$$

Now, let j run through the set $\{1,...,n_\beta+n_\delta\}$. Let

$$
\tilde{I}=\frac{1}{T}\sum_{t=1}^{T}\left[\begin{array}{cc}
\frac{\xi_*\nabla_\beta\tilde{f}_{1t}+(1-\xi_*)\nabla_\beta\tilde{f}_{2t}}{\tilde{f}_t}\frac{\xi_*\nabla_{\beta'}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\beta'}\tilde{f}_{2t}}{\tilde{f}_t} & \frac{\xi_*\nabla_\beta\tilde{f}_{1t}+(1-\xi_*)\nabla_\beta\tilde{f}_{2t}}{\tilde{f}_t}\frac{\xi_*\nabla_{\delta_1'}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\delta_1'}\tilde{f}_{2t}}{\tilde{f}_t} \\
\frac{\xi_*\nabla_{\delta_1}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\delta_1}\tilde{f}_{2t}}{\tilde{f}_t}\frac{\xi_*\nabla_{\beta'}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\beta'}\tilde{f}_{2t}}{\tilde{f}_t} & \frac{\xi_*\nabla_{\delta_1}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\delta_1}\tilde{f}_{2t}}{\tilde{f}_t}\frac{\xi_*\nabla_{\delta_1'}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\delta_1'}\tilde{f}_{2t}}{\tilde{f}_t}
\end{array}\right].
$$

Then, (B.30) can be written as

$$
\tilde{I}\left[\begin{array}{c}\nabla_{\delta_{2k}}\hat{\beta}(\tilde{\delta}) \\ \nabla_{\delta_{2k}}\hat{\delta}_1(\tilde{\delta})\end{array}\right]+\left[\begin{array}{c}
\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_\beta\tilde{f}_{1t}+(1-\xi_*)\nabla_\beta\tilde{f}_{2t}}{\tilde{f}_t}\frac{\xi_*\nabla_{\delta_{2k}}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\delta_{2k}}\tilde{f}_{2t}}{\tilde{f}_t} \\
\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\delta_1}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\delta_1}\tilde{f}_{2t}}{\tilde{f}_t}\frac{\xi_*\nabla_{\delta_{2k}}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\delta_{2k}}\tilde{f}_{2t}}{\tilde{f}_t}
\end{array}\right], \qquad \text{(B.31)}
$$

where $\tilde{I}$ is the observed information evaluated at the null estimates. Now we turn to the first term at $\delta_2=\tilde{\delta}$ in (B.26). It equals

$$
\left\{\frac{1}{T}\sum_{t=1}^{T}\left[\xi_*\frac{\nabla_{\theta_j}\nabla_{\theta'}\tilde{f}_{1t}}{\tilde{f}_t}+(1-\xi_*)\frac{\nabla_{\theta_j}\nabla_{\theta'}\tilde{f}_{2t}}{\tilde{f}_t}+\left(\frac{\nabla_{\theta_j}\tilde{f}_{1t}-\nabla_{\theta_j}\tilde{f}_{2t}}{\tilde{f}_t}\right)\nabla_{\theta'}\hat{\xi}_{t|t-1}+\nabla_{\theta_j}\hat{\xi}_{t|t-1}\left(\frac{\nabla_{\theta'}\tilde{f}_{1t}-\nabla_{\theta'}\tilde{f}_{2t}}{\tilde{f}_t}\right)\right]\right\}\nabla_{\delta_{2k}}\hat{\theta}(\tilde{\delta}).
$$

All the terms inside the curly brackets are $O_p(T^{-1/2})$. Their effects are dominated by $\tilde{I}$ which is positive definite in large sample. Combining this fact with (B.26) and (B.31) , we obtain

$$
\tilde{I}\left[\begin{array}{c}\nabla_{\delta_{2k}}\hat{\beta}(\tilde{\delta}) \\ \nabla_{\delta_{2k}}\hat{\delta}_1(\tilde{\delta})\end{array}\right]=-\left[\begin{array}{c}
\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_\beta\tilde{f}_{1t}+(1-\xi_*)\nabla_\beta\tilde{f}_{2t}}{\tilde{f}_t}\frac{\xi_*\nabla_{\delta_{2k}}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\delta_{2k}}\tilde{f}_{2t}}{\tilde{f}_t} \\
\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\delta_1}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\delta_1}\tilde{f}_{2t}}{\tilde{f}_t}\frac{\xi_*\nabla_{\delta_{2k}}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\delta_{2k}}\tilde{f}_{2t}}{\tilde{f}_t}
\end{array}\right]+O_p(T^{-1/2}). \qquad \text{(B.32)}
$$

To solve this, it is important to check $\xi_*\nabla_{\delta_{1k}}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\delta_{1k}}\tilde{f}_{2t}$ and $\xi_*\nabla_{\delta_{2k}}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\delta_{2k}}\tilde{f}_{2t}$. By the results in (B.13) and (B.18), we have

$$
\nabla_{\delta_{2k}}\tilde{P}_{t-1|t-1}=\left(\frac{1-\xi_*}{\xi_*}\right)\nabla_{\delta_{1k}}\tilde{P}_{t-1|t-1},
$$

and

$$\nabla_{\delta_{2k}}\tilde{x}_{t-1|t-1} = \left(\frac{1-\xi_*}{\xi_*}\right)\nabla_{\delta_{1k}}\tilde{x}_{t-1|t-1}.$$

Applying (4.3) and the above two equalities, we have

$$\xi_*\nabla_{\delta_{2k}}\tilde{f}_{1t} + (1-\xi_*)\nabla_{\delta_{2k}}\tilde{f}_{2t} = \left(\frac{1-\xi_*}{\xi_*}\right)\left[\xi_*\nabla_{\delta_{1k}}\tilde{f}_{1t} + (1-\xi_*)\nabla_{\delta_{1k}}\tilde{f}_{2t}\right]. \tag{B.33}$$

Then we can solve the unknowns in (B.32) and get

$$\begin{bmatrix} \nabla_{\delta_{2k}}\hat{\beta}(\tilde{\delta}) \\ \nabla_{\delta_{2k}}\hat{\delta}_1(\tilde{\delta}) \end{bmatrix} = \left(\frac{\xi_*-1}{\xi_*}\right)\begin{bmatrix} 0 \\ e_k \end{bmatrix} + O_p(T^{-1/2}).$$

□

Proof of Lemma A.3.2. View the quantities in (B.26) as functions of δ_2, p and q, and differentiate it with respect to the l-th component of δ_2 $(l = 1, ..., n_\delta)$:

$$\begin{aligned} &\mathcal{M}^{(3)}_{jkl}(p,q,\delta_2) \\ =\;& \frac{1}{T}\sum_{t=1}^{T}\left\{\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{M}_{jt}}{\hat{B}_t} - \frac{\nabla_{\delta_{2k}}\hat{M}_{jt}\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t^2} - \frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t^2}\hat{M}_{jt} - \frac{\nabla_{\delta_{2l}}\hat{M}_{jt}\nabla_{\delta_{2k}}\hat{B}_t}{\hat{B}_t^2} + 2\frac{\nabla_{\delta_{2k}}\hat{B}_t\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t^3}\hat{M}_{jt}\right\} \\ =\;& 0, \end{aligned} \tag{B.34}$$

where

$$\begin{aligned} \nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{M}_{jt} \;=\;& \sum_{s=1}^{n_\beta+2n_\delta}\Big\{\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\hat{f}_{1t} + \hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t} \\ & -\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\hat{f}_{2t} + (1-\hat{\xi}_{t|t-1})\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t} \\ & +(\nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_s}\hat{\xi}_{t|t-1} + (\nabla_{\theta_j}\hat{f}_{1t} - \nabla_{\theta_j}\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1} \\ & +(\nabla_{\theta_s}\hat{f}_{1t} - \nabla_{\theta_s}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta'}\hat{\xi}_{t|t-1} + (\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_j}\hat{\xi}_{t|t-1} \\ & +(\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_s}\hat{\xi}_{t|t-1} + (\hat{f}_{1t} - \hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\hat{\theta}_s(\delta_2)\nabla_{\delta_{2l}}\hat{\theta}(\delta_2) \\ & +\Big\{\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{1t} + \left(1-\hat{\xi}_{t|t-1}\right)\nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{2t} + \left(\nabla_{\theta_j}\hat{f}_{1t} - \nabla_{\theta_j}\hat{f}_{2t}\right)\nabla_{\theta'}\hat{\xi}_{t|t-1} + \\ & +\nabla_{\theta_j}\hat{\xi}_{t|t-1}\left(\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta'}\hat{f}_{2t}\right) + (\hat{f}_{1t} - \hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}(\delta_2) \end{aligned}$$

and

$$\begin{aligned}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{B}_t &= \sum_{s=1}^{n_\beta+2n_\delta}\Big\{\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\hat{f}_{1t}+\hat{\xi}_{t|t-1}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\hat{f}_{2t}+(1-\hat{\xi}_{t|t-1})\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t}\\ &\quad+(\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_s}\hat{\xi}_{t|t-1}+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\hat{\theta}_s(\delta_2)\nabla_{\delta_{2l}}\hat{\theta}(\delta_2)\\ &\quad+\Big\{\hat{\xi}_{t|t-1}\nabla_{\theta'}\hat{f}_{1t}+(1-\hat{\xi}_{t|t-1})\nabla_{\theta'}\hat{f}_{2t}+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}(\delta_2).\end{aligned}$$

To study (B.34), we can start with the third term $T^{-1}\sum_{t=1}^{T}\left[\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{B}_t/\hat{B}_t^2\right]\hat{M}_{jt}$. At $\tilde{\theta}$, it equals

$$\begin{aligned}&\sum_{u=1}^{n_\beta+2n_\delta}\sum_{s=1}^{n_\beta+2n_\delta}\Big\{\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_j}\tilde{f}_{2t}}{\tilde{f}_t^2}[\nabla_{\theta_u}\tilde{\xi}_{t|t-1}\nabla_{\theta_s}\tilde{f}_{1t}+\xi_*\nabla_{\theta_s}\nabla_{\theta_u}\tilde{f}_{1t}-\nabla_{\theta_u}\tilde{\xi}_{t|t-1}\nabla_{\theta_s}\tilde{f}_{2t}\\ &\qquad+(1-\xi_*)\nabla_{\theta_s}\nabla_{\theta_u}\tilde{f}_{2t}+(\nabla_{\theta_u}\tilde{f}_{1t}-\nabla_{\theta_u}\tilde{f}_{2t})\nabla_{\theta_s}\tilde{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\hat{\theta}_s(\tilde{\delta})\nabla_{\delta_{2l}}\hat{\theta}_u(\tilde{\delta})\\ &+\sum_{s=1}^{n_\beta+2n_\delta}\Big\{\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_j}\tilde{f}_{2t}}{\tilde{f}_t^2}[\xi_*\nabla_{\theta_s}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_s}\tilde{f}_{2t}]\Big\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\tilde{\delta}).\end{aligned}$$

Because $\nabla_{\delta_{2k}}\hat{\theta}_s(\tilde{\delta})$ and $\nabla_{\delta_{2l}}\hat{\theta}_u(\tilde{\delta})$ are $O_p(T^{-1/2})$ except for the following four cases $s=n_\beta+k$, $s=$

$n_\beta + n_\delta + k$, $u = n_\beta + l$ and $u = n_\beta + n_\delta + l$, the preceding display equals

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_j}\tilde{f}_{2t}}{\tilde{f}_t^2}\Big\{\nabla_{\delta_{1l}}\tilde{\xi}_{t|t-1}\nabla_{\delta_{1k}}\tilde{f}_{1t}+\xi_*\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\tilde{f}_{1t}-\nabla_{\delta_{1l}}\tilde{\xi}_{t|t-1}\nabla_{\delta_{1k}}\tilde{f}_{2t}\\
&+(1-\xi_*)\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\tilde{f}_{2t}+(\nabla_{\delta_{1l}}\tilde{f}_{1t}-\nabla_{\delta_{1l}}\tilde{f}_{2t})\nabla_{\delta_{1k}}\tilde{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\hat{\delta}_{1k}(\tilde{\delta})\nabla_{\delta_{2l}}\hat{\delta}_{1l}(\tilde{\delta})\\
+\;&\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_j}\tilde{f}_{2t}}{\tilde{f}_t^2}\Big\{\nabla_{\delta_{2l}}\tilde{\xi}_{t|t-1}\nabla_{\delta_{1k}}\tilde{f}_{1t}+\xi_*\nabla_{\delta_{1k}}\nabla_{\delta_{2l}}\tilde{f}_{1t}-\nabla_{\delta_{2l}}\tilde{\xi}_{t|t-1}\nabla_{\delta_{1k}}\tilde{f}_{2t}\\
&+(1-\xi_*)\nabla_{\delta_{1k}}\nabla_{\delta_{2l}}\tilde{f}_{2t}+(\nabla_{\delta_{2l}}\tilde{f}_{1t}-\nabla_{\delta_{2l}}\tilde{f}_{2t})\nabla_{\delta_{1k}}\tilde{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\hat{\delta}_{1k}(\tilde{\delta})\nabla_{\delta_{2l}}\hat{\delta}_{2l}(\tilde{\delta})\\
+\;&\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_j}\tilde{f}_{2t}}{\tilde{f}_t^2}\Big\{\nabla_{\delta_{1l}}\tilde{\xi}_{t|t-1}\nabla_{\delta_{2k}}\tilde{f}_{1t}+\xi_*\nabla_{\delta_{2k}}\nabla_{\delta_{1l}}\tilde{f}_{1t}-\nabla_{\delta_{1l}}\tilde{\xi}_{t|t-1}\nabla_{\delta_{2k}}\tilde{f}_{2t}\\
&+(1-\xi_*)\nabla_{\delta_{2k}}\nabla_{\delta_{1l}}\tilde{f}_{2t}+(\nabla_{\delta_{1l}}\tilde{f}_{1t}-\nabla_{\delta_{1l}}\tilde{f}_{2t})\nabla_{\delta_{2k}}\tilde{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\hat{\delta}_{2k}(\tilde{\delta})\nabla_{\delta_{2l}}\hat{\delta}_{1l}(\tilde{\delta})\\
+\;&\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_j}\tilde{f}_{2t}}{\tilde{f}_t^2}\Big\{\nabla_{\delta_{2l}}\tilde{\xi}_{t|t-1}\nabla_{\delta_{2k}}\tilde{f}_{1t}+\xi_*\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\tilde{f}_{1t}-\nabla_{\delta_{2l}}\tilde{\xi}_{t|t-1}\nabla_{\delta_{2k}}\tilde{f}_{2t}\\
&+(1-\xi_*)\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\tilde{f}_{2t}+(\nabla_{\delta_{2l}}\tilde{f}_{1t}-\nabla_{\delta_{2l}}\tilde{f}_{2t})\nabla_{\delta_{2k}}\tilde{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\hat{\delta}_{2k}(\tilde{\delta})\nabla_{\delta_{2l}}\hat{\delta}_{2l}(\tilde{\delta})\\
+\;&\sum_{s=1}^{n_\beta+n_\delta}\left\{\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_j}\tilde{f}_{2t}}{\tilde{f}_t^2}[\xi_*\nabla_{\theta_s}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_s}\tilde{f}_{2t}]\right\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\tilde{\delta})+O_p(T^{-1/2}).
\end{aligned}$$

Apply $\nabla_{\delta_{2l}}\hat{\delta}_{1l}(\tilde{\delta}) = (\xi_*-1)/\xi_* + O_p(T^{-1/2})$, $\nabla_{\delta_{2k}}\hat{\delta}_{1k}(\tilde{\delta}) = (\xi_*-1)/\xi_* + O_p(T^{-1/2})$, $\nabla_{\delta_{2l}}\hat{\delta}_{2l}(\tilde{\delta}) = \nabla_{\delta_{2k}}\hat{\delta}_{2k}(\tilde{\delta}) = 1$ and rearrange terms, the preceding display further reduces to

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_j}\tilde{f}_{2t}}{\tilde{f}_t^2}\\
&\times\Bigg\{\left(\frac{1-\xi_*}{\xi_*}\right)^2\left[\xi_*\frac{\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\tilde{f}_{1t}}{\tilde{f}_t}+(1-\xi_*)\,\frac{\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\tilde{f}_{2t}}{\tilde{f}_t}\right]+\left[\xi_*\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\tilde{f}_{1t}}{\tilde{f}_t}+(1-\xi_*)\,\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\tilde{f}_{2t}}{\tilde{f}_t}\right]\\
&-\left(\frac{1-\xi_*}{\xi_*}\right)\left[\xi_*\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{1l}}\tilde{f}_{1t}}{\tilde{f}_t}+(1-\xi_*)\,\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{1l}}\tilde{f}_{2t}}{\tilde{f}_t}+\xi_*\frac{\nabla_{\delta_{1k}}\nabla_{\delta_{2l}}\tilde{f}_{1t}}{\tilde{f}_t}+(1-\xi_*)\,\frac{\nabla_{\delta_{1k}}\nabla_{\delta_{2l}}\tilde{f}_{2t}}{\tilde{f}_t}\right]\\
&+\frac{1}{\xi_*^2}\left[\nabla_{\delta_{1k}}\tilde{\xi}_{t|t-1}\left(\frac{\nabla_{\theta_l}\bar{f}_{1t}}{\bar{f}_t}-\frac{\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t}\right)+\left(\frac{\nabla_{\theta_k}\bar{f}_{1t}}{\bar{f}_t}-\frac{\nabla_{\theta_k}\bar{f}_{2t}}{\bar{f}_t}\right)\nabla_{\delta_{1l}}\tilde{\xi}_{t|t-1}\right]\Bigg\}\\
&+\sum_{s=1}^{n_\beta+n_\delta}\left\{\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_j}\tilde{f}_{2t}}{\tilde{f}_t^2}[\xi_*\nabla_{\theta_s}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_s}\tilde{f}_{2t}]\right\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\tilde{\delta})+O_p(T^{-1/2}).
\end{aligned}$$

The above display leads to $(n_\beta+n_\delta)$ equations with $j = 1, ..., n_\beta+n_\delta$. These equations can be written collectively as

$$\tilde{I}\begin{bmatrix} \nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\tilde{\delta}) \\ \nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\tilde{\delta}) \end{bmatrix} + \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_\beta \tilde{f}_{1t}+(1-\xi_*)\nabla_\beta \tilde{f}_{2t}}{\tilde{f}_t^2}\tilde{U}_{kl,t} \\ \frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\delta_1}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\delta_1}\tilde{f}_{2t}}{\tilde{f}_t^2}\tilde{U}_{kl,t} \end{bmatrix} + O_p(T^{-1/2}),$$

where

$$\begin{aligned}\tilde{U}_{kl,t} = & \left(\frac{1-\xi_*}{\xi_*}\right)^2\left[\xi_*\frac{\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\tilde{f}_{1t}}{\tilde{f}_t}+(1-\xi_*)\frac{\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\tilde{f}_{2t}}{\tilde{f}_t}\right]+\left[\xi_*\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\tilde{f}_{1t}}{\tilde{f}_t}+(1-\xi_*)\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\tilde{f}_{2t}}{\tilde{f}_t}\right] \\ & -\left(\frac{1-\xi_*}{\xi_*}\right)\left[\xi_*\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{1l}}\tilde{f}_{1t}}{\tilde{f}_t}+(1-\xi_*)\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{1l}}\tilde{f}_{2t}}{\tilde{f}_t}+\xi_*\frac{\nabla_{\delta_{1k}}\nabla_{\delta_{2l}}\tilde{f}_{1t}}{\tilde{f}_t}+(1-\xi_*)\frac{\nabla_{\delta_{1k}}\nabla_{\delta_{2l}}\tilde{f}_{2t}}{\tilde{f}_t}\right] \\ & +\frac{1}{\xi_*^2}\left[\nabla_{\delta_{1k}}\tilde{\xi}_{t|t-1}\left(\frac{\nabla_{\theta_l}\bar{f}_{1t}}{\bar{f}_t}-\frac{\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t}\right)+\left(\frac{\nabla_{\theta_k}\bar{f}_{1t}}{\bar{f}_t}-\frac{\nabla_{\theta_k}\bar{f}_{2t}}{\bar{f}_t}\right)\nabla_{\delta_{1l}}\tilde{\xi}_{t|t-1}\right].\end{aligned}$$

This completes the analysis of the third term in (B.34). As shown below, the other terms in (B.34) are all asymptotically negligible. Note that at $\delta_2 = \tilde{\delta}$, $\nabla_{\delta_{2k}}\hat{B}_t$ can be rewritten as

$$\begin{aligned} & \sum_{s=1}^{n_\beta}\nabla_{\theta_s}\tilde{f}_{1t}\nabla_{\delta_{2k}}\hat{\theta}_s(\tilde{\delta})+\sum_{s=n_\beta+1}^{n_\beta+2n_\delta}\left[\xi_*\nabla_{\theta_s}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_s}\tilde{f}_{2t}\right]\nabla_{\delta_{2k}}\hat{\theta}_s(\tilde{\delta}) \\ = & \sum_{s=1}^{n_\beta}\nabla_{\theta_s}\tilde{f}_{1t}\nabla_{\delta_{2k}}\hat{\beta}_s(\tilde{\delta})+\sum_{s=1,s\neq k}^{n_\delta}\left[\xi_*\nabla_{\theta_s}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_s}\tilde{f}_{2t}\right]\nabla_{\delta_{2k}}\hat{\delta}_{1s}(\tilde{\delta}) \\ & +\left[\xi_*\nabla_{\delta_{1k}}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\delta_{1k}}\tilde{f}_{2t}\right]\left[\nabla_{\delta_{2k}}\hat{\delta}_{1k}(\tilde{\delta})+\left(\frac{1-\xi_*}{\xi_*}\right)\right]. \end{aligned} \quad \text{(B.35)}$$

This representation is useful because, corresponding to the three terms on the right hand side, we have $\nabla_{\delta_{2k}}\hat{\beta}_s(\tilde{\delta}) = O_p(T^{-1/2})$, $\nabla_{\delta_{2k}}\hat{\delta}_{1s}(\tilde{\delta}) = O_p(T^{-1/2})$ when $s \neq k$ and $\nabla_{\delta_{2k}}\hat{\delta}_{1k}(\tilde{\delta})+(1-\xi_*)/\xi_* = O_p(T^{-1/2})$ by the previous results. So, the second, fourth and fifth terms are all $O_p(T^{-1/2})$ in (B.34) after applying (B.35) to $\nabla_{\delta_{2k}}\hat{B}_t$. □

Proof of Lemma A.3.3. View the quantities in (B.34) as functions of δ_2, p and q and differentiate them with respect to the h-th element of δ_2 $(h=1,...,n_\delta)$:

$$\mathcal{M}^{(4)}_{jklh}(p,q,\delta_2) = \frac{1}{T}\sum_{t=1}^{T}\Bigg\{\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{M}_{jt}}{\hat{B}_t} - \frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{M}_{jt}\nabla_{\delta_{2h}}\hat{B}_t}{\hat{B}_t^2} \tag{B.36}$$
$$- \frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2h}}\hat{M}_{jt}\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t^2} - \frac{\nabla_{\delta_{2k}}\hat{M}_{jt}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{B}_t}{\hat{B}_t^2} + 2\frac{\nabla_{\delta_{2k}}\hat{M}_{jt}\nabla_{\delta_{2l}}\hat{B}_t\nabla_{\delta_{2h}}\hat{B}_t}{\hat{B}_t^3}$$
$$- \frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{B}_t}{\hat{B}_t^2}\hat{M}_{jt} - \frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t^2}\nabla_{\delta_{2h}}\hat{M}_{jt} + 2\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{B}_t\nabla_{\delta_{2h}}\hat{B}_t}{\hat{B}_t^3}\hat{M}_{jt}$$
$$- \frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2h}}\hat{B}_t}{\hat{B}_t^2}\nabla_{\delta_{2l}}\hat{M}_{jt} - \frac{\nabla_{\delta_{2k}}\hat{B}_t}{\hat{B}_t^2}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{M}_{jt} + 2\frac{\nabla_{\delta_{2k}}\hat{B}_t\nabla_{\delta_{2h}}\hat{B}_t}{\hat{B}_t^3}\nabla_{\delta_{2l}}\hat{M}_{jt}$$
$$+ 2\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2h}}\hat{B}_t\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t^3}\hat{M}_{jt} + 2\frac{\nabla_{\delta_{2k}}\hat{B}_t\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{B}_t}{\hat{B}_t^3}\hat{M}_{jt}$$
$$+2\frac{\nabla_{\delta_{2k}}\hat{B}_t\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t^3}\nabla_{\delta_{2h}}\hat{M}_{jt} - 6\frac{\nabla_{\delta_{2k}}\hat{B}_t\nabla_{\delta_{2l}}\hat{B}_t\nabla_{\delta_{2h}}\hat{B}_t}{\hat{B}_t^4}\hat{M}_{jt}.\Bigg\} = 0.$$

Among the fifteen terms, only the 1st and the 6th term involve third order derivatives. They will be analyzed later. Among the remaining terms, we have the following five cases: (1) The 4th, 7th and 9th terms involve second order derivatives of $\hat{B}_t$ and first order derivatives of $\hat{M}_{jt}$, which lead to: $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\tilde{f}_{at}/\tilde{f}_t)(\nabla_{\theta_{j_1}}\nabla_{\theta_{j_2}}\tilde{f}_{bt}/\tilde{f}_t)$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\tilde{f}_{at}/\tilde{f}_t)(\nabla_{\theta_{i_2}}\tilde{f}_{bt}/\tilde{f}_t)\nabla_{\theta_{j_1}}\tilde{\xi}_{t|t-1}\nabla_{\theta_{j_2}}\tilde{\xi}_{t|t-1}$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\tilde{f}_{at}/\tilde{f}_t)(\nabla_{\theta_{j_1}}\nabla_{\theta_{j_2}}\tilde{f}_{bt}/\tilde{f}_t)$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\tilde{f}_{at}/\tilde{f}_t)(\nabla_{\theta_{j_1}}\tilde{f}_{bt}/\tilde{f}_t)\nabla_{\theta_{j_2}}\tilde{\xi}_{t|t-1}$ and $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\tilde{f}_{at}/\tilde{f}_t)\nabla_{\theta_{j_1}}\tilde{\xi}_{t|t-1}$, where $1\leq i_1,i_2,j_1,j_2\leq n_\beta+2n_\delta$, $a=1,2$ and $b=1,2$. They are all $O_p(1)$ by Lemma A.2. (2) The 2rd, 3rd and 10th terms consist of first order derivatives of $\hat{B}_t$ and second order derivatives of $\hat{M}_{jt}$. They lead to: $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\tilde{f}_{at}/\tilde{f}_t)(\nabla_{\theta_{j_1}}\nabla_{\theta_{j_2}}\nabla_{\theta_{j_3}}\tilde{f}_{bt}/\tilde{f}_t)$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\tilde{f}_{at}/\tilde{f}_t)(\nabla_{\theta_{j_1}}\tilde{f}_{bt}/\tilde{f}_t)\nabla_{\theta_{j_2}}\tilde{\xi}_{t|t-1}$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\tilde{f}_{at}/\tilde{f}_t)(\nabla_{\theta_{i_2}}\tilde{f}_{bt}/\tilde{f}_t)\nabla_{\theta_{j_1}}\nabla_{\theta_{j_2}}\tilde{\xi}_{t|t-1}$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\tilde{f}_{at}/\tilde{f}_t)(\nabla_{\theta_{j_1}}\tilde{f}_{bt}/\tilde{f}_t)$ and $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\tilde{f}_{at}/\tilde{f}_t)(\nabla_{\theta_{j_1}}\tilde{f}_{bt}/\tilde{f}_t)\nabla_{\theta_{j_2}}\tilde{\xi}_{t|t-1}$, which are all $O_p(1)$. These three terms are thus $O_p(T^{-1/2})$ after applying (B.35) to the first order derivatives of $\hat{B}_t$. (3) The 5th, 11th and 14th terms consist of: $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\tilde{f}_{at}/\tilde{f}_t)(\nabla_{\theta_{i_2}}\tilde{f}_{bt}/\tilde{f}_t)(\nabla_{\theta_{j_1}}\nabla_{\theta_{j_2}}\tilde{f}_{ct}/\tilde{f}_t)$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\tilde{f}_{at}/\tilde{f}_t)(\nabla_{\theta_{i_2}}\tilde{f}_{bt}/\tilde{f}_t)(\nabla_{\theta_{i_3}}\tilde{f}_{ct}/\tilde{f}_t)\nabla_{\theta_{j_1}}\tilde{\xi}_{t|t-1}$, which are all $O_p(1)$. Consequently, these three terms are $O_p(T^{-1/2})$ after applying (B.35). (4) The 8th, 12th and 13th terms lead to: $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\tilde{f}_{at}/\tilde{f}_t)(\nabla_{\theta_{j_1}}\tilde{f}_{bt}/\tilde{f}_t)(\nabla_{\theta_{j_2}}\tilde{f}_{ct}/\tilde{f}_t)$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\tilde{f}_{at}/\tilde{f}_t)(\nabla_{\theta_{i_2}}\tilde{f}_{bt}/\tilde{f}_t)(\nabla_{\theta_{j_1}}\nabla_{\theta_{j_2}}\tilde{f}_{ct}/\tilde{f}_t)$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\tilde{f}_{at}/\tilde{f}_t)(\nabla_{\theta_{i_2}}\tilde{f}_{bt}/\tilde{f}_t)(\nabla_{\theta_{i_3}}\tilde{f}_{ct}/\tilde{f}_t)\nabla_{\theta_{j_1}}\tilde{\xi}_{t|t-1}$, which are all $O_p(1)$. (5) The 15th term consists of $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\tilde{f}_{at}/\tilde{f}_t)(\nabla_{\theta_{i_2}}\tilde{f}_{bt}/\tilde{f}_t)(\nabla_{\theta_{i_3}}\tilde{f}_{ct}/\tilde{f}_t)(\nabla_{\theta_{i_4}}\tilde{f}_{ct}/\tilde{f}_{ct})$. This term is $O_p\left(T^{-1/2}\right)$ after applying (B.35).

To analyze the remaining two terms in (B.36), we need third order derivatives of $\hat{M}_{jt}$ and $\hat{B}_t$:

$$
\begin{aligned}
&\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{M}_{jt} \\
&= \sum_{u=1}^{n_\beta+2n_\delta}\sum_{s=1}^{n_\beta+2n_\delta}\Big\{\nabla_{\theta_u}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t} + \nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\hat{f}_{1t} + \hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{1t} \\
&+ \nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{1t} - \nabla_{\theta_u}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t} - \nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\hat{f}_{2t} \\
&+ (1-\hat{\xi}_{t|t-1})\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{2t} - \nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{2t} \\
&+ (\nabla_{\theta_j}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta_j}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_s}\hat{\xi}_{t|t-1} + (\nabla_{\theta_j}\nabla_{\theta_u}\hat{f}_{1t} - \nabla_{\theta_j}\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1} \\
&+ (\nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta_u}\hat{\xi}_{t|t-1} + (\nabla_{\theta_j}\hat{f}_{1t} - \nabla_{\theta_j}\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1} \\
&+ (\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_u}\hat{\xi}_{t|t-1} + (\nabla_{\theta_s}\hat{f}_{1t} - \nabla_{\theta_s}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1} \\
&+ (\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_j}\hat{\xi}_{t|t-1} + (\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{1t} - \nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta'}\hat{\xi}_{t|t-1} \\
&+ (\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_s}\hat{\xi}_{t|t-1} + (\nabla_{\theta_u}\hat{f}_{1t} - \nabla_{\theta_u}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1} \\
&+ (\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta_u}\hat{\xi}_{t|t-1} \\
&+ (\hat{f}_{1t} - \hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\hat{\theta}_s(\delta_2)\nabla_{\delta_{2l}}\hat{\theta}_u(\delta_2)\nabla_{\delta_{2h}}\hat{\theta}(\delta_2) \\
&+ \sum_{u=1}^{n_\beta+2n_\delta}\sum_{s=1}^{n_\beta+2n_\delta}\Big\{\nabla_{\theta_u}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\hat{f}_{1t} + \hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{1t} - \nabla_{\theta_u}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\hat{f}_{2t} \\
&+ (1-\hat{\xi}_{t|t-1})\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{2t} + (\nabla_{\theta_j}\nabla_{\theta_u}\hat{f}_{1t} - \nabla_{\theta_j}\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\theta_s}\hat{\xi}_{t|t-1} \\
&+ (\nabla_{\theta_j}\hat{f}_{1t} - \nabla_{\theta_j}\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta_u}\hat{\xi}_{t|t-1} + (\nabla_{\theta_s}\hat{f}_{1t} - \nabla_{\theta_s}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_u}\hat{\xi}_{t|t-1} \\
&+ (\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{1t} - \nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\theta_j}\hat{\xi}_{t|t-1} + (\nabla_{\theta_u}\hat{f}_{1t} - \nabla_{\theta_u}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_s}\hat{\xi}_{t|t-1} \\
&+ (\hat{f}_{1t} - \hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta_u}\hat{\xi}_{t|t-1}\Big\}[\nabla_{\delta_{2k}}\nabla_{\delta_{2h}}\hat{\theta}_s(\delta_2)\nabla_{\delta_{2l}}\hat{\theta}_u(\delta_2) + \nabla_{\delta_{2k}}\hat{\theta}_s(\delta_2)\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\theta}_u(\delta_2)] \\
&+ \sum_{s=1}^{n_\beta+2n_\delta}\Big\{\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t} + \nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\hat{f}_{1t} + (1-\hat{\xi}_{t|t-1})\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t} \\
&- \nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\hat{f}_{2t} + (\nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_s}\hat{\xi}_{t|t-1} + (\nabla_{\theta_j}\hat{f}_{1t} - \nabla_{\theta_j}\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1} \\
&+ (\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_j}\hat{\xi}_{t|t-1} + (\nabla_{\theta_s}\hat{f}_{1t} - \nabla_{\theta_s}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta'}\hat{\xi}_{t|t-1} \\
&+ (\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_s}\hat{\xi}_{t|t-1} + (\hat{f}_{1t} - \hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2h}}\hat{\theta}(\delta_2)\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\delta_2), \\
&+ \Big\{\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{1t} + (1-\xi_{t|t-1})\nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{2t} + (\nabla_{\theta_j}\hat{f}_{1t} - \nabla_{\theta_j}\hat{f}_{2t})\nabla_{\theta'}\hat{\xi}_{t|t-1} \\
&+ (\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_j}\hat{\xi}_{t|t-1} + (\hat{f}_{1t} - \hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\theta}(\delta_2),
\end{aligned}
$$

and

$$
\begin{aligned}
&\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{B}_t \\
&= \sum_{s=1}^{n_\beta+2n_\delta}\sum_{u=1}^{n_\beta+2n_\delta}\Big\{\nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\hat{f}_{1t}+\nabla_{\theta_u}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t}+\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{1t} \\
&+\hat{\xi}_{t|t-1}\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\hat{f}_{2t}-\nabla_{\theta_u}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t}+(1-\hat{\xi}_{t|t-1})\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{2t} \\
&-\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{2t}+(\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_s}\hat{\xi}_{t|t-1}+(\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1} \\
&+(\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta_u}\hat{\xi}_{t|t-1}+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2l}}\hat{\theta}_u(\delta_2)\nabla_{\delta_{2k}}\hat{\theta}_s(\delta_2)\nabla_{\delta_{2h}}\hat{\theta}(\delta_2) \\
&+\sum_{s=1}^{n_\beta+2n_\delta}\sum_{u=1}^{n_\beta+2n_\delta}\Big\{\nabla_{\theta_u}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\hat{f}_{1t}+\hat{\xi}_{t|t-1}\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\theta_u}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\hat{f}_{2t}+(1-\hat{\xi}_{t|t-1})\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{2t} \\
&+(\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\theta_s}\hat{\xi}_{t|t-1}+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta_u}\hat{\xi}_{t|t-1}\Big\}\times \\
&\Big[\nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\hat{\theta}_u(\delta_2)\nabla_{\delta_{2k}}\hat{\theta}_s(\delta_2)+\nabla_{\delta_{2l}}\hat{\theta}_u(\delta_2)\nabla_{\delta_{2h}}\nabla_{\delta_{2k}}\hat{\theta}_s(\delta_2)\Big] \\
&+\sum_{s=1}^{n_\beta+2n_\delta}\Big\{\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\hat{f}_{1t}+\hat{\xi}_{t|t-1}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t}+(1-\hat{\xi}_{t|t-1})\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t}-\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\hat{f}_{2t} \\
&+(\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_s}\hat{\xi}_{t|t-1}+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2h}}\hat{\theta}(\delta_2)\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\delta_2) \\
&+\Big[\hat{\xi}_{t|t-1}\nabla_{\theta'}\hat{f}_{1t}+(1-\hat{\xi}_{t|t-1})\nabla_{\theta'}\hat{f}_{2t}+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big]\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\theta}(\delta_2).
\end{aligned}
$$

Consider the 1st term in (B.36). In the expression of $\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{M}_{jt}$, only the last two lines involve third order derivatives of $\hat{\theta}(\delta_2)$. These derivatives are multiplied by (after division by $\tilde{f}_t$): $T^{-1}\sum_{t=1}^{T}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\tilde{f}_{at}/\tilde{f}_t$ and $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\tilde{f}_{at}/\tilde{f}_t)\nabla_{\theta_{j_1}}\tilde{\xi}_{t|t-1}$, where $a=1,2$. They are $O_p(T^{-1/2})$ by Lemma A.2. The remaining components of $\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{M}_{jt}$ lead to: $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\tilde{f}_{at}/\tilde{f}_t)$ for $a=1,2$ and $k\le 4$ and $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\tilde{f}_{at}/\tilde{f}_t)(\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_m}}\tilde{\xi}_{t|t-1})$ for $a=1,2$ and $k+m\le 4$. They are all $o_p(1)$ by Lemma A.2. Therefore the contribution of $\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{M}_{jt}$ to (B.36) is $o_p(1)$. Finally, we turn to the 6th term in (B.36). In the expression for $\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{B}_t$, only the final line involves third order derivatives of $\hat{\theta}(\delta_2)$. It can be analyzed in the same way as the second term in (B.26); see Step 2 of the proof there. The remaining components, multiplied by $\hat{M}_{jt}$, lead to: $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\tilde{f}_{at}/\tilde{f}_t)(\nabla_{\theta_{i_2}}\tilde{f}_{bt}/\tilde{f}_t)(\nabla_{\theta_{j_1}}\nabla_{\theta_{j_2}}\tilde{\xi}_{t|t-1})$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\tilde{f}_{at}/\tilde{f}_t)(\nabla_{\theta_{j_1}}\tilde{f}_{bt}/\tilde{f}_t)$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\tilde{f}_{at}/\tilde{f}_t)(\nabla_{\theta_{j_1}}\tilde{f}_{bt}/\tilde{f}_t)(\nabla_{\theta_{j_2}}\tilde{\xi}_{t|t-1})$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}\tilde{f}_{at}/\tilde{f}_t)(\nabla_{\theta_{j_1}}\tilde{f}_{bt}/\tilde{f}_t)$ and $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\tilde{f}_{at}/\tilde{f}_t)(\nabla_{\theta_{i_2}}\tilde{f}_{bt}/\tilde{f}_t)(\nabla_{\theta_{j_1}}\tilde{\xi}_{t|t-1})$ for $a=1,2$ and $b=1,2$. They are all $O_p(1)$ by Lemma

A.2. This implies the desired result. □

Proof of Lemma 4. The first order derivative of $\mathcal{L}(p,q,\delta_2)$ with respect to δ_{2j} gives

$$\begin{aligned}
&\mathcal{L}_j^{(1)}(p,q,\delta_2)\\
=\;& \nabla_{\delta_{2j}}\mathcal{L}(p,q,\delta_2)\\
=\;& \sum_{t=1}^{T}\frac{1}{\hat{B}_t}\left(\nabla_{\theta'}\hat{f}_{1t}\hat{\xi}_{t|t-1}+\nabla_{\theta'}\hat{f}_{2t}(1-\hat{\xi}_{t|t-1})+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\theta'}\hat{\xi}_{t|t-1}\right)\nabla_{\delta_{2j}}\hat{\theta}(\delta_2)\\
=\;& \sum_{s=1}^{n_\beta+n_\delta}\left\{\sum_{t=1}^{T}\frac{1}{\hat{B}_t}\left(\nabla_{\theta_s}\hat{f}_{1t}\hat{\xi}_{t|t-1}+\nabla_{\theta_s}\hat{f}_{2t}(1-\hat{\xi}_{t|t-1})+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\theta_s}\hat{\xi}_{t|t-1}\right)\right\}\nabla_{\delta_{2j}}\hat{\theta}_s(\delta_2)\\
&+\sum_{t=1}^{T}\frac{1}{\hat{B}_t}\left(\nabla_{\delta_{2j}}\hat{f}_{1t}\hat{\xi}_{t|t-1}+\nabla_{\delta_{2j}}\hat{f}_{2t}(1-\hat{\xi}_{t|t-1})+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\delta_{2j}}\hat{\xi}_{t|t-1}\right),
\end{aligned}$$

where the third equality follows from the definition of $\nabla_{\delta_{2j}}\hat{\theta}(\delta_2)$. The term inside the curly brackets is zero because of the first order conditions determining $\hat{\beta}(\delta_2)$ and $\hat{\delta}_1(\delta_2)$. Thus,

$$\mathcal{L}_j^{(1)}(p,q,\delta_2)=\sum_{t=1}^{T}\frac{\hat{L}_{jt}}{\hat{B}_t},$$

where $\hat{B}_t$ is defined in (B.1) and $\hat{L}_{jt}=\nabla_{\delta_{2j}}\hat{f}_{1t}\hat{\xi}_{t|t-1}+\nabla_{\delta_{2j}}\hat{f}_{2t}(1-\hat{\xi}_{t|t-1})+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\delta_{2j}}\hat{\xi}_{t|t-1}$. When $\delta_2=\tilde{\delta}$, by (B.33), we have

$$\hat{L}_{jt}=\left(\frac{1-\xi_*}{\xi_*}\right)\hat{M}_{(n_\beta+j)t} \tag{B.37}$$

and

$$\mathcal{L}_j^{(1)}(p,q,\tilde{\delta})=\left(\frac{1-\xi_*}{\xi_*}\right)\mathcal{M}_{n_\beta+j}^{(1)}(p,q,\tilde{\delta})=0.$$

Consider the second result in the lemma. Differentiate $\mathcal{L}_j^{(1)}(p,q,\delta_2)$ with respect to the k-th component of δ_2:

$$\mathcal{L}_{jk}^{(2)}(p,q,\delta_2)=\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\hat{L}_{jt}}{\hat{B}_t}-\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\hat{B}_{jt}}{\hat{B}_t}\hat{L}_{jt},$$

where

$$\begin{aligned}
\nabla_{\delta_{2k}}\hat{L}_{jt} \;=\;& \{\hat{\xi}_{t|t-1}\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{f}_{1t}+(1-\hat{\xi}_{t|t-1})\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{f}_{2t}+(\nabla_{\delta_{2j}}\hat{f}_{1t}-\nabla_{\delta_{2j}}\hat{f}_{2t})\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&+(\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta'}\hat{f}_{2t})\nabla_{\delta_{2j}}\hat{\xi}_{t|t-1}+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{\xi}_{t|t-1}\}\nabla_{\delta_{2k}}\hat{\theta}(\delta_2).
\end{aligned}$$

Because $\mathcal{M}^{(2)}_{(n_\beta+j)k}(p,q,\delta_2)=0$, we have

$$\begin{aligned} T^{-1/2}\mathcal{L}^{(2)}_{jk}(p,q,\delta_2) &= T^{-1/2}\mathcal{L}^{(2)}_{jk}(p,q,\delta_2)-T^{1/2}\mathcal{M}^{(2)}_{(n_\beta+j)k}(p,q,\delta_2) \\ &= T^{-1/2}\sum_{t=1}^{T}\left\{\frac{\nabla_{\delta_{2k}}\hat{L}_{jt}}{\hat{B}_t}-\left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2k}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\} \\ &\quad -T^{-1/2}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\hat{B}_t}{\hat{B}_t^2}\left\{\hat{L}_{jt}-\left(\frac{1-\xi_*}{\xi_*}\right)\hat{M}_{(n_\beta+j)t}\right\}. \end{aligned}$$

The second summation on the right hand side equals 0 when evaluated at $\delta_2=\tilde{\delta}$ by (B.37). Consider the first summation, at $\delta_2=\tilde{\delta}$, $T^{-1/2}\sum_{t=1}^{T}\left[\nabla_{\delta_{2k}}\hat{L}_{jt}/\hat{B}_t\right]$ equals

$$\begin{aligned} \sum_{s=1}^{n_\beta+2n_\delta}\Bigg\{T^{-1/2}\sum_{t=1}^{T}\Bigg[&\frac{\xi_*\nabla_{\delta_{2j}}\nabla_{\theta_s}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\delta_{2j}}\nabla_{\theta_s}\tilde{f}_{2t}}{\tilde{f}_t} \\ &+\frac{(\nabla_{\delta_{2j}}\tilde{f}_{1t}-\nabla_{\delta_{2j}}\tilde{f}_{2t})\nabla_{\theta_s}\tilde{\xi}_{t|t-1}+(\nabla_{\theta_s}\tilde{f}_{1t}-\nabla_{\theta_s}\tilde{f}_{2t})\nabla_{\delta_{2j}}\tilde{\xi}_{t|t-1}}{\tilde{f}_t}\Bigg]\Bigg\}\nabla_{\delta_{2k}}\hat{\theta}_s(\tilde{\delta}). \end{aligned}$$

The terms in the curly brackets are all $O_p(1)$ while $\nabla_{\delta_{2k}}\hat{\theta}_s(\tilde{\delta})=O_p(T^{-1/2})$ unless $s=n_\beta+k$ or $s=n_\beta+n_\delta+k$. Similarly, at $\delta_2=\tilde{\delta}$, $T^{-1/2}\sum_{t=1}^{T}\left[\nabla_{\delta_{2k}}\hat{M}_{(n_\beta+j)t}/\hat{B}_t\right]$ equals

$$\begin{aligned} \sum_{s=1}^{n_\beta+2n_\delta}\Bigg\{T^{-1/2}\sum_{t=1}^{T}\Bigg[&\frac{\xi_*\nabla_{\delta_{1j}}\nabla_{\theta_s}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\delta_{1j}}\nabla_{\theta_s}\tilde{f}_{2t}}{\tilde{f}_t} \\ &+\frac{(\nabla_{\delta_{1j}}\tilde{f}_{1t}-\nabla_{\delta_{1j}}\tilde{f}_{2t})\nabla_{\theta_s}\tilde{\xi}_{t|t-1}+(\nabla_{\theta_s}\tilde{f}_{1t}-\nabla_{\theta_s}\tilde{f}_{2t})\nabla_{\delta_{1j}}\tilde{\xi}_{t|t-1}}{\tilde{f}_t}\Bigg]\Bigg\}\nabla_{\delta_{2k}}\hat{\theta}_s(\tilde{\delta}), \end{aligned}$$

and terms in the curly brackets are all $O_p(1)$ while $\nabla_{\delta_{2k}}\hat{\theta}_s(\tilde{\delta})=O_p(T^{-1/2})$ unless $s=n_\beta+k$ or $s=n_\beta+n_\delta+k$. Combining the two preceding displays, we have

$$T^{-1/2}\mathcal{L}^{(2)}_{jk}(p,q,\tilde{\delta})=T^{-1/2}\sum_{t=1}^{T}\tilde{U}_{jk,t}+o_p(1).$$

Consider the third order derivatives. We have

$$\begin{aligned}
& T^{-3/4}\mathcal{L}^{(3)}_{jkl}(p,q,\delta_2) - T^{1/4}\left(\frac{1-\xi_*}{\xi_*}\right)\mathcal{M}^{(3)}_{(n_\beta+j)kl}(p,q,\delta_2) \\
= \; & T^{-3/4}\sum_{t=1}^{T}\left\{\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{L}_{jt}}{\hat{B}_t} - \left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\} \\
& -T^{-3/4}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t}\left\{\frac{\nabla_{\delta_{2k}}\hat{L}_{jt}}{\hat{B}_t} - \left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2k}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\} \\
& -T^{-3/4}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\hat{B}_t}{\hat{B}_t}\left\{\frac{\nabla_{\delta_{2l}}\hat{L}_{jt}}{\hat{B}_t} - \left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2l}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\} \\
& -T^{-3/4}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t^2}\left\{\hat{L}_{jt} - \left(\frac{1-\xi_*}{\xi_*}\right)\hat{M}_{(n_\beta+j)t}\right\} \\
& +2T^{-3/4}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\hat{B}_t\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t^2}\left\{\hat{L}_{jt} - \left(\frac{1-\xi_*}{\xi_*}\right)\hat{M}_{(n_\beta+j)t}\right\},
\end{aligned} \tag{B.38}$$

where

$$\begin{aligned}
& \nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{L}_{jt} \\
= \; & \sum_{s=1}^{n_\beta+2n_\delta}\Big\{\hat{\xi}_{t|t-1}\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t} + (1-\hat{\xi}_{t|t-1})\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t} + (\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{f}_{1t} - \nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{f}_{2t})\nabla_{\theta'}\hat{\xi}_{t|t-1} \\
& +(\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\delta_{2j}}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_s}\hat{\xi}_{t|t-1} + (\nabla_{\delta_{2j}}\hat{f}_{1t} - \nabla_{\delta_{2j}}\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1} \\
& +(\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\delta_{2j}}\hat{\xi}_{t|t-1} + (\nabla_{\theta_s}\hat{f}_{1t} - \nabla_{\theta_s}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{\xi}_{t|t-1} \\
& +(\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta'}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{\xi}_{t|t-1} + (\hat{f}_{1t} - \hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\hat{\theta}_s(\delta_2)\nabla_{\delta_{2l}}\hat{\theta}(\delta_2) \\
& +\Big\{\hat{\xi}_{t|t-1}\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{f}_{1t} + (1-\hat{\xi}_{t|t-1})\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{f}_{2t} + (\nabla_{\delta_{2j}}\hat{f}_{1t} - \nabla_{\delta_{2j}}\hat{f}_{2t})\nabla_{\theta'}\hat{\xi}_{t|t-1} \\
& +(\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta'}\hat{f}_{2t})\nabla_{\delta_{2j}}\hat{\xi}_{t|t-1} + (\hat{f}_{1t} - \hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}(\delta_2).
\end{aligned}$$

The first summation in (B.38) consists of the following: $T^{-3/4}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_u}}\tilde{f}_{at}/\tilde{f}_t)\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_v}}\tilde{\xi}_{t|t-1}$ with $u+v\leq 3$. They are $O_p(T^{-1/4})$ by the first result in Lemma A.2. Combining this result with Lemma A.3, it follows that this summation is $O_p(T^{-1/4})$. The remaining two summations in (B.38) have the same structure. They are both $O_p(T^{-1/4})$ after applying (B.35).

Consider the fourth order derivatives. We have

$$
\begin{aligned}
& T^{-1}\mathcal{L}^{(4)}_{jklm}(p,q,\delta_2) - \left(\frac{1-\xi_*}{\xi_*}\right)\mathcal{M}^{(4)}_{(n_\beta+j)klm}(p,q,\delta_2) \qquad \text{(B.39)}\\
= \; & T^{-1}\sum_{t=1}^{T}\left\{\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{L}_{jt}}{\hat{B}_t} - \left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\}\\
& -T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2m}}\hat{B}_t}{\hat{B}_t}\left\{\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{L}_{jt}}{\hat{B}_t} - \left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\}\\
& -T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{B}_t}{\hat{B}_t}\left\{\frac{\nabla_{\delta_{2k}}\hat{L}_{jt}}{\hat{B}_t} - \left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2k}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\}\\
& -T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t}\left\{\frac{\nabla_{\delta_{2m}}\nabla_{\delta_{2k}}\hat{L}_{jt}}{\hat{B}_t} - \left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2m}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\}\\
& +2T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2m}}\hat{B}_t}{\hat{B}_t}\left\{\frac{\nabla_{\delta_{2k}}\hat{L}_{jt}\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t^2} - \left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2k}}\hat{M}_{(n_\beta+j)t}\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t^2}\right\}\\
& -T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t}\left\{\frac{\nabla_{\delta_{2m}}\hat{L}_{jt}}{\hat{B}_t} - \left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2m}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\}\\
& -T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2m}}\hat{B}_t}{\hat{B}_t}\left\{\frac{\nabla_{\delta_{2l}}\hat{L}_{jt}}{\hat{B}_t} - \left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2l}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\}\\
& -T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\hat{B}_t}{\hat{B}_t}\left\{\frac{\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{L}_{jt}}{\hat{B}_t} - \left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\}\\
& +2T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\hat{B}_t}{\hat{B}_t}\left\{\frac{\nabla_{\delta_{2l}}\hat{L}_{jt}\nabla_{\delta_{2m}}\hat{B}_t}{\hat{B}_t^2} - \left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2m}}\hat{B}_t\nabla_{\delta_{2l}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t^2}\right\}\\
& -T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{B}_t}{\hat{B}_t^2}\left\{\hat{L}_{jt} - \left(\frac{1-\xi_*}{\xi_*}\right)\hat{M}_{(n_\beta+j)t}\right\},
\end{aligned}
$$

where

$$
\begin{aligned}
&\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{L}_{jt}\\
&=\sum_{u=1}^{n_\beta+2n_\delta}\sum_{s=1}^{n_\beta+2n_\delta}\{\hat{\xi}_{t|t-1}\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{1t}+(1-\hat{\xi}_{t|t-1})\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{2t}\\
&\quad+(\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&\quad+(\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_u}\hat{\xi}_{t|t-1}+(\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{f}_{1t}-\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{f}_{2t})\nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&\quad+(\nabla_{\delta_{2j}}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\delta_{2j}}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_s}\hat{\xi}_{t|t-1}+(\nabla_{\delta_{2j}}\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\delta_{2j}}\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&\quad+(\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta_u}\hat{\xi}_{t|t-1}+(\nabla_{\delta_{2j}}\hat{f}_{1t}-\nabla_{\delta_{2j}}\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&\quad+(\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\delta_{2j}}\hat{\xi}_{t|t-1}+(\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&\quad+(\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_u}\hat{\xi}_{t|t-1}+(\nabla_{\theta_s}\hat{f}_{1t}-\nabla_{\theta_s}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&\quad+(\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{\xi}_{t|t-1}+(\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&\quad+(\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta'}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta_u}\hat{\xi}_{t|t-1}+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1}\}\times\\
&\nabla_{\delta_{2k}}\hat{\theta}_s(\delta_2)\nabla_{\delta_{2l}}\hat{\theta}_u(\delta_2)\nabla_{\delta_{2m}}\hat{\theta}(\delta_2)\\
&\quad+\sum_{u=1}^{n_\beta+2n_\delta}\sum_{s=1}^{n_\beta+2n_\delta}\{\hat{\xi}_{t|t-1}\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{1t}+(1-\hat{\xi}_{t|t-1})\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{2t}\\
&\quad+(\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{f}_{1t}-\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{f}_{2t})\nabla_{\theta_u}\hat{\xi}_{t|t-1}\\
&\quad+(\nabla_{\delta_{2j}}\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\delta_{2j}}\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\theta_s}\hat{\xi}_{t|t-1}+(\nabla_{\delta_{2j}}\hat{f}_{1t}-\nabla_{\delta_{2j}}\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta_u}\hat{\xi}_{t|t-1}\\
&\quad+(\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\delta_{2j}}\hat{\xi}_{t|t-1}+(\nabla_{\theta_s}\hat{f}_{1t}-\nabla_{\theta_s}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_u}\hat{\xi}_{t|t-1}\\
&\quad+(\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{\xi}_{t|t-1}+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta_u}\hat{\xi}_{t|t-1}\}\times\\
&[\nabla_{\delta_{2m}}\nabla_{\delta_{2k}}\hat{\theta}_s(\delta_2)\nabla_{\delta_{2l}}\hat{\theta}_u(\delta_2)+\nabla_{\delta_{2k}}\hat{\theta}_s(\delta_2)\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\theta}_u(\delta_2)]\\
&\quad+\sum_{s=1}^{n_\beta+2n_\delta}\{\hat{\xi}_{t|t-1}\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t}+(1-\hat{\xi}_{t|t-1})\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t}+(\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{f}_{1t}-\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{f}_{2t})\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&\quad+(\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_s}\hat{\xi}_{t|t-1}+(\nabla_{\delta_{2j}}\hat{f}_{1t}-\nabla_{\delta_{2j}}\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&\quad+(\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\delta_{2j}}\hat{\xi}_{t|t-1}+(\nabla_{\theta_s}\hat{f}_{1t}-\nabla_{\theta_s}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&\quad+(\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta'}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{\xi}_{t|t-1}+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1}\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\delta_2)\nabla_{\delta_{2m}}\hat{\theta}(\delta_2)\\
&\quad+\sum_{s=1}^{n_\beta+2n_\delta}\{\hat{\xi}_{t|t-1}\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{f}_{1t}+(1-\hat{\xi}_{t|t-1})\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{f}_{2t}+(\nabla_{\delta_{2j}}\hat{f}_{1t}-\nabla_{\delta_{2j}}\hat{f}_{2t})\nabla_{\theta_s}\hat{\xi}_{t|t-1}\\
&\quad+(\nabla_{\theta_s}\hat{f}_{1t}-\nabla_{\theta_s}\hat{f}_{2t})\nabla_{\delta_{2j}}\hat{\xi}_{t|t-1}+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{\xi}_{t|t-1}\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\theta}_s(\delta_2).
\end{aligned}
$$

When $\delta_2 = \tilde{\delta}$, all the terms in (B.39) are $o_p(1)$ except the 3rd, 6th and 7th terms. These three terms share the same structure and it is suffice to study the first of them:

$$-T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{B}_t}{\hat{B}_t}\left\{\frac{\nabla_{\delta_{2k}}\hat{L}_{jt}}{\hat{B}_t}-\left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2k}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\}. \tag{B.40}$$

According to the previous study, for $\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{B}_t$ it is sufficient to consider

$$\begin{aligned}
&\left(\frac{1-\xi_*}{\xi_*}\right)^2\left[\xi_*\frac{\nabla_{\delta_{1l}}\nabla_{\delta_{1m}}\tilde{f}_{1t}}{\tilde{f}_t}+(1-\xi_*)\frac{\nabla_{\delta_{1l}}\nabla_{\delta_{1m}}\tilde{f}_{2t}}{\tilde{f}_t}\right]+\left[\xi_*\frac{\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\tilde{f}_{1t}}{\tilde{f}_t}+(1-\xi_*)\frac{\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\tilde{f}_{2t}}{\tilde{f}_t}\right]\\
&-\left(\frac{1-\xi_*}{\xi_*}\right)\left[\xi_*\frac{\nabla_{\delta_{2l}}\nabla_{\delta_{1m}}\tilde{f}_{1t}}{\tilde{f}_t}+(1-\xi_*)\frac{\nabla_{\delta_{2l}}\nabla_{\delta_{1m}}\tilde{f}_{2t}}{\tilde{f}_t}+\xi_*\frac{\nabla_{\delta_{1l}}\nabla_{\delta_{2m}}\tilde{f}_{1t}}{\tilde{f}_t}+(1-\xi_*)\frac{\nabla_{\delta_{1l}}\nabla_{\delta_{2m}}\tilde{f}_{2t}}{\tilde{f}_t}\right]\\
&+\frac{1}{\xi_*^2}\left[\nabla_{\delta_{1l}}\tilde{\xi}_{t|t-1}\left(\frac{\nabla_{\theta_m}\bar{f}_{1t}}{\bar{f}_t}-\frac{\nabla_{\theta_m}\bar{f}_{2t}}{\bar{f}_t}\right)+\left(\frac{\nabla_{\theta_l}\bar{f}_{1t}}{\bar{f}_t}-\frac{\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t}\right)\nabla_{\delta_{1m}}\tilde{\xi}_{t|t-1}\right]\\
&+\sum_{s=1}^{n_\beta+n_\delta}\left\{[\xi_*\nabla_{\theta_s}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_s}\tilde{f}_{2t}]\right\}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\theta}_s(\tilde{\delta}),
\end{aligned}$$

and for

$$\frac{\nabla_{\delta_{2k}}\hat{L}_{jt}}{\hat{B}_t}-\left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2k}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t},$$

it is sufficient to consider

$$\begin{aligned}
&\left(\frac{1-\xi_*}{\xi_*}\right)^2\left[\xi_*\frac{\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\tilde{f}_{1t}}{\tilde{f}_t}+(1-\xi_*)\frac{\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\tilde{f}_{2t}}{\tilde{f}_t}\right]+\left[\xi_*\frac{\nabla_{\delta_{2j}}\nabla_{\delta_{2k}}\tilde{f}_{1t}}{\tilde{f}_t}+(1-\xi_*)\frac{\nabla_{\delta_{2j}}\nabla_{\delta_{2k}}\tilde{f}_{2t}}{\tilde{f}_t}\right]\\
&-\left(\frac{1-\xi_*}{\xi_*}\right)\left[\xi_*\frac{\nabla_{\delta_{2j}}\nabla_{\delta_{1k}}\tilde{f}_{1t}}{\tilde{f}_t}+(1-\xi_*)\frac{\nabla_{\delta_{2j}}\nabla_{\delta_{1k}}\tilde{f}_{2t}}{\tilde{f}_t}+\xi_*\frac{\nabla_{\delta_{1j}}\nabla_{\delta_{2k}}\tilde{f}_{1t}}{\tilde{f}_t}+(1-\xi_*)\frac{\nabla_{\delta_{1j}}\nabla_{\delta_{2k}}\tilde{f}_{2t}}{\tilde{f}_t}\right]\\
&+\frac{1}{\xi_*^2}\left[\nabla_{\delta_{1j}}\tilde{\xi}_{t|t-1}\left(\frac{\nabla_{\theta_k}\bar{f}_{1t}}{\bar{f}_t}-\frac{\nabla_{\theta_k}\bar{f}_{2t}}{\bar{f}_t}\right)+\left(\frac{\nabla_{\theta_j}\bar{f}_{1t}}{\bar{f}_t}-\frac{\nabla_{\theta_j}\bar{f}_{2t}}{\bar{f}_t}\right)\nabla_{\delta_{1k}}\tilde{\xi}_{t|t-1}\right].
\end{aligned}$$

So, at $\delta_2 = \tilde{\delta}$, (B.40) equals

$$\begin{aligned}
&-T^{-1}\sum_{t=1}^{T}\left[\tilde{U}_{lm,t}+\frac{1}{\tilde{f}_t}\sum_{s=1}^{n_\beta+n_\delta}\left\{[\xi_*\nabla_{\theta_s}\tilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_s}\tilde{f}_{2t}]\right\}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\theta}_s(\tilde{\delta})\right]\tilde{U}_{jk,t}+o_p(1)\\
=\;&-T^{-1}\sum_{t=1}^{T}\tilde{U}_{lm,t}\tilde{U}_{jk,t}-T^{-1}\sum_{t=1}^{T}\left[\tilde{U}_{jk,t}\frac{\xi_*\nabla_{(\beta',\delta_1')}\tilde{f}_{1t}+(1-\xi_*)\nabla_{(\beta',\delta_1')}\tilde{f}_{2t}}{\tilde{f}_t}\right]\tilde{I}^{-1}\tilde{D}_{lm}+o_p(1)\\
=\;&-\left[\tilde{V}_{jklm}-\tilde{D}_{jk}'\tilde{I}^{-1}\tilde{D}_{lm}\right]+o_p(1).
\end{aligned}$$

Consequently,

$$
\begin{aligned}
&T^{-1}\mathcal{L}^{(4)}_{jklm}(p,q,\tilde{\delta}) - T^{-1}\left(\frac{1-\xi_*}{\xi_*}\right)\mathcal{M}^{(4)}_{(n_\beta+j)klm}(p,q,\tilde{\delta}) \\
&= -\left\{\tilde{V}_{jklm} - \tilde{D}'_{jk}\tilde{I}^{-1}\tilde{D}_{lm} + \tilde{V}_{jmkl} - \tilde{D}'_{jm}\tilde{I}^{-1}\tilde{D}_{kl} + \tilde{V}_{kmjl} - \tilde{D}'_{km}\tilde{I}^{-1}\tilde{D}_{jl}\right\} + o_p(1).
\end{aligned}
$$

This proves the last result of the lemma. □

The next lemma is the same as Lemma A.3 in Qu and Zhuo (2015) and will be used in the proof of Lemma 5 for establishing the stochastic equicontinuity. Let "*" signify that the quantity is evaluated at the true parameter value.

Lemma A.4. *Let Assumptions 1-6 and the null hypothesis hold. Let* $z_t(\rho) = T^{-1/2}\sum_{s=1}^{t-1}\rho^{t-s}\varepsilon_{js}\varepsilon_{it}$, *where* $\varepsilon_{it} = \nabla_{\delta_{1i}} f^*_{1t}/f^*_t$ *and* $\varepsilon_{js} = \nabla_{\delta_{1j}} f^*_{1s}/f^*_s$. *Then, for any* ρ, ρ_1 *and* ρ_2 *satisfying* $\epsilon - 1 \le \rho_1 \le \rho \le \rho_2 \le 1-\epsilon$, *we have*

$$
E\left(\left|\sum_{t=1}^{T}[z_t(\rho) - z_t(\rho_1)]\right|^2 \left|\sum_{t=1}^{T}[z_t(\rho_2) - z_t(\rho)]\right|^2\right) \le C\left(\rho-\rho_1\right)^2, \tag{B.41}
$$

where C *is a finite constant that depends only on* $0<\epsilon<1/2$ *and the moments of* ε_{it} *and* ε_{js} *up to the fourth order.*

Proof of A.4. See the proof of Lemma A.3 in Qu and Zhuo (2015).

Proof of Lemma 5. Apply the mean value theorem:

$$
T^{-1/2}\sum_{t=1}^{T}\tilde{U}_{jk,t} = T^{-1/2}\sum_{t=1}^{T}U_{jk,t} + \left\{T^{-1}\sum_{t=1}^{T}\nabla_{\theta'}\bar{U}_{jk,t}\right\}T^{1/2}(\tilde{\theta}-\theta_*), \tag{B.42}
$$

where $U_{jk,t}$ and $\bar{U}_{jk,t}$ have the same definition as $\tilde{U}_{jk,t}$ but evaluated at the true value θ_* and some value $\bar{\theta}$ that lies between $\tilde{\theta}$ and θ_*, respectively. We establish the weak convergence of the first term of (B.42) in two steps. First, for any $\epsilon \le p,\ q \le 1-\epsilon$, $T^{-1/2}\sum_{t=1}^{T}U_{jk,t}$ satisfies the central limit theorem. Second, to verify its stochastic equicontinuity, it suffices to consider the following term in its definition (4.6):

$$
\begin{aligned}
&T^{-1/2}\frac{1}{\xi_*^2}\sum_{t=1}^{T}\nabla_{\delta_{1j}}\xi_{t|t-1}\left(\frac{\nabla_{\delta_{1k}}f_{1t}}{f_t} - \frac{\nabla_{\delta_{1k}}f_{2t}}{f_t}\right) \\
&= \left(\frac{1-p}{1-q}\right)\left\{T^{-1/2}\sum_{t=1}^{T}\left(\sum_{s=1}^{t-1}\rho^s\left(\frac{\nabla_{\delta_{1j}}f_{1t-s}}{f_t} - \frac{\nabla_{\delta_{1j}}f_{2t-s}}{f_t}\right)\right)\left(\frac{\nabla_{\delta_{1k}}f_{1t}}{f_t} - \frac{\nabla_{\delta_{1k}}f_{2t}}{f_t}\right)\right\},
\end{aligned}
$$

where the quantities are all evaluated at the true value θ_*, and the equality follows from (B.6) and (2.4). Denote the quantity inside the curly brackets as $W(\rho)$. Note that we have $|\rho| \leq 1-2\epsilon$. Then, Lemma A.3 implies, for any $\rho_1 \leq \rho \leq \rho_2$, we have $E[|W(\rho_1)-W(\rho)|^2\,|W(\rho)-W(\rho_2)|^2] \leq C_2\,(\rho_1-\rho_2)^2$, where C_2 is a finite constant. We can apply the similar proof to the third and fourth components in $T^{-1/2}\sum_{t=1}^{T} U_{jk,t}$. Then the condition required in Theorem 13.5 in Billingsley (1999; c.f. the Display (13.14) in p. 143) is satisfied.

The second term in (B.42) equals, by the mean value theorem,

$$
\begin{aligned}
&-\left\{T^{-1}\sum_{t=1}^{T}\left(\frac{\xi_*\nabla_{(\beta',\delta_1')}f_{1t}+(1-\xi_*)\nabla_{(\beta',\delta_1')}f_{2t}}{f_t}\right)U_{jk,t}\right\}I^{-1}\left\{T^{-1/2}\sum_{t=1}^{T}\left(\frac{\xi_*\nabla_{(\beta',\delta_1')}f_{1t}+(1-\xi_*)\nabla_{(\beta',\delta_1')}f_{2t}}{f_t}\right)\right\}+o_p(1)\\
&=-D_{jk}I^{-1}\left\{T^{-1/2}\sum_{t=1}^{T}\left(\frac{\xi_*\nabla_{(\beta',\delta_1')}f_{1t}+(1-\xi_*)\nabla_{(\beta',\delta_1')}f_{2t}}{f_t}\right)\right\}+o_p(1),
\end{aligned}
$$

where the quantities are all evaluated at the true value θ_* and the equality holds because of the uniform law of large numbers. The term inside the last curly brackets is independent of p and q and satisfies the central limit theorem. Combining the above results for the two terms in (B.42), it follows that $T^{-1/2}\sum_{t=1}^{T}\tilde{U}_{jk,t}$ converges weakly over $\epsilon \leq p,\, q \leq 1-\epsilon$. The covariance function follows immediately. □

Proof of Proposition 1. Let $\eta = T^{-1/4}(\delta_2-\tilde{\delta})$. The expansion (4.5) can be equivalently represented in matrix notation as

$$
\begin{aligned}
&\mathcal{L}(p,q,\delta_2)-\mathcal{L}(p,q,\tilde{\delta})\\
&=\frac{1}{2!}\left(\eta^{\otimes 2}\right)'\left[T^{-1/2}vec\mathcal{L}^{(2)}(p,q,\tilde{\delta})\right]+\frac{1}{3!}\left(\eta^{\otimes 3}\right)'\times O_p\left(T^{-1/4}\right)-\frac{1}{8}\left(\eta^{\otimes 2}\right)'\left[\Omega(p,q)+o_p(1)\right]\left(\eta^{\otimes 2}\right).
\end{aligned}
$$

Because $\Omega(p,q)$ is positive definite, the right hand side will be negative with probability approaching 1 unless $\eta = O_p(1)$. Thus, for any $\varepsilon > 0$, we can choose $M < \infty$ such that $P(\|\eta\| \leq M) \geq 1-\varepsilon$ for sufficiently large T. Restricting to this set, we have

$$
\begin{aligned}
&\sup_{(p,q)\in\Lambda_\epsilon}\sup_{\|\eta\|\leq M}\left[\mathcal{L}(p,q,\delta_2)-\mathcal{L}(p,q,\tilde{\delta})\right]\\
&=\sup_{(p,q)\in\Lambda_\epsilon}\sup_{\|\eta\|\leq M}\left\{\left(\eta^{\otimes 2}\right)'\left[T^{-1/2}vec\mathcal{L}^{(2)}(p,q,\tilde{\delta})\right]-\frac{1}{4}\left(\eta^{\otimes 2}\right)'\Omega(p,q)\left(\eta^{\otimes 2}\right)\right\}+o_p(1) \qquad \text{(B.43)}\\
&\Longrightarrow \sup_{(p,q)\in\Lambda_\epsilon}\sup_{\|\eta\|\leq M}\left\{\left(\eta^{\otimes 2}\right)'G(p,q)-\frac{1}{4}\left(\eta^{\otimes 2}\right)'\Omega(p,q)\left(\eta^{\otimes 2}\right)\right\},
\end{aligned}
$$

where the convergence follows from Lemma (B.43) and that the supremum operator is continuous when taken over a compact set. Finally, the result follows because ε can be made arbitrarily small. $\square$