Estimating Finite-Horizon Life-Cycle Models: A Quasi-Bayesian Approach*

Xiaobin Liu
Singapore Management University

November 16, 2017

Abstract

This paper proposes a quasi-Bayesian approach for structural parameters in finite-horizon life-cycle models. This approach circumvents the numerical evaluation of the gradient of the objective function and alleviates the local optimum problem. The asymptotic normality of the estimators with and without approximation errors is derived. The proposed estimators reach the semiparametric efficiency bound in the general methods of moment (GMM) framework. Both the estimators and the corresponding asymptotic covariance are readily computable. The estimation procedure is easy to parallel so that the graphic processing unit (GPU) can be used to enhance the computational speed. The estimation procedure is illustrated using a variant of the model in Gourinchas and Parker (2002).

JEL classification: C11, C12, D91.
Keywords: Finite-horizon life-cycle model, Structural estimation, Quasi-Bayesian estimator, Method of simulated moment, Numerical solution, GPU computation.

1 Introduction

Life-cycle models (also known as dynamic structural models) have been used extensively in macroeconomics, labor economics, industrial organizations, demographics, household finance, and many other fields; see Pakes (1994) and Rust (1994) for excellent reviews. The life-cycle model with finite-horizon is a subclass that has been found to have a great number of applications. For a sample of references, see Gourinchas and Parker (2002), Jørgensen (2017), Cagetti (2003), Browning and Ejrnæs (2009), Kaplan and Violante (2014), Li et al. (2016), Fagereng, Gottlieb and Guiso (2017), Koijen, Nijman and Irker (2009), and Fischer and Stamos (2013).

A popular technique used to estimate finite-horizon life-cycle models in the literature is based on the log-linearized approximations to Euler equations. However, it has been argued that this approach can result in estimation bias; see Ludvigson and Paxson (2001), Carroll

*Xiaobin Liu, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903. Email: xb.liu.2013@phdecons.smu.edu.sg. I am in debt to my supervisor Jun Yu for his guidance and encouragement. I also thank Thomas Sargent, Peter Phillips, Han Hong, Yonghui Zhang, Tao Zeng, Yong Li, Yichong Zhang, Jungho Lee, Liang Jiang and participants in Professor Jun Yu's econometric lunch workshop, Sargent's Brownbag seminar, Ph.D. workshop in Econometrics in SMU, YEAP conference, China Meeting of Econometric Society, Third Guangzhou Econometrics Workshop and seminar audiences at Singapore Management University, Renmin University and Sun Yat-sen University. All errors are my own.

(2001) and Jøgensen (2016). To deal with this bias, empirical researchers have increasingly adopted the method of simulated moments (MSM) introduced by Duffie and Singleton (1993). Gourinchas and Parker (2002), hereafter GP, were the first to using MSM to estimate the preference parameters in a life-cycle model. Li et al. (2016) studied optimal life-cycle housing and nonhousing consumption using MSM. Fagereng, Gottlieb and Guiso (2017) applied MSM to estimate structural parameters and studied portfolio choice over the life-cycle. In these papers, the estimation procedure was divided into two stages. During the first stage, GMM or calibration was used to estimate parameters of exogenous processes such as the income process. During the second stage the structural parameters were estimated using MSM.

However, since MSM uses iterative optimization algorithms, there are four challenges to its use for estimating finite-horizon life-cycle models. First, the model has to be solved numerically at each iteration. Solving finite-horizon life-cycle models is time consuming and inconvenient because of the nonstationary policy functions. Second, one has to use numerical differentiation to evaluate the gradient of the objective function for parameter updating. Numerical differentiation requires more restrictive assumptions on the objective function and the computation is also cumbersome. Third, due to the complexity of the models, there may exist local optimums. Fourth, typically two-step estimation is necessary, which complicates the asymptotic behavior of the estimator.

The present paper develops a quasi-Bayesian method for estimating structural parameters in finite-horizon life-cycle models during the second stage. Following Chernozhukov and Hong (2003), hereafter CH, we build the quasi-posterior density function based on first-stage estimates and the GMM objective function. The new estimator is obtained by minimizing the Bayesian risk function consisting of the quasi-posterior density function and a net loss function. By doing this, the optimization problem is converted into a sampling one, which avoids the numerical evaluation for the gradient of the objective function and alleviates the local optimum problem; see CH for examples where the local optimum problem was carefully explained.

The asymptotic behavior of the proposed estimator is studied in two cases. First, when the policy functions are analytically available, the asymptotic normality of this estimator is derived. There is a bias in the asymptotic mean that depends on the net loss function. We also show that the estimator reaches the efficiency bound in the framework of GMM. When the net loss function is symmetric, the bias term becomes zero. In particular, if the net loss function is quadratic, the estimator becomes the posterior mean and the associated asymptotic covariance can be approximated by the posterior covariance. This is advantageous in computation since the posterior mean and posterior covariance can be simultaneously computed from the quasi-posterior samples.

Second, when the policy functions are not analytically available, we propose to approximate them over a set of grid points. We show that the magnitude of approximation errors depends both on the number of grid points (j) and the number of observations (N). While the approximation errors associated with a numerical method accumulate as the number of observations grows, it is found that they decrease as the number of grid points (j) increases. Interestingly, the results obtained for the case with analytical solutions still hold true in this case when the approximation errors decrease at a speed faster than the number of observations. This result shows that, even in the presence of approximation errors, the estimation approach is attractive from both the theoretical and computational viewpoints. In practice, most finite-horizon life-cycle models require numerical solutions, making the proposed estimation method useful in practical applications.

In terms of the computational effort, the new estimate requires extensive sampling. It should be noted that Markov Chain Monte Carlo (MCMC) does not work well here. This is because, to use MCMC, such as the Gibbs-sampler and Metropolis-Hasting sampler, one needs to update samples sequentially many times and at each updating the objective function has to be numerically evaluated. Instead of using MCMC, the importance sampling strategy is employed. The algorithm used by Creel and Kristensen (2016) is extended to construct a proposal distribution for important sampling. There are two computational advantages in the proposed algorithm. First, it is easy to parallelize and hence GPU can be used. Second, it is made to be adaptive to the dataset.

This paper makes four contributions to the literature. First, a quasi-Bayesian estimation approach is proposed for finite-horizon life-cycle models. The quasi-Bayesian estimator has desirable properties both in terms of asymptotic behavior and computation. Second, the method extends the seminal work of CH to life-cycle models and is related to a growing strand of literature on approximate Bayesian computation. Third, the econometric problem in the presence of approximation errors caused by numerical methods is carefully studied. The results complement Fernández-Villaverde, Rubio-Ramírez and Santos (2006), hereafter FRS, and Ackerberg, Geweke and Hahn (2009). The present paper considers the problem in the GMM framework while FRS and Ackerberg, Geweke and Hahn (2009) consider the problem in the likelihood setting. If an empirical researcher would like to be agnostic about the error distribution, a GMM framework will be more attractive than the full likelihood approach. Finally, the proposed adaptive algorithm makes use of GPU to enhance computational efficiency and is applicable to other complicated models with moment conditions.

Throughout the paper, a version of the model in GP is used for illustration, but other types of life-cycle models can also be considered. As long as the assumptions listed in the paper are satisfied, the theoretical results can be applied and the estimation algorithm remains useful.

The rest of the paper proceeds as follows. Section 2 introduces the illustrative model in detail. Section 3 presents the first-stage estimation for parameters of the exogenous process and the latent dynamic state variable filtering. Section 4 examines the second-stage estimation, including the definition of the estimator, the asymptotic behavior and the related algorithm to compute the estimator. Section 5 reports results from Monte Carlo studies, including models with and without dynamic latent state. Section 6 concludes. Appendices contain the details of proofs, numerical method used and other related computations.

2 An Illustrative Model

Let us first define a discrete-time life-cycle model for households. Households work until an exogenously given retirement age, T_r. At each working age, the utility function is the constant relative risk aversion (CRRA) utility function, i.e.,

$$u(C;\rho_0) = \begin{cases} \frac{C^{1-\rho_0}}{1-\rho_0} & \rho_0 \neq 1 \\ \log C & \rho_0 = 1 \end{cases},$$

where C is the consumption level and ρ_0 is the risk aversion. The number of household is N^{obs}. By forward looking from the initial working age t_i, household i ($\in \{1, ..., N^{obs}\}$) chooses the level of consumption $C_{i,t}$ to solve the optimization problem

$$\max_{C_{i,\tau}} E_{t_i}\left[\sum_{\tau=t_i}^{T_r}\beta_0^{\tau-t_i}v\left(\boldsymbol{z}_{i,\tau};\boldsymbol{\eta}_0\right)u\left(C_{i,\tau};\rho_0\right)+\beta_0^{T_r+1-t}\tilde{V}_{T_r+1}\left(M_{i,T_r+1},\boldsymbol{z}_{i,T_r+1};\boldsymbol{\eta}_0,\rho_0,\kappa_0\right)\right] \tag{1}$$

$$s.t.\ M_{i,t+1}=R\left(M_{i,t}-C_{i,t}\right)+Y_{i,t+1}, t_i\le t\le T_r-1, \tag{2}$$

$$M_{i,T_r+1}=R\left(M_{i,T_r}-C_{i,T_r}\right), \tag{3}$$

$$C_{i,t}\in\left(0,M_{i,t}\right], \tag{4}$$

$$M_{i,t_i} \text{ given,}$$

where the subscript τ indicates that the associated variable realizes at age τ and the subscript i indicates that the variable belongs to household i, β_0 the subject discount factor, $C_{i,\tau}$ the consumption level, $M_{i,\tau}$ the liquid wealth, R the gross interest rate, $\boldsymbol{z}_{i,\tau}$ a vector of characteristics and $v\left(\boldsymbol{z};\boldsymbol{\eta}_0\right)$ a shifter in utility, which can be interpreted as a taste shifter in which the individual characteristic information $\boldsymbol{z}$ plays a role. In many applications, $v\left(\boldsymbol{z};\boldsymbol{\eta}_0\right)$ is a specific function that summarizes the impact of the individual characteristics $\boldsymbol{z}$.

The equations (2) and (3) are wealth accumulation equations before and after retirement. As in GP, the income process, $Y_{i,t+1}$, is assumed to follow the following stochastic process.

Income process: Income process is defined as

$$\begin{cases}Y_{i,t}=P_{i,t}\epsilon_{i,t},\\ P_{i,t}=G_tP_{i,t-1}\varsigma_{i,t},\end{cases}\quad t_i\le t\le T_r, \tag{5}$$

where $P_{i,t}$ denotes the latent permanent component of $Y_{i,t}$ and $P_{i,T_r+1}=P_{i,T_r}$ since there is no income at age T_r+1, $\epsilon_{i,t}$ the transitory component, G_t the real gross permanent income growth, $\varsigma_{i,t}$ the permanent income shock. Specifically,

$$\epsilon_{i,t}=\begin{cases}\mu_0, & \text{with probability } p_0,\\ \xi_{i,t}, & \text{with probability } 1-p_0,\end{cases}\quad \text{where } \log\xi_{i,t}\overset{i.i.d.}{\sim}N\left(0,\sigma_{\epsilon 0}^2\right),$$

$$\log\varsigma_{i,t}\overset{i.i.d.}{\sim}N\left(0,\sigma_{\varsigma 0}^2\right),$$

where μ_0 can be zero or some other small values, $\log\varsigma_{i,t}$ and $\log\xi_{i,t}$ independent across i and t. The parameters for the income process are denoted as $\boldsymbol{\chi}_0^{inc}=\left(\mu_0,p_0,\sigma_{\epsilon 0}^2,\sigma_{\varsigma 0}^2,\{G_t\}_{t=t_{min}}^{T_r}\right)'$, where $t_{min}=\min_{1\le i\le N^{obs}}\{t_i\}$.

Characteristics information vector: The characteristics vector at age t of household i, $\boldsymbol{z}_{i,t}$, can be deterministic or stochastic. The parameters involved in $\boldsymbol{z}_{i,t}$ are denoted as $\boldsymbol{\chi}_0^{cha}$. According to Jøgensen (2017) and GP, researchers can examine the impact of different characteristics such as the number of children or family size on the marginal utility.

Retirement: When household i retires at T_r, for the tractability of the problem (1), following GP, the retirement value function is assumed to be

$$\tilde{V}_{T_r+1}\left(M_{i,T_r+1},\boldsymbol{z}_{i,T_r+1};\ \boldsymbol{\eta}_0,\rho_0,\kappa_0\right)=\kappa_0v\left(\boldsymbol{z}_{i,T_r+1};\boldsymbol{\eta}_0\right)\frac{\left(M_{i,T_r+1}+H_{i,T_r+1}\right)^{1-\rho_0}}{1-\rho_0},$$

where κ_0 is the motivation to retire, $M_{i,T_{r+1}}$ the liquid wealth at age T_r+1, H_{i,T_r+1} the illiquid wealth after retirement and $H_{i,T_r+1}=hP_{i,T_r+1}$, i.e., H_{i,T_r+1} is proportional to the permanent component at T_r+1. Since there is no income at T_r+1, we let $P_{i,T_r+1}=P_{i,T_r}$.

The Bellman equation for model (1) is

$$\begin{aligned}\tilde{V}_t\left(M_{i,t}, P_{i,t}, \boldsymbol{z}_{i,t}; \boldsymbol{\theta}_0, \boldsymbol{\chi}_0\right) = &\max_{C_{i,t}\in(0,M_{i,t}]} \left\{v\left(\boldsymbol{z}_{i,t}; \boldsymbol{\eta}_0\right) u\left(C_{i,t}; \rho_0\right)\right.\\ &\left.+\beta_0 E_t\left[\tilde{V}_{t+1}\left(M_{i,t+1}, P_{i,t+1}, \boldsymbol{z}_{i,t+1}; \boldsymbol{\theta}_0, \boldsymbol{\chi}_0\right)\right]\right\} \qquad (6)\\ s.t.\ &M_{i,t+1} = R\left(M_{i,t} - C_{i,t}\right) + Y_{i,t+1}, t_i \leq t \leq T_r - 1,\\ &M_{i,T_r+1} = R\left(M_{i,T_r} - C_{i,T_r}\right),\\ &C_{i,t} \in (0, M_{i,t}] \text{ with } M_{i,t_i} \text{ given,}\end{aligned}$$

where $\boldsymbol{\chi}_0 = \left(\left(\boldsymbol{\chi}_0^{inc}\right)', \left(\boldsymbol{\chi}_0^{cha}\right)', R\right)'$, $\boldsymbol{\theta}_0 = (\boldsymbol{\eta}_0', \rho_0, \beta_0, \kappa_0, h)' \in \boldsymbol{\Theta} \subset R^d$. At age $T_r + 1$,

$$\tilde{V}_{T_r+1}\left(M_{i,T_r+1}, P_{i,T_r+1}, \boldsymbol{z}_{i,T_r+1}; \boldsymbol{\theta}_0, \boldsymbol{\chi}_0\right) = \tilde{V}_{T_r+1}\left(M_{i,T_r+1}, \boldsymbol{z}_{i,T_r+1}; \boldsymbol{\eta}_0, \rho_0, \kappa_0, h\right).$$

According to the model setup, the data that economists obtain are $\{M_{i,t}, C_{i,t}, Y_{i,t}, \boldsymbol{z}_{i,t}\}_{t=t_i}^{T_r+1}$ for household i. Therefore, for the Bellman equation (6), economists cannot directly solve it since it involves latent state variable $P_{i,t}$, which is only observed by household i. Thus, we instead study the ratio form of the Bellman equation (6).

The setup of the problem, combined with the retirement value function, makes the problem homogeneous of degree $1 - \rho_0$ in $P_{i,t}$. Thus, we define the normalized value functions as follows.

$$V_t\left(m_{i,t}, \boldsymbol{z}_{i,t}; \boldsymbol{\theta}_0, \boldsymbol{\chi}_0\right) = \frac{1}{P_{i,t}^{1-\rho_0}} \tilde{V}_t\left(M_{i,t}, P_{i,t}; \boldsymbol{\theta}_0, \boldsymbol{\chi}_0\right),$$

$$\begin{aligned}V_{T_r+1}\left(m_{i,T_r+1}, \boldsymbol{z}_{i,T_r+1}; \boldsymbol{\theta}_0, \boldsymbol{\chi}_0\right) &= \frac{1}{P_{i,T_r+1}^{1-\rho_0}} \tilde{V}_{T_r+1}\left(M_{i,T_r+1}, \boldsymbol{z}_{i,T_r+1}; \boldsymbol{\eta}_0, \rho_0, \kappa_0\right)\\ &= \kappa_0 v\left(\boldsymbol{z}_{i,T_r+1}; \boldsymbol{\eta}_0\right) \frac{\left(m_{i,T_r+1} + h\right)^{1-\rho_0}}{1 - \rho_0}.\end{aligned}$$

We also normalize the variables of household i at age t by $P_{i,t}$, denoted by lowercase letters, e.g., $m_{i,t} \equiv M_{i,t}/P_{i,t}$, $c_{i,t} \equiv C_{i,t}/P_{i,t}$. Accordingly, the wealth accumulation equations can be expressed as

$$\begin{aligned}m_{i,t+1} &= \left(m_{i,t} - c_{i,t}\right) \frac{R}{G_{t+1}\varsigma_{i,t+1}} + \epsilon_{i,t+1}, t_i \leq t \leq T_r - 1,\\ m_{i,T_r+1} &= R\left(m_{i,T_r} - c_{i,T_r}\right).\end{aligned}$$

The ratio-form Bellman equation (6) is

$$\begin{aligned}V_t\left(m_{i,t}, \boldsymbol{z}_{i,t}; \boldsymbol{\theta}_0, \boldsymbol{\chi}_0\right) = &\max_{c_{i,t}} \left\{v\left(\boldsymbol{z}_{i,t}; \boldsymbol{\eta}_0\right) u\left(c_{i,\tau}; \rho_0\right)\right.\\ &\left.+\beta_0 E_t\left[\left(G_{t+1}\varsigma_{i,t+1}\right)^{1-\rho_0} V_{t+1}\left(m_{i,t+1}, \boldsymbol{z}_{i,t+1}; \boldsymbol{\theta}_0, \boldsymbol{\chi}_0\right)\right]\right\} \qquad (7)\\ s.t.\ &m_{i,t+1} = \left(m_{i,t} - c_{i,t}\right) \frac{R}{G_{i,t+1}\varsigma_{i,t+1}} + \epsilon_{i,t+1}, t_i \leq t \leq T_r - 1,\\ &m_{i,T_r+1} = R\left(m_{i,T_r} - c_{i,T_r}\right),\\ &c_{i,t} \in (0, m_{i,t}].\end{aligned}$$

Therefore, economists can solve the model (7) without the knowledge of latent state variable $P_{i,t}$.

Remark 2.1. *In the Bellman equation (7), the structural parameter $\boldsymbol{\theta}_0$ is the same as that in the original problem (6). We can solve the model by deriving the analytical solutions or using numerical methods conditional on the value of $\boldsymbol{\theta}_0$ and $\boldsymbol{\chi}_0$. The Euler equations for problem (7) are*

$$c_{i,t}^{-\rho_0} = \beta_0 R E_{\varsigma_{i,t+1},\epsilon_{i,t+1},\boldsymbol{z}_{i,t+1}} \left[\frac{v\left(\boldsymbol{z}_{i,t+1};\boldsymbol{\eta}_0\right)}{v\left(\boldsymbol{z}_{i,t};\boldsymbol{\eta}_0\right)} \left(G_{t+1}\varsigma_{i,t+1}\right)^{-\rho_0} c_{i,t+1}^{-\rho_0} \right], t_i \leq t \leq T_r - 1,$$

which are necessary to derive the optimal policies at each age by backward optimization. In particular, the endogenous grid method (EGM) described in detail in Appendix B.2 can be applied here.

3 First-Stage Estimation and Latent State Filtering

Following GP and based on the discussion in the previous section, the parameters are divided into to two parts, the nuisance parameters $\boldsymbol{\chi}_0 = \left(\left(\boldsymbol{\chi}_0^{inc}\right)', \left(\boldsymbol{\chi}_0^{cha}\right)', R\right)'$ and structural parameters $\boldsymbol{\theta}_0$. Data include a panel dataset used during the second stage estimation, $\left\{C_{i,t}^d, M_{i,t}^d, Y_{i,t}^d, \boldsymbol{z}_{i,t}^d,\right\}_{t=t_i}^{T_r}, i = 1, \ldots, N^{obs}$ and an additional one with sample size J used during the first stage. In the panel dataset with N^{obs} households, $C_{i,t}^d$, $M_{i,t}^d$, $Y_{i,t}^d$ and $\boldsymbol{z}_{i,t}^d$ are respectively the consumption level, liquid wealth, income level and characteristic information vector of household i at age t, respectively.

At the first stage, conditional on the additional dataset, GMM or calibration is used to estimate $\boldsymbol{\chi}$, denoted as $\widehat{\boldsymbol{\chi}}$. The following assumption is imposed for the first-stage estimator.

ASSUMPTION 1. *In the first-stage estimation, the nuisance parameters $\boldsymbol{\chi}_0 = \left(\left(\boldsymbol{\chi}_0^{inc}\right)', \left(\boldsymbol{\chi}_0^{cha}\right)', R\right)' \in \boldsymbol{\Psi}$ can be obtained by GMM based on the additional dataset. The estimator $\widehat{\boldsymbol{\chi}}$ satisfies,*

$$\sqrt{J}\left(\widehat{\boldsymbol{\chi}} - \boldsymbol{\chi}_0\right) \xrightarrow{d} N\left(\mathbf{0}, \Sigma_\chi\right), \tag{8}$$

where Σ_χ is the covariance matrix.

Remark 3.1. *If the calibration approach is used in the first stage, then we simply treat $\widehat{\boldsymbol{\chi}} = \boldsymbol{\chi}_0$ without considering the dispersion caused by estimation, i.e., $\Sigma_\chi = \mathbf{0}$. This approach is frequently used in empirical literature such as Li et al. (2016) and Jøgensen (2017).*

Define $F_{i,t}$ as the information set up to age t for household i. The income process (5) can be rewritten as

$$\begin{cases} \log Y_{i,t} = \log P_{i,t} + \log \epsilon_{i,t}, \\ \log P_{i,t} = \log \widehat{G}_t + \log P_{i,t-1} + \log \varsigma_{i,t}, \end{cases} \quad t_i \leq t \leq T_r - 1,$$

where $\log \epsilon_{i,t} \overset{i.i.d.}{\sim} N\left(0, \hat{\sigma}_\epsilon^2\right)$ and $\log \varsigma_{i,t} \overset{i.i.d.}{\sim} N\left(0, \hat{\sigma}_\varsigma^2\right)$. This is the standard linear state-space model with Gaussian errors so that the Kalman filter can be used to obtain the distribution of $P_{i,t}$ conditional on $F_{i,t}$ and $\widehat{\boldsymbol{\chi}}$. When $\mu = 0$, the observations with zero income level can be considered as missing variables since the estimate $\hat{p}$ for p_0 is very small and thus zero-valued observation is rare. If $\mu \neq 0$ and is very small, then we can set up the threshold value to check whether there exists a shock. Via the Kalman filter, the mean and variance of $P_{i,t}$ conditional on $F_{i,t}$ are obtained. Denote the expectation of a random variable with respect to $P_{i,t}$ up to the information at age t as $E_{P_{i,t}}\left(\cdot|F_{i,t}\right)$.

4 Second-stage Estimation

4.1 Estimator

In this section, given $\widehat{\boldsymbol{\chi}}$ from the frist stage, the estimator for $\boldsymbol{\theta}_0$ will be constructed. In this subsection we deal with the case in which there exists a close-form solution for optimal policy at each age. In the next subsection we deal with the case where optimal policies are not analytically available.

Given any generic $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\boldsymbol{\chi} \in \boldsymbol{\Psi}$, the analytical solutions for the optimal policy functions for the Bellman equation (7) is assumed to exist and denoted as $c_t\left(m_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \boldsymbol{\chi}\right)$ for household i at age t, where $m_{i,t}^d \equiv M_{i,t}^d / P_{i,t}$. For economists, $P_{i,t}$ is unobservable. Hence, taking $P_{i,t}$ into account, conditional on the information up to age t, it is natural to assume that the household i chooses the optimal consumption level according to

$$C_t\left(M_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \boldsymbol{\chi}\right) = E_{P_{i,t}}\left[c_t\left(\frac{M_{i,t}^d}{P_{i,t}}, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \boldsymbol{\chi}\right) P_{i,t} \middle| F_{i,t}\right], \tag{9}$$

where $E_{P_{i,t}}\left(\cdot | F_{i,t}\right)$ is the expectation with respect to $P_{i,t}$ based on the filtering at the first-stage estimation.

Remark 4.1. *The conditional expectation of equation (9) is more natural than the unconditional expectation used in GP, in which the Monte Carlo method was used based on the paths simulated from the initial working age and hence the information up to age t was discarded. Jøgensen (2017) treated the mean of $\log P_{i,t}$ obtained by the Kalman filter as the true value of $\log P_{i,t}$, which also ignored the variance information of $\log P_{i,t}$. In Appendix B.5, these two approaches are compared with that based on equation (9). The evidence shows that equation (9) is superior to the other two approaches.*

In the following assumption, a moment condition is introduced.

ASSUMPTION 2. ***(Identification)*** *The unique parameter $\boldsymbol{\theta}_0$ is in the interior of a compact convex subset $\boldsymbol{\Theta}$ of the Euclidean space R^d. For household i, assume*

$$\begin{aligned} E\left[C_{i,t}^d - C_t\left(M_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}_0, \boldsymbol{\chi}_0\right)\right] =& E\left[g_t\left(M_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}_0, \boldsymbol{\chi}_0\right)\right] \\ =& E\left[g_{i,t}\left(\boldsymbol{\theta}_0; \boldsymbol{\chi}_0\right)\right] = 0, \end{aligned} \tag{10}$$

where $t = t_i, ..., T_r$, $C_{i,t}^d$ is the observed consumption level and $C_t\left(M_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}_0, \boldsymbol{\chi}_0\right)$ is defined in equation (9).

Remark 4.2. *Assumption 2 is the identification assumption of the structural parameters $\boldsymbol{\theta}_0$. The assumption ensures the parameters are point-identified, which is also adopted by Hansen (1982) and Duffie and Singleton (1993).*

According to equation (10), we can have at most T_m moment conditions, where $T_m = T_r - t_{min} + 1$ and $t_{min} = \min\{t_i\}_{i=1}^{N^{obs}}$. Based on $\widehat{\boldsymbol{\chi}}$ from the first stage, the objective function is

$$L_N\left(\boldsymbol{\theta}\right) = L_N\left(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}\right) = -\frac{N}{2}\left[\lambda_N \bar{g}_N\left(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}\right)\right]' W_N\left(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}\right) \lambda_N \bar{g}_N\left(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}\right), \tag{11}$$

where the *total number of observations* $N=\sum_{t=t_{min}}^{T_r}N_t$ with N_t the sample size at age t from $t=t_{min}$ to $t=T_r$,

$$\begin{aligned}\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) &= \left(\bar{g}_{t_{min}}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}),\ldots,\bar{g}_{T_r}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\right)' \\ &= \left(\frac{1}{N_{t_{min}}}\sum_{i=1}^{N_{t_{min}}}g_{i,t_{min}}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}),\ldots,\frac{1}{N_{T_r}}\sum_{i=1}^{N_{T_r}}g_{i,T_r}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\right)',\end{aligned}$$

$$W_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})=V_N^{-1}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}),$$

where,

$$\begin{aligned}V_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) =& \zeta_N\sum_{i=1}^{N^{obs}}\lambda_N\tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\,\tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\,\lambda_N'\zeta_N' \\ &+\frac{N}{J}\lambda_N\bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\,\widehat{\Sigma}_\chi\bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\,\lambda_N',\end{aligned} \tag{12}$$

in which $\bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})$ is the first-order derivative of $\bar{g}_N(\boldsymbol{\theta};\boldsymbol{\chi})$ with respect to $\boldsymbol{\chi}$,

$$\tilde{g}_i(\boldsymbol{\theta};\boldsymbol{\chi})=\underbrace{\left(0,\ldots,0,g_{i,t_i}(\boldsymbol{\theta};\boldsymbol{\chi}),\ldots,g_{i,T_r}(\boldsymbol{\theta};\boldsymbol{\chi})\right)'}_{T_m\text{ elements}},$$

$$\lambda_N=diag\left(\sqrt{\frac{N_{t_{min}}}{N}},\ldots,\sqrt{\frac{N_{T_r}}{N}}\right)=diag\left(\sqrt{\lambda_{N,t_{min}}},\ldots,\sqrt{\lambda_{N,T_r}}\right),$$

$$\zeta_N=diag\left(\sqrt{\frac{1}{N_{t_{min}}}},\ldots,\sqrt{\frac{1}{N_{T_r}}}\right).$$

The use of the weighting matrices λ_N and ζ_N is because households may have different initial working ages.

Following CH, the quasi-Bayesian estimators (QBE), also called Laplace type estimators (LTE), is constructed. Although the objective function in (11) is not a probability density function, it is transformed into a proper one by

$$p_N(\boldsymbol{\theta})=\frac{e^{L_N(\boldsymbol{\theta})}\pi(\boldsymbol{\theta})}{\int_\Theta e^{L_N(\boldsymbol{\theta})}\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}}, \tag{13}$$

where $\pi(\boldsymbol{\theta})$ is the prior information. The $p_N(\boldsymbol{\theta})$ in equation (13) is called the *quasi-posterior* density function. Based on $p_N(\boldsymbol{\theta})$, given the penalty or loss function $\varrho_N(u)$, the corresponding risk function is

$$R_N(\boldsymbol{\xi})=\int_\Theta\varrho_N(\boldsymbol{\theta}-\boldsymbol{\xi})\,p_N(\boldsymbol{\theta})\,d\boldsymbol{\theta}. \tag{14}$$

Following CH, the following assumptions are imposed on the loss function $\varrho_N(u)$.

ASSUMPTION 3. *The loss function $\varrho_N:R^d\to R_+$ satisfies:*

(i) $\varrho_N(u)=\varrho\left(\sqrt{N}u\right)$, *where $\varrho(u)\geq 0$ and $\varrho(u)=0$ if and only if $u=0$;*

(ii) *ϱ is convex and $\varrho(h) \leq 1 + |h|^p$ for some $p \geq 1$;*

(iii) *$\varphi(\xi) = \int_{R^d} \varrho(u - \xi) e^{-u'au} du$ is minimized uniquely at some $\tau \in R^d$ for any finite $a > 0$.*

Given the loss function $\varrho_N(u)$, based on risk function (14), the QBE for $\boldsymbol{\theta}_0$ is defined below.

Definition 4.1. *The QBE is the one minimizing the risk function $R_N(\boldsymbol{\xi})$ in (14):*

$$\widehat{\boldsymbol{\theta}} = \arg \inf_{\boldsymbol{\xi} \in \boldsymbol{\Theta}} R_N(\boldsymbol{\xi}). \tag{15}$$

4.2 Asymptotic Theory for analytical Solution for Optimal Policy

In this subsection, the asymptotic behavior of the estimator $\widehat{\boldsymbol{\theta}}$ defined in (15) is studied. The following assumptions are imposed.

ASSUMPTION 4. *The function $g_t\left(M_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \boldsymbol{\chi}\right)$ defined in (10) satisfies the following conditions: (i) $g_t(\cdot; \boldsymbol{\theta}, \boldsymbol{\chi})$ and $\nabla_\theta g_t(\cdot; \boldsymbol{\theta}, \boldsymbol{\chi})$ are Borel measurable for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\boldsymbol{\chi} \in \boldsymbol{\Psi}$; (ii) given $\boldsymbol{\chi} \in \boldsymbol{\Psi}$, $\nabla_\theta g_t\left(M_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \boldsymbol{\chi}\right)$ is continuously differentiable on $\boldsymbol{\Theta}$; (iii) $\nabla_{\theta\theta} g_t(\cdot; \boldsymbol{\theta}, \boldsymbol{\chi})$ is Borel measurable for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\boldsymbol{\chi} \in \boldsymbol{\Psi}$.*

ASSUMPTION 5. *$G(\boldsymbol{\theta}, \boldsymbol{\chi}) = \nabla_\theta E\left[g_t\left(M_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \boldsymbol{\chi}\right)\right]$ is continuous on $\boldsymbol{\Theta}$ and $\boldsymbol{\chi}$. $G(\boldsymbol{\theta}_0, \boldsymbol{\chi}_0)$ is finite and has full rank.*

ASSUMPTION 6. *$\lim_{N\to\infty} \lambda_N = \lambda$, $\lim_{N\to\infty} N/J = \gamma$ for some constants $\lambda, \gamma \in R^+$,*

Remark 4.3. *Assumptions 4 and 5 are similar to those in Hansen (1982). The assumptions on the moment vector are essential for the study of asymptotic behavior of the estimator. Assumption 6 implies N_t is proportional to the total number of observations N. Assumption 6 also implicates that N is proportional to the number of households in the dataset, N^{obs}.*

When GMM is adopted during the first stage, the following two assumptions are imposed.

ASSUMPTION 7. *The first-order derivative of $g_t\left(M_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \boldsymbol{\chi}\right)$ with respect to $\boldsymbol{\chi}$, $g_{t,\chi}\left(M_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \boldsymbol{\chi}\right)$ satisfies the following conditions: (i) $g_{t,\chi}(\cdot; \boldsymbol{\theta}, \boldsymbol{\chi})$ and $\nabla_\theta g_{t,\chi}(\cdot; \boldsymbol{\theta}, \boldsymbol{\chi})$ are Borel measurable for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\boldsymbol{\chi} \in \boldsymbol{\Psi}$; (ii) given $\boldsymbol{\chi} \in \boldsymbol{\Psi}$, $\nabla_\theta g_{t,\chi}\left(M_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \boldsymbol{\chi}\right)$ is continuously differentiable on $\boldsymbol{\Theta}$; (iii) $\nabla_{\theta\theta} g_{t,\chi}(\cdot; \boldsymbol{\theta}, \boldsymbol{\chi})$ is Borel measurable for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\boldsymbol{\chi} \in \boldsymbol{\Psi}$.*

ASSUMPTION 8. *$G_\chi(\boldsymbol{\theta}, \boldsymbol{\chi}) = \nabla_\chi E\left[g_t\left(M_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \boldsymbol{\chi}\right)\right]$ is continuous on $\boldsymbol{\Theta}$ and $\boldsymbol{\chi}$. $G_\chi(\boldsymbol{\theta}_0, \boldsymbol{\chi}_0)$ is finite and full rank.*

Remark 4.4. *Assumptions 7 and 8 are similar to Assumptions 5 and 6. They are associated with $g_{t,\chi}\left(M_{i,t}^d,\ \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta},\ \boldsymbol{\chi}\right)$ and necessary because the estimation error due to GMM must be taken into account. These two assumptions are not required if the calibration is used during the first stage.*

Finally, there are also some restrictions on the prior information $\pi(\boldsymbol{\theta})$.

ASSUMPTION 9. *$\pi(\boldsymbol{\theta})$ is continuous and uniformly positive over $\boldsymbol{\Theta}$*

In this paper, only GMM is used during the first stage because the calibration is a special case of GMM as explained in Remark 3.1. Based on the discussion above, we define

$$g_i(\boldsymbol{\theta};\boldsymbol{\chi}) = \underbrace{\left(g_{i,t_{min}}(\boldsymbol{\theta};\boldsymbol{\chi}),\ldots,g_{i,T_r}(\boldsymbol{\theta};\boldsymbol{\chi})\right)'}_{T_m \text{ elements}}.$$

Furthermore, according to the standard assumption that households are independent across i, we have the following lemma and theorems.

Lemma 4.1. *Under Assumptions 5–8, $V_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})$ defined in equation (12) has the following property, uniformly over $\boldsymbol{\Theta}$,*

$$\begin{aligned} V_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) \xrightarrow{p} & \lambda E\left[g_i(\boldsymbol{\theta};\boldsymbol{\chi}_0)\, g_i(\boldsymbol{\theta};\boldsymbol{\chi}_0)'\right]\lambda' \\ & + \gamma\lambda E\left[g_{i,\chi}(\boldsymbol{\theta};\boldsymbol{\chi}_0)\right]\Sigma_\chi E\left[g_{i,\chi}(\boldsymbol{\theta};\boldsymbol{\chi}_0)'\right]\lambda' = V(\boldsymbol{\theta}). \end{aligned}$$

Theorem 4.1. *Under Assumptions 1–9, for the estimator $\widehat{\boldsymbol{\theta}}$ defined in (15),*

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} \tau + \mathbb{N}(\mathbf{0},\Sigma_\theta),$$

where

$$\Sigma_\theta = \left[G_\theta'\lambda'\left(\lambda\Sigma_g\lambda' + \gamma\lambda G_\chi'\Sigma_\chi G_\chi\lambda'\right)^{-1}\lambda G_\theta\right]^{-1},$$

$$\tau = \arg\inf_{\alpha\in R^d}\left\{\int_{R^d}\rho(\alpha - u)\, f\left(u;\mathbf{0},G_\theta'\lambda' W(\boldsymbol{\theta}_0)\lambda G_\theta\right)du\right\},$$

where $f(\cdot,\boldsymbol{\mu},\boldsymbol{\Omega})$ is the multivariate normal density with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Omega}$, $G_\theta = \nabla_\theta E\left[g_i(\boldsymbol{\theta}_0;\boldsymbol{\chi}_0)\right]$, $G_\chi = \nabla_\chi E\left[g_i(\boldsymbol{\theta}_0;\boldsymbol{\chi}_0)\right]$, $\Sigma_g = E\left[g_i(\boldsymbol{\theta}_0;\boldsymbol{\chi}_0)\, g_i(\boldsymbol{\theta}_0;\boldsymbol{\chi}_0)'\right]$.

Remark 4.5. *If the calibration is used during the first stage, then we have*

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} \tau + \mathbb{N}\left(\mathbf{0},\left(G_\theta'\lambda'\left(\lambda\Sigma_g\lambda'\right)^{-1}\lambda G_\theta\right)^{-1}\right).$$

Since there is no need to take estimation error into account, the second term in the optimal weighting matrix disappears in the calibration.

Usually τ is difficult to evaluate at $\boldsymbol{\theta}_0$ since the value of $\boldsymbol{\theta}_0$ is unknown. However, if we choose the quadratic loss function, according to CH and the Bayesian literature, the estimator in Definition 4.1 becomes the mean of the quasi-posterior distribution in (13), which is called the *quasi-posterior mean* and defined as

$$\bar{\boldsymbol{\theta}} = E_{p_N}[\boldsymbol{\theta}] = \int_{\boldsymbol{\Theta}}\boldsymbol{\theta}\, p_N(\boldsymbol{\theta})\, d\boldsymbol{\theta}. \tag{16}$$

The corollary below follows Theorem 4.1.

Corollary 4.2. *Under Assumptions 1–9, given $\varrho_N(\cdot) = N\cdot u^2$ and the estimator $\bar{\boldsymbol{\theta}}$ defined in (16),*

$$\sqrt{N}\left(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathbb{N}(\mathbf{0},\Sigma_\theta),$$

with $\Sigma_\theta = \left[G_\theta'\lambda'\left(\lambda\Sigma_g\lambda' + \gamma\lambda G_\chi'\Sigma_\chi G_\chi\lambda'\right)^{-1}\lambda G_\theta\right]^{-1}$, where the variables are the same as in Theorem 4.1. Meanwhile, Σ_θ has the following property.

$$N\cdot E_{p_N}\left[\left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}\right)'\right] = \Sigma_\theta + o_p(1).$$

Remark 4.6. *From Corollary 4.2 with samples from $p_N(\boldsymbol{\theta})$, both the estimator and the asymptotic covariance, which are the mean and covariance of quasi-posterior distribution, can be simultaneously calculated. This is in contrast to extremum estimators where the estimator and the asymptotic covariance are obtained separately.*

4.3 Asymptotic Theory for Numerical Solution for Optimal Policy

In most cases, there is no analytical solution for the Bellman equation (7). Numerical methods are needed to solve the model inevitably introducing approximation errors. In this subsection, we develop conditions under which the results obtained in the last subsection continue to hold when numerical solutions are used.

Given the values of $\boldsymbol{\theta}$ and $\boldsymbol{\chi}$, the (infeasible) exact solution for the policy function at age t for household i is denoted as $c_t\left(m_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \boldsymbol{\chi}\right)$. Denote the numerical approximation by $c_t^j\left(m_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \boldsymbol{\chi}\right)$ where j is the number of grid points in the finite range of $m_{i,t}^d$ based on which we can evaluate other optimal policies by using interpolation methods. The numerical solution $c_t^j\left(m_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \boldsymbol{\chi}\right)$ is indexed by j because the approximation admits refinements, i.e., when j goes to infinity, $c_t^j\left(m_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \boldsymbol{\chi}\right)$ converges to $c_t\left(m_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \boldsymbol{\chi}\right)$.

With the numerical solution, neither the exact objective function (11) nor the quasi-posterior density in (13) can be evaluated. Before we introduce our estimation procedure, let us first fix some new notations.

The approximated optimal consumption level for household i at age t is

$$C_t^j\left(M_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \boldsymbol{\chi}\right) = E_{P_{i,t}}\left[c_t^j\left(\frac{M_{i,t}^d}{P_{i,t}}, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \boldsymbol{\chi}\right) P_{i,t} \middle| F_{i,t}\right]. \tag{17}$$

The sample moment becomes

$$\begin{aligned} C_{i,t}^d - C_t^j\left(M_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}_0, \boldsymbol{\chi}_0\right) =& g_t^j\left(M_{i,t}^d,\ \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}_0,\ \boldsymbol{\chi}_0\right) \\ =& g_{i,t}^j\left(\boldsymbol{\theta}_0;\ \boldsymbol{\chi}_0\right), \end{aligned} \tag{18}$$

for household i at age t, where $t = t_i, ..., T_r$. Then the approximate objective function is defined as

$$L_N^j\left(\boldsymbol{\theta}\right) = -\frac{N}{2}\left[\lambda_N \bar{g}_N^j\left(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}\right)\right]' W_N^j\left(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}\right) \lambda_N \bar{g}_N^j\left(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}\right), \tag{19}$$

where

$$\begin{aligned} \bar{g}_N^j\left(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}\right) &= \left(\bar{g}_{t_{min}}^j\left(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}\right), \ldots, \bar{g}_{T_r}^j\left(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}\right)\right)' \\ &= \left(\frac{1}{N_{t_{min}}} \sum_{i=1}^{N_{t_{min}}} g_{i,t_{min}}^j\left(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}\right), \ldots, \frac{1}{N_{T_r+1}} \sum_{i=1}^{N_{T_r+1}} g_{i,T_r}^j\left(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}\right)\right)', \end{aligned}$$

$$W_N^j\left(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}\right) = \left[V_N^j\left(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}\right)\right]^{-1},$$

$$
\begin{aligned}
V_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) = & \zeta_N \sum_{i=1}^{N^{obs}} \lambda_N \tilde{g}_i^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\,\tilde{g}_i^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\,\lambda_N'\zeta_N' \\
& + \frac{N}{J}\lambda_N \bar{g}_{N,\chi}^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\,\widehat{\Sigma}_\chi \bar{g}_{N,\chi}^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\,\lambda_N',
\end{aligned}
\tag{20}
$$

$$
\tilde{g}_i^j(\boldsymbol{\theta};\boldsymbol{\chi}) = \underbrace{\left(0,\ldots,0,g_{i,t_i}^j(\boldsymbol{\theta};\boldsymbol{\chi}),\ldots,g_{i,T_r}^j(\boldsymbol{\theta};\boldsymbol{\chi})\right)'}_{T_m \text{ elements}}.
$$

Remark 4.7. *Based on the approximated objective function (19), one can use MSM to obtain the extremum estimator. If so, one must implement an iterative optimization algorithm in which the value and gradient of the objective function have to be numerically evaluated for each parameter updating. These computational efforts and their cost are demanding. Further, as pointed out in CH, sometimes the maximum estimator is the local optimum, not the global one.*

Based on equation (19), we can define the *approximated quasi-posterior* as

$$
p_N^j(\boldsymbol{\theta}) = \frac{e^{L_N^j(\boldsymbol{\theta})}\pi(\boldsymbol{\theta})}{\int_\Theta e^{L_N^j(\boldsymbol{\theta})}\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}}. \tag{21}
$$

Given the loss function $\varrho_N(u)$, the risk function and estimator corresponding to the approximated quasi-posterior is

$$
R_N^j(\boldsymbol{\xi}) = \int_\Theta \varrho_N(\boldsymbol{\theta}-\boldsymbol{\xi})\,p_N^j(\boldsymbol{\theta})\,d\boldsymbol{\theta}, \tag{22}
$$

$$
\widehat{\boldsymbol{\theta}}^j = \arg\inf_{\boldsymbol{\xi}\in\boldsymbol{\Theta}} R_N^j(\boldsymbol{\xi}). \tag{23}
$$

Other variables remain the same as those in the case with the analytical solution.

Following FRS and Ackerberg, Geweke and Hahn (2009), the following assumption is imposed on numerical methods.

ASSUMPTION 10. *For all j, $\boldsymbol{\chi}$ and $\boldsymbol{z}$, over a finite range of m, $c_t^j(m,\boldsymbol{z};\ \boldsymbol{\theta},\boldsymbol{\chi})$ is continuous on m and continuously differentiable at all points except at a finite number of points.*

Remark 4.8. *Assumption 10 ensures the continuity of $c_t^j(m,\boldsymbol{z};\boldsymbol{\theta},\widehat{\boldsymbol{\chi}})$ at all points and differentiability except at a finite number of points in the finite range of m. The lack of differentiability makes it possible to use numerical methods with kinks at a finite number of points. Such methods include the linear interpolation or the approximation within space spanned by linear basis functions. This assumption is satisfied naturally by most solution methods for dynamic economic models.*

FRS studied the econometric problem of computed dynamic models. They found that under some mild conditions, as the approximated policy functions converged to the exact ones, the approximated likelihood also converged to the exact likelihood. Meanwhile, as more data are included, a better approximation is required. Ackerberg, Geweke and Hahn (2009) examined the impact of approximation errors on a classical estimate of a simple time series model. They found the approximation errors are required to vanish at a certain speed as the sample size goes to infinity. Following Ackerberg, Geweke and Hahn (2009), the approximation error is defined as

$$\Delta_j = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta},\boldsymbol{\chi}\in\boldsymbol{\Psi}} \left\{ \max_{\boldsymbol{z},m,t} \left\{ \left\| c_t^j\left(m,\boldsymbol{z};\boldsymbol{\theta},\boldsymbol{\chi}\right) - c_t\left(m,\boldsymbol{z};\boldsymbol{\theta},\boldsymbol{\chi}\right)\right\|, \right.\right.$$
$$\left.\left. \left\| C_{t,\chi}^j\left(M,\boldsymbol{z};\boldsymbol{\theta},\boldsymbol{\chi}\right) - C_{t,\chi}\left(M,\boldsymbol{z};\boldsymbol{\theta},\boldsymbol{\chi}\right)\right\| \right\}\right\}. \tag{24}$$

Remark 4.9. *Unlike Ackerberg, Geweke and Hahn (2009), we do not need to consider the approximation error associated with the first and second-order derivatives of the objective function. Note that* $t \in [t_{min}, T_r + 1]$ *and from the dataset, the normalized wealth* m *and characteristic vector* $\boldsymbol{z}$ *are all bounded. Thus, given any generic* $\boldsymbol{\theta}$ *and* $\boldsymbol{\chi}$*,* Δ_j *is controlled by the number of grid points* j*. Furthermore, if the calibration is adopted during the first stage, we do not have to consider the approximation error of* $C_{t,\chi}^j\left(m,\boldsymbol{z};\boldsymbol{\theta},\boldsymbol{\chi}\right)$*.*

In accordance with Ackerberg, Geweke and Hahn (2009), the approximation error should disappear asymptotically, i.e., $j \to \infty$, as $N \to \infty$. Given Assumptions 1–10, the following theorem hold.

Theorem 4.3. *Under Assumptions 1–10, for the estimator* $\widehat{\boldsymbol{\theta}}^j$ *defined in (23), if as* $N \to \infty$*,*

$$N\Delta_j \to 0,$$

then,

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}^j - \boldsymbol{\theta}_0\right) \xrightarrow{d} \tau + \mathbb{N}\left(\mathbf{0}, \Sigma_\theta\right),$$

with

$$\Sigma_\theta = \left[G_\theta'\lambda'\left(\lambda\Sigma_g\lambda' + \gamma\lambda G_\chi'\Sigma_\chi G_\chi\lambda'\right)^{-1}\lambda G_\theta\right]^{-1}.$$

Remark 4.10. *An approximate optimal policy for every household at every age inevitably introduces the approximation error. As the total number of observations increases, the error will accumulate. Theorem 4.3 requires that the accumulative approximation error be smaller than the sampling error, and thus is negligible. The detailed relationship between* j *and* N *in different numerical methods is left for future studies.*

Similarly, given the quadratic loss function, the approximated quasi-posterior mean is defined as

$$\bar{\boldsymbol{\theta}}^j = E_{p_N^j}\left[\boldsymbol{\theta}\right] = \int_{\boldsymbol{\Theta}} \boldsymbol{\theta} p_N^j\left(\boldsymbol{\theta}\right) d\boldsymbol{\theta}. \tag{25}$$

Corollary 4.4. *Under Assumptions 1–10, given the quadratic loss function* $\varrho_N(\cdot)$ *and the estimator* $\bar{\boldsymbol{\theta}}^j$ *defined in (25), if* $N\Delta_j \to 0$ *as* $N \to \infty$*, then,*

$$\sqrt{N}\left(\bar{\boldsymbol{\theta}}^j - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathbb{N}\left(\mathbf{0}, \Sigma_\theta\right),$$

with $\Sigma_\theta = \left[G_\theta'\lambda'\left(\lambda\Sigma_g\lambda' + \gamma\lambda G_\chi'\Sigma_\chi G_\chi\lambda'\right)^{-1}\lambda G_\theta\right]^{-1}$*, where the variables are the same as in Theorem 4.1. Meanwhile,* Σ_θ *has the following property.*

$$N \cdot E_{p_N^j}\left[N\left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}^j\right)\left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}^j\right)'\right] = \Sigma_\theta + o_p(1), \tag{26}$$

where $E_{p_N^j}$ *is the expectation with respect to* $p_N^j(\boldsymbol{\theta})$*.*

Theorem 4.3 and Corollary 4.4 are important because they show that when the approximation errors disappears at a speed faster than the total number of observations, the approximated estimator shares the desirable properties of the estimator when policy functions are analytically available.

This result is related to that in FRS and Ackerberg, Geweke and Hahn (2009) with two differences. First, both papers focus on the likelihood inference, whereas the estimation framework is GMM in the present paper. Second, the disappearance rate in Theorem 4.3 is also different. In Ackerberg, Geweke and Hahn (2009), a static simple time series model is studied and the rate of the approximation errors is required to be faster than the square root of the time span, i.e., $o\left(T^{1/2}\right)$. The present paper focuses on the life-cycle model with finite horizon and the speed of the approximation error is required to be faster than the total number of the observations, i.e., $o\left(N\right)$.

Remark 4.11. *Theorem 4.3 and Corollary 4.4 show that only the approximation error of* $c_t^j\left(m_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \widehat{\boldsymbol{\chi}}\right)$ *and* $C_{t,\chi}^j\left(M_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \widehat{\boldsymbol{\chi}}\right)$ *need to be considered. If the calibration is used at the first stage, the approximation error of* $C_{t,\chi}^j\left(M_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \widehat{\boldsymbol{\chi}}\right)$ *can be ignored. However, if an optimization approach is used, other types of approximation errors, such as those in calculating the first- and second-order derivatives of the objective function, require careful attention, which may be very complicated and difficult to control in practice.*

Remark 4.12. *Equation (26) can be used to compute the asymptotic covariance. On the one hand, it is the by-product of samples from the approximated quasi-posterior distribution. On the other hand, it avoids numerical evaluations of* G_θ *and* G_χ*.*

4.4 Estimation

The theoretical results in previous subsections are attractive. However, sampling from the quasi-posterior distribution remains a difficult problem. The MCMC method does not work well here since it requires sampling sequentially many times and numerically evaluating the objective function at each updating. Instead of MCMC, importance sampling is used together with GPU to enhance the computational speed.

In practice, it is very hard to find a good proposal distribution for the importance sampling. Direct sampling from the prior can be computationally inefficient. Recognizing this problem, we adapt the algorithm proposed in Creel and Kristensen (2016) to estimate finite-horizon life-cycle models. The algorithm for the estimation is summarized in Algorithms 1 and 2. Both algorithms request a great number of quasi-posterior density evaluations. The usual CPU time will be high. Thanks to the availability of GPU, we can solve the model numerically given a great number of parameter values and do the interpolation in parallel.

In Algorithm 1, δ and $\exp(L)$ are close to zero. They are threshold values for the search of area and selection of particles with significant quasi-posterior density values, respectively. Specifically, steps 10–24 ensure that the shrinking sampling area is sufficiently narrow given K_1 and δ, and that they are adaptive to different datasets. Besides, step 25 selects particles in S with significant quasi-posterior density values, denoted as $\tilde{S}$. Step 26–29 uniformly draw K particles from $\tilde{S}$ and construct the proposal distribution for important sampling, which is a mixture of normal distributions.

Algorithm 1 Construction of Proposal Distribution

Input: The number of samples K_1, the selected number of particles K_2, the covariance for the random perturbation Σ, the tolerance level δ, the threshold value L, the number of component in proposal distribution K.

Set up $i = 0$

for $k = 1$ to K_1 **do**

  Draw $\boldsymbol{\theta}_k^i \sim \pi(\boldsymbol{\theta})$.

  Compute $\omega_k^i = L_N\left(\boldsymbol{\theta}_k^i\right) + \log \pi\left(\boldsymbol{\theta}_k^i\right)$.

end for

Set up the set of particles $S = \emptyset$.

Compute $V_1 = \max \omega_k^0$

Compute $V_2 = V_1 + 2\delta$

while $|V_1 - V_2| < \delta$ **do**

  Sort $\left\{\omega_k^i\right\}_{k=1}^{K_1}$ in descending order.

  Select the first K_2 of the sorted ω_k^i and associated $\boldsymbol{\theta}_k^i$, obtain $\left\{\tilde{\omega}_k^i\right\}_{k=1}^{K_2}$ and $\left\{\tilde{\boldsymbol{\theta}}_k^i\right\}_{k=1}^{K_2}$

  $S = S \bigcup \left\{\tilde{\boldsymbol{\theta}}_k^i\right\}_{k=1}^{K_2}$.

  for $k = 1$ to K_2 **do**

    Compute $\omega_{Norm}^k = \frac{e^{\tilde{\omega}_k^i}}{\sum_{k=1}^{K} e^{\tilde{\omega}_k^i}}$.

  end for

  for $k = 1$ to K_1 **do**

    Draw $\tilde{\tilde{\boldsymbol{\theta}}}_k^i \sim Multinomial\left(\left\{\tilde{\boldsymbol{\theta}}_k^i\right\}_{k=1}^{K_2}, \left\{\omega_{Norm}^k\right\}_{k=1}^{K_2}\right)$

    Compute $\boldsymbol{\theta}_k^{i+1} = \tilde{\tilde{\boldsymbol{\theta}}}_k^i + \epsilon_k^{i+1}, \epsilon_k^{i+1} \sim N(0, \Sigma)$.

    Compute $\omega_k^{i+1} = L_N\left(\boldsymbol{\theta}_k^{i+1}\right) + \log \pi\left(\boldsymbol{\theta}_k^{i+1}\right)$.

  end for

  Compute $V_1 = V_2$.

  Compute $V_2 = \max \omega_k^{i+1}$.

end while

Select the particle points in S that satisfies $\omega_k^i - V_2 > L$, obtain $\tilde{S}$.

for $k = 1$ to K **do**

  Draw $\boldsymbol{\theta}_k^{IS}$ from $\tilde{S}$ uniformly.

end for

Define the importance sampling density as the mixture of densities associated with each drawn $\boldsymbol{\theta}_k^{IS}$:

$$q(\boldsymbol{\theta}) = \sum_{k=1}^{K} p_k q_k\left(\boldsymbol{\theta} | \boldsymbol{\theta}_k^{IS}\right),$$

where $p_k = e^{\omega_k} / \sum_{k=1}^{K} e^{\omega_k}$, $\omega_k == L_N\left(\boldsymbol{\theta}_k^{IS}\right) + \log \pi\left(\boldsymbol{\theta}_k^{IS}\right)$, and $q_k\left(\boldsymbol{\theta} | \boldsymbol{\theta}_k^{IS}\right) = N\left(\boldsymbol{\theta}_k^{IS}, \Sigma\right)$. Or $p_k = \frac{1}{K}$, for $k = 1, \ldots, K$.

Output: $q(\boldsymbol{\theta})$.

Algorithm 2 Estimator Calculation

Input: The number of samples K_3, the proposal distribution $q(\boldsymbol{\theta})$.
for $k = 1$ to K_3 **do**
    Draw $\boldsymbol{\theta}^{(k)} \sim q(\boldsymbol{\theta})$.
    Compute $\tilde{\omega}^{(k)} = e^{L_N\left(\boldsymbol{\theta}^{(k)}\right)} \pi\left(\boldsymbol{\theta}^{(k)}\right)$.
end for
Compute the estimator

$$\widehat{\bar{\boldsymbol{\theta}}} = \frac{\sum_{k=1}^{K_3} \omega^{(k)} \boldsymbol{\theta}^{(k)}}{\sum_{k=1}^{K_3} \omega^{(k)}},$$

$$\widehat{Var(\boldsymbol{\theta})} = \frac{1}{\sum_{k=1}^{K_3} \omega^{(k)}} \sum_{k=1}^{K_3} \omega^{(k)} \left(\boldsymbol{\theta}^{(k)} - \widehat{\bar{\boldsymbol{\theta}}}\right)\left(\boldsymbol{\theta}^{(k)} - \widehat{\bar{\boldsymbol{\theta}}}\right)',$$

where $\omega^{(k)} = \tilde{\omega}^{(k)} / q\left(\boldsymbol{\theta}^{(k)}\right)$.
Output: $\widehat{\bar{\boldsymbol{\theta}}}$, $\widehat{Var(\boldsymbol{\theta})}$.

In Algorithm 2, when $K_3 \to \infty$, $\widehat{\bar{\boldsymbol{\theta}}} \to \bar{\boldsymbol{\theta}}$, $\widehat{Var(\boldsymbol{\theta})} \to Var(\boldsymbol{\theta})$, where $Var(\boldsymbol{\theta})$ is the quasi-posterior covariance with respect to $p_N(\boldsymbol{\theta})$, since

$$\widehat{\bar{\boldsymbol{\theta}}} = \frac{\sum_{k=1}^{K_3} \omega^{(k)} \boldsymbol{\theta}^{(k)}}{\sum_{k=1}^{K_3} \omega^{(k)}} \to \int_{\Theta} \boldsymbol{\theta} p_n(\boldsymbol{\theta})\, d\boldsymbol{\theta} = \bar{\boldsymbol{\theta}},$$

$$\begin{aligned}
\widehat{Var(\boldsymbol{\theta})} &= \frac{1}{\sum_{k=1}^{K_3} \omega^{(k)}} \sum_{k=1}^{K_3} \omega^{(k)} \left(\boldsymbol{\theta}^{(k)} - \widehat{\bar{\boldsymbol{\theta}}}\right)\left(\boldsymbol{\theta}^{(k)} - \widehat{\bar{\boldsymbol{\theta}}}\right)' \\
&\to \int_{\Theta} \boldsymbol{\theta}\boldsymbol{\theta}' p_n(\boldsymbol{\theta})\, d\boldsymbol{\theta} + \bar{\boldsymbol{\theta}}\bar{\boldsymbol{\theta}}' \\
&\equiv \int_{\Theta} \left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}\right)' p_N(\boldsymbol{\theta})\, d\boldsymbol{\theta}.
\end{aligned}$$

Remark 4.13. *The numerical evaluation of the quasi-posterior density values is costly computationally. GPU can enhance the computational speed greatly since it can solve the model and compute the density values in parallel given a great number of sampled parameters. Steps 10–24 are adaptive since the area with the largest posterior density values will be automatically found given the dataset, δ and K_1.*

5 Monte Carlo Studies

In this section, two models are studied to examine the performance of the new approach. One is the life-cycle model without exogenous dynamic latent state. The other one is a simplified version of the illustrative model.

Table 1: The Values of Parameters Used to Simulate Data

T	β	ρ	R	y	σ_ϵ^2
10	0.96	2	1.03	0.5	0.04

5.1 The Case without Dynamic Latent State

The households are faced with the same utility maximization problem, i.e.,

$$\max_{\{c_t\}_{t=0}^T} E_0\left[\sum_{t=0}^T \beta^t \frac{c_t^{1-\rho}}{1-\rho}\right], \tag{27}$$
$$s.t.\ m_{t+1} = R\left(m_t - c_t\right) + y\epsilon_{t+1}, 0 \leq t < T,$$
$$c_t \in \left(0, m_t\right], \text{with } m_0 \text{ given},$$

where β is the subjective discount factor, ρ the risk aversion of the households, R the gross interest rate, y the income level for the households from period $t = 0$ to $t = T$, ϵ_{t+1} the income shock associated with the income at each period and $\epsilon_{t+1} \overset{i.i.d.}{\sim} \log N\left(-\frac{\sigma_\epsilon^2}{2}, \sigma_\epsilon^2\right)$, m_t the liquid wealth at the beginning of period t and c_t the consumption level that chosen by the households, which is in the budget constraint $(0, m_t]$. Thus, the Euler equations for the life-cycle model are

$$c_t^{-\rho} = R\beta E_t\left[c_{t+1}\left(m_{t+1}\right)^{-\rho}\right], m_{t+1} = R\left(m_t - c_t\right) + y\epsilon_{t+1}, 0 \leq t \leq T-1, \tag{28}$$

where at period T, $c_T = m_T$, which results from the households seeking to consume all their wealth at the last period. There are no close-form solutions for the optimal consumptions, thus a numerical method is required. Conditional on the values of parameters, EGM is used to construct the grid of the optimal consumption at each period. The detail is illustrated in Appendix B.2.

In this study, the true values of the parameters are reported in Table 1. Conditional on the values listed in Table 1, we solve the model numerically and simulate a data set $\left\{c_{i,t}^*, m_{i,t}^d\right\}_{t=0}^T$ for each household i, where the initial wealth $m_{0,i}^d$ is drawn from a truncated normal distribution with mean 5 and variance 100 ranging from 0 to infinity, i.e., $N(5, 100)\, I\{x > 0\}$, where I is the indicator function. The optimal consumption $c_{i,t}^*$ is interpolated based on the consumption grid obtained from numerical solving. The measurement error is added, $c_{i,t}^d = c_{i,t}^* + \varepsilon_{i,t}$, $\varepsilon_{i,t} \overset{i.i.d.}{\sim} N\left(0, \sigma_\varepsilon^2\right)$, where $\sigma_\varepsilon^2 = 0.005^\dagger$. The numbers of households simulated are $N^{obs} = 1000, 1500, 2000, 3000$, respectively and the number of replications for each case is 200. For each replication, the simulated noisy data $\left\{c_{i,t}^d, m_{i,t}^d\right\}_{t=1}^T$ are used to estimate the parameters ρ and β.

In order to estimate the parameters, the priors for the two parameters are set to

$$\beta \sim U(0.5, 1), \rho \sim U(0, 15),$$

† Jøgensen (2017) estimated the variance of measurement error, which was approximately 0.46. But the sample size he used ranged from 150,000 to 800,000. Since the sample sizes in Monte Carlo studies are between 1000 and 3000, the variance of measurement error is proportionally set as 0.005 in terms of the variance of sample moments.

Table 2: The Bias and RMSE of the Estimator for β and ρ

	β		ρ	
	Bias	RMSE	Bias	RMSE
$N^{obs} = 1000$	-1.3602×10^{-3}	3.89×10^{-3}	0.2311	0.6780
$N^{obs} = 1500$	-1.4685×10^{-3}	3.407×10^{-3}	0.2535	0.6008
$N^{obs} = 2000$	-6.2943×10^{-4}	2.683×10^{-3}	0.1081	0.4692
$N^{obs} = 3000$	4.3860×10^{-4}	2.2×10^{-3}	0.0715	0.3926

where $U(a, b)$ is the uniform distribution ranging from a to b. For β, based on the economic theory, it should satisfy $\beta \in (0, 1)$ and usually it is assumed to be around 0.9. Thus the prior for β is uninformative. Besides, for the risk averse parameter, ρ, the range between 0 and 15 is also quite uninformative.

Algorithms 1 and 2 are applied to estimate the model (for more details of the estimation, please refer to Appendix B.3) and the bias and root mean square error (RMSE) are computed for each parameter in every scenario. The bias and RMSE are defined in Appendix B.1. The results are listed in Table 2. It is obvious that as the sample size increases, the bias of both parameters decreases. Further, the RMSE of both parameters also decreases and the magnitude of all the RMSE is proportional to the square root of the sample size approximately. This simulation study justifies the asymptotic theory and the usefulness of the algorithm.

5.2 The Case with Dynamic Latent State

In this subsection, a simplified life-cycle model in GP is considered to examine the performance of the new approach. The model is defined in the following. The household i is faced with the following optimization problem,

$$\max_{C_{i,\tau}} E_{t_i} \left[\sum_{\tau=t_0}^{T_r} \beta^{\tau-t_0} \frac{C_{i,\tau}^{1-\rho}}{1-\rho} + \kappa \beta^{T_r+1-t_0} \frac{\left(M_{i,T_r+1} + H_{i,T_{r+1}}\right)^{1-\rho}}{1-\rho} \right] \tag{29}$$

$$\begin{aligned} s.t. M_{i,t+1} &= R\left(M_{i,t} - C_{i,t}\right) + Y_{i,t+1}, t_0 \le t \le T_r - 1 \\ M_{i,T_r+1} &= R\left(M_{i,T_r} - C_{i,T_r}\right), t = T_r, \\ C_{i,t} &\in \left(0, M_{i,t}\right], \text{with } M_{i,t_0} \text{ given.} \end{aligned}$$

The model specification is almost the same as the illustrative model except that all households start to work at the same age and the marginal utility shifter is not included. The income process is also the same and is defined as,

$$\begin{cases} Y_{i,t} = P_{i,t} \epsilon_{i,t}, \\ P_{i,t} = G_t P_{i,t-1} \varsigma_{i,t}, \end{cases} \quad t_i \le t \le T_r,$$

$$\epsilon_{i,t} = \begin{cases} \mu, & \text{with probability } p, \\ \xi_{i,t}, & \text{with probability } 1-p, \end{cases} \quad \text{where } \log \xi_{i,t} \overset{i.i.d.}{\sim} N\left(0, \sigma_\epsilon^2\right),$$

$$\log \varsigma_{i,t} \overset{i.i.d.}{\sim} N\left(0, \sigma_\varsigma^2\right).$$

Table 3: Parameter values used to simulate data.

$\{G_t\}_{t=1}^{10}$	R	σ_ϵ^2	σ_ς^2	p	μ	β	ρ	T_r	γ_1	t_0
Figure 1	1.03	0.04	0.02	0.03	10^{-6}	0.96	2	10	0.07	1

The parameters of the income process are given and the ratio-form Bellman equation is now,

$$V_t(m_{i,t};\boldsymbol{\theta}) = \max_{c_{i,t}} \left\{ \frac{c_{i,\tau}^{1-\rho}}{1-\rho} + +\beta E_t \left[(G_{t+1}N_{i,t+1})^{1-\rho} V_{t+1}(m_{i,t+1};\boldsymbol{\theta}) \right] \right\} \tag{30}$$

$$s.t. m_{i,t+1} = (m_{i,t} - c_{i,t}) \frac{R}{G_{i,t+1}\varsigma_{i,t+1}} + \epsilon_{i,t+1}, t_i \leq t \leq T_r - 1,$$

$$m_{i,T_r+1} = R(m_{i,T_r} - c_{i,T_r}), t = T_r,$$

$$c_{i,t} \in (0, m_{i,t}],$$

with

$$\begin{aligned} V_{T_r+1}(m_{i,T_r+1};\boldsymbol{\theta}) &= \kappa \frac{(m_{i,T_r+1} + h)^{1-\rho}}{1-\rho} \\ &= \frac{1}{(1-\rho)\kappa^{-\frac{1}{\rho}}} \left(\kappa^{-\frac{1}{\rho}} m_{i,T_r+1} + \kappa^{-\frac{1}{\rho}} h \right)^{1-\rho} \\ &= \frac{1}{(1-\rho)\gamma_1} (\gamma_1 m_{i,T_r+1} + \gamma_0)^{1-\rho}, \end{aligned}$$

where $c_{i,t}$ and $m_{i,t}$ are the normalized values of consumption level $C_{i,t}$ and wealth $M_{i,t}$, respectively. For simplicity, γ_0 is equal to 0, which is consistent with the result obtained by GP. The value function after retirement becomes

$$V_{T_r+1}(m_{i,T_r+1};\boldsymbol{\theta}) = \frac{1}{(1-\rho)\gamma_1} (\gamma_1 m_{i,T_r+1})^{1-\rho}. \tag{31}$$

The structural parameter is now $\boldsymbol{\theta} = \{\beta, \rho, \gamma_1\}$. The values of parameters for the simulation are listed in Table 3.

The values of $\{G_t\}_{t=1}^{10}$ are described in the left panel of Figure 1, which is the same as Jøgensen (2016). The discount factor β, gross interest rate R, income shock probability p, variance of transitory shock σ_ϵ^2, retirement rule parameter γ_1 and variance of the shock to permanent income σ_ς^2 are approximately equal to those in GP. Following Jøgensen (2016), the risk aversion ρ equals 2 and the value of μ is very close to zero.

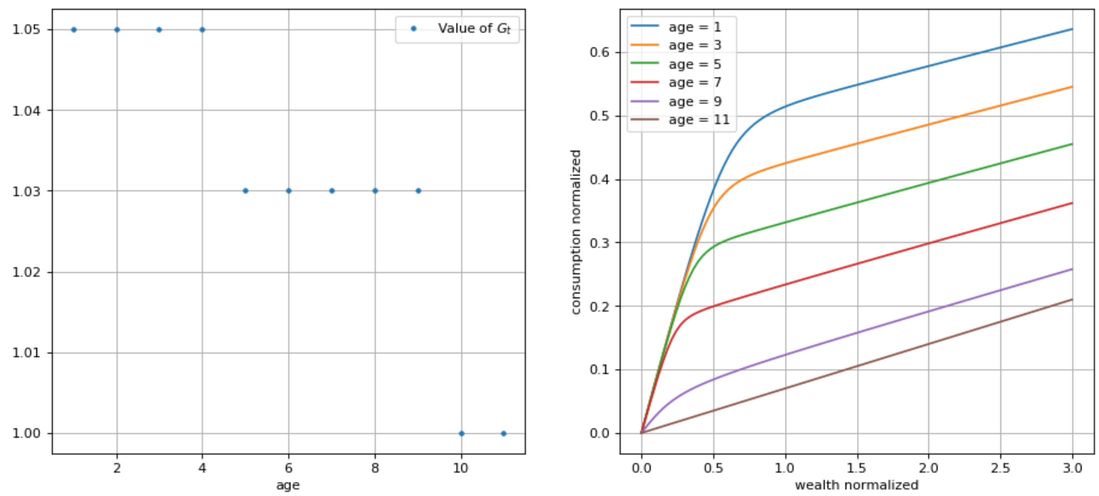
For this model, the corresponding ratio-form Euler equations are

$$c_{i,t}^{-\rho} = \max\left\{ m_{i,t}^{-\rho}, \beta R E_{\varsigma_{i,t+1},\epsilon_{i,t+1}} \left[(G_{t+1}\varsigma_{i,t+1})^{-\rho} c_{i,t+1} (m_{i,t+1})^{-\rho} \right] \right\}, t_0 \leq t \leq T_r - 1,$$

$$c_{i,T_r}^{-\rho} = \max\left\{ m_{i,T_r}^{-\rho}, \beta R (\gamma_1 m_{i,T_r+1})^{-\rho} \right\}, \text{ at age } T_r.$$

EGM is used to solve the model (for more details, one can refer to Appendix B.2). The solution of the ratio-form model is presented in the right panel of Figure 1.

Figure 1: The values of G_t and the policy functions for Bellman equation in ratio form



Notes: The left panel presents the plots of the value of G_t at different ages. The right panel is the numerical solution of the ratio-form model (30).

In the simulation, we assume at age $t = 1$, the corresponding permanent component of income $P_{i,1}^d$ for every household is drawn from a log-normal distribution, i.e.,

$$\log P_{i,1}^d \sim N\left(0, \sigma_\varsigma^2\right), \forall i = 1, ..., N^{obs},$$

where N^{obs} is the number of simulated households. We then simulate an income panel dataset $\left\{Y_{i,t}^d, P_{i,t}^d\right\}_{t=1}^{10}$ for each household i. Meanwhile, household's initial wealth at age 1, $M_{i,1}^d$, is sampled from a truncated normal distribution with mean 1 and variance 1 ranging from 0 to infinity, i.e., $M_{i,1}^d \sim N(1,1) I\{x > 0\}$, for $i = 1, \ldots, N^{obs}$, where I is the indicator function.

The Bellman equation in ratio form is solved by EGM and we obtain the consumption grid at each period. At each t, we normalize the wealth $m_{i,t}^d = \frac{M_{i,t}^d}{P_{i,t}^d}$ and use the grid to interpolate the corresponding optimal ratio-form consumption $c_{i,t}^*$. We then compute the optimal consumption level as $C_{i,t}^* = c_{i,t}^* P_{i,t}^d$ and obtain $\left\{C_{i,t}^*, M_{i,t}^d, Y_{i,t}^d, P_{i,t}^d\right\}_{t=1}^{10}$ for each household i. Following the simulation procedure in the last subsection, we add the measurement error, $C_{i,t}^d = C_{i,t}^* + \varepsilon_{i,t}$, $\varepsilon_{i,t} \overset{i.i.d.}{\sim} N\left(0, \sigma_\varepsilon^2\right)$, $\sigma_\varepsilon^2 = 0.008$. Finally we have $\left\{C_{i,t}^d, M_{i,t}^d, Y_{i,t}^d\right\}_{t=1}^{10}$, for $i = 1, \ldots, N^{obs}$, which is used for estimation.

In order to obtain the sample moment vector, the Kalman filter is used to filter the income observations to obtain the mean and variance for $P_{i,t}$ at each t for household i. The Kalman filter for income process is documented in detail in Appendix B.4.

To estimate the parameters ρ, β, γ_1, the following priors are used,

$$\rho \sim U(0, 15), \beta \sim U(0.5, 1), \gamma_1 \sim U(0, 1).$$

It is quite intuitive that households must use their wealth to support their lives after retirement and they would not consume all their liquid wealth in the first year after retirement. Thus,

Table 4: The bias and RMSE of the estimator

		$N^{obs} = 1500$	$N^{obs} = 2000$	$N^{obs} = 3000$
β	Bias	2.8583×10^{-4}	-3.0266×10^{-5}	-3.7344×10^{-5}
	RMSE	2.8394×10^{-3}	2.7242×10^{-3}	1.9599×10^{-3}
ρ	Bias	-3.4112×10^{-2}	-1.8676×10^{-3}	-1.5472×10^{-2}
	RMSE	0.1726	0.1676	0.1321
γ_1	Bias	5.7411×10^{-5}	2.2704×10^{-5}	-6.8096×10^{-6}
	RMSE	2.0429×10^{-4}	1.6755×10^{-4}	1.3224×10^{-4}

the bound is quite reasonable and uninformative. For the priors for ρ and β, they are also uninformative as argued earlier.

We use Algorithms 1 and 2 to do the estimation. In the estimation, we set $K_1 = K_3 = 38400$, $K_2 = 1280$, $\Sigma = diag\,(0.0001, 0.04, 0.0001)$, $\delta = 0.5$, $L = -10$, $K = 7680$ and the number of grids in EGM is 100. The sample sizes considered here are $N^{obs} = 1500, 2000, 3000$, respectively. The number of replications is 50. The biases and RMSE of the estimation are reported in Table 4.

The results in Table 4 have similar patterns to the outputs in the preceding subsection. The bias for all parameters decreases as the sample size increases. Further, the RMSE is approximately proportional to the square root of sample size as predicted by theory. In summary, the results in Table 4 still justify the asymptotic theory.

6 Conclusion

In this paper, a quasi-Bayesian estimator is introduced for structural parameters in finite-horizon life-cycle models. The asymptotic normality of the estimator is derived when an analytical solution for the model exists. When the policy functions are not analytically available, it is shown that if the approximation errors caused by numerical solving vanish fast enough, the estimator remains to be asymptotically normal. Further, it is shown that the estimator reaches the semiparametric efficiency bound in the GMM framework. In the proposed method, the usual optimization procedure is converted into a sampling procedure, thereby avoiding the numerical evaluation for the gradient of objective function and alleviating the local optimum problem. The estimator and associated asymptotic covariance can be computed simultaneously. The estimation procedure is also easy to parallelize, facilitating a GPU-based and adaptive algorithm to enhance computational efficiency. The estimation procedure is also illustrated based on a variant of the model in GP.

In general our estimator is less efficient than the full likelihood-based procedures, such as those proposed by FRS and Ackerberg, Geweke, and Hahn (2009). However, our procedure is less stringent about the model specification. For example, the distribution is left unspecified in our approach. Hence, our set up may be more appealing to empirical researchers who are agnostic about distributional behaviors of the errors.

There are many possible extensions for this method. For example, finite-horizon life-cycle models with endogenous discrete choices can be considered since these models have received considerable attention recently; see Iskhakov et.al. (2017), Kaplan and Violante (2014) and references therein. Meanwhile, the present paper only focuses on the estimation. There also remains plenty of work related to inference. These topics are left for future research.

References

Ackerberg, Daniel, John Geweke, and Jinyong Hahn, Comments on Convergence Properties of the Likelihood of Computed Dynamic Models, *Econometrica* 77:6 (2009), 2009-2017.

Browning, Martin, and Mette Ejrnæs, Consumption and Children, *The Review of Economics and Statistics* 91:1 (2009), 93-111.

Carroll, Christopher D., Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis, *The Quarterly Journal of Economics* 112:1 (1997), 1-55.

Carroll, Christopher D., Death to the Log-Linearized Consumption Euler Equation! (And Very Poor Health to the Second-Order Approximation), *Advances in Macroeconomics* 1:1 (2001).

Carroll, Christopher D., The Method of Endogenous Grid Points for Solving Dynamic Stochastic Optimization Problems, *Economics Letters* 91:3 (2006), 312-320.

Cagetti, Marco, Wealth Accumulation over the Life Cycle and Precautionary Savings, *Journal of Business & Economic Statistics* 21:3 (2003), 339-353.

Campbell, John Y., Household Finance, *The Journal of Finance*, 61:4 (2006), 1553-1604.

Chernozhukov, Victor, and Han Hong, An MCMC Approach to Classical Estimation, *Journal of Econometrics* 115:2 (2003), 293-346.

Creel, Michael, and Dennis Kristensen, ABC of SV: Limited Information Likelihood Inference in Stochastic Volatility Jump-Diffusion Models, *Journal of Empirical Finance* 31 (2015), 85-108.

Creel, Michael, and Dennis Kristensen, On Selection of Statistics for Approximate Bayesian Computing (or the Method of Simulated Moments), *Computational Statistics & Data Analysis* 100 (2016), 99-114.

Creel, Michael, Jiti Gao, Han Hong, and Dennis Kristensen, Bayesian Indirect Inference and the ABC of GMM, arXiv preprint arXiv:1512.07385(2015).

Duffle, Darrell, and Kenneth J. Singleton, Simulated Moments Estimation of Markov Models of Asset Prices, *Econometrica* 61:4 (1993), 929-952.

Fagereng, Andreas, Charles Gottlieb, and Luigi Guiso, Asset Market Participation and Portfolio Choice over the Life-Cycle, *The Journal of Finance* 72:2 (2017), 705-750.

Fernández-Villaverde, Jesús, Juan F. Rubio-Ramirez, and Manuel S. Santos, Convergence Properties of the Likelihood of Computed Dynamic Models, *Econometrica* 74:1 (2006), 93-119.

Fischer, Marcel, and Michael Z. Stamos, Optimal Life Cycle Portfolio Choice with Housing Market Cycles, *The Review of Financial Studies* 26:9 (2013), 2311-2352.

Gourinchas, Pierre-Olivier and Jonathan A. Parker, Consumption over the Life Cycle, *Econometrica* 70:1 (2002), 47-89.

Hansen, Lars Peter, Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica* (1982) 1029-1054.

Iskhakov, Fedor, Thomas H. Jørgensen, John Rust, and Bertel Schjerning, Estimating Discrete-Continuous Choice Models: The Endogenous Grid Method with Taste Shocks, *Quantitative Economics* 8:2 (2017), 317-365.

Jørgensen, Thomas H., Euler Equation Estimation: Children and Credit Constraints, *Quantitative Economics* 7:3 (2016), 935-968.

Jørgensen, Thomas H., Life-Cycle Consumption and Children: Evidence from a Structural Estimation, *Oxford Bulletin of Economics and Statistics* 79:5 (2017), 717-746.

Kaplan, Greg, and Giovanni L. Violante, A Model of the Consumption Response to Fiscal Stimulus Payments, *Econometrica* 82:4 (2014), 1199-1239.

Koijen, Ralph S., Theo E. Nijman, and Bas J. Werker, When Can Life Cycle Investors Benefit from Time-Varying Bond Risk Premia?, *The Review of Financial Studies* 23:2 (2009), 741-780.

Li, Wenli, Haiyong Liu, Fang Yang, and Rui Yao, Housing over Time and over the Life Cycle: A Structural Estimation, *International Economic Review* 57:4 (2016), 1237-1260.

Ludvigson, Sydney, and Christina H. Paxson, Approximation Bias in Linearized Euler Equations, *Review of Economics and Statistics* 83:2 (2001), 242-256.

Appendices

A Proof of Lemmas and Theorems

A.1 The Proof of Lemma 4.1

As in (12),

$$\begin{aligned} V_N\left(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}\right) = & \zeta_N \sum_{i=1}^{N^{obs}} \lambda_N \tilde{g}_i\left(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}\right) \tilde{g}_i\left(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}\right)' \lambda_N' \zeta_N' \\ & + \frac{N}{J} \lambda_N \bar{g}_{N,\chi}\left(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}\right) \widehat{\Sigma}_\chi \bar{g}_{N,\chi}\left(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}\right)' \lambda_N', \end{aligned}$$

For the first term, by Assumption 2 and Assumption 7, as $N \to \infty$, $\widehat{\boldsymbol{\chi}} \to \boldsymbol{\chi}_0$. And in the framework of the structural model, $\{\tilde{g}_i\left(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}\right)\}_{i=1}^{N^{obs}}$ are independent across i. Combined with Assumption 1, 5 and 6, we have

$$\zeta_N \sum_{i=1}^{N^{obs}} \lambda_N \tilde{g}_i\left(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}\right) \tilde{g}_i\left(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}\right)' \lambda_N \zeta_N \xrightarrow{p} E\left[\lambda g_i\left(\boldsymbol{\theta};\boldsymbol{\chi}_0\right) g_i\left(\boldsymbol{\theta};\boldsymbol{\chi}_0\right)' \lambda'\right],$$

where

$$g_i\left(\boldsymbol{\theta};\boldsymbol{\chi}\right) = \underbrace{\left(g_{i,t_{min}}\left(\boldsymbol{\theta};\boldsymbol{\chi}\right), \ldots, g_{i,T_r}\left(\boldsymbol{\theta};\boldsymbol{\chi}\right)\right)'}_{T_m \text{ elements}}.$$

Similarly, by Assumption 2 and 7, as $N \to \infty$, $\widehat{\Sigma}_\chi \xrightarrow{p} \Sigma_\chi$, $\widehat{\boldsymbol{\chi}} \to \boldsymbol{\chi}_0$. Combined with Assumption 1, 8 and 9, we can have

$$\frac{N}{J} \lambda_N \bar{g}_{N,\chi}\left(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}\right) \widehat{\Sigma}_\chi \bar{g}_{N,\chi}\left(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}\right)' \lambda_N', \xrightarrow{p} \gamma \lambda E\left[g_{i,\chi}\left(\boldsymbol{\theta};\boldsymbol{\chi}_0\right)\right] \Sigma_\chi E\left[g_{i,\chi}\left(\boldsymbol{\theta};\boldsymbol{\chi}_0\right)'\right] \lambda'.$$

A.2 The Proof of Theorem 4.1

We define

$$M\left(\boldsymbol{\theta}\right) = -\frac{1}{2} E\left[g_i\left(\boldsymbol{\theta};\boldsymbol{\chi}_0\right)\right]' \lambda' W\left(\boldsymbol{\theta}\right) \lambda E\left[g_i\left(\boldsymbol{\theta};\boldsymbol{\chi}_0\right)\right],$$

where $W\left(\boldsymbol{\theta}\right) = V^{-1}\left(\boldsymbol{\theta}\right) = \left\{\lambda E\left[g_i\left(\boldsymbol{\theta};\boldsymbol{\chi}_0\right) g_i\left(\boldsymbol{\theta};\boldsymbol{\chi}_0\right)'\right] \lambda + \gamma \lambda E\left[g_{i,\chi}\left(\boldsymbol{\theta};\boldsymbol{\chi}_0\right)\right] \Sigma_\chi E\left[g_{i,\chi}\left(\boldsymbol{\theta};\boldsymbol{\chi}_0\right)'\right] \lambda'\right\}^{-1}$, where $V\left(\boldsymbol{\theta}\right)$ defined in Lemma 4.1. From the definition of criterion function (11), under Assumption 1- 10 , we have

$$\frac{1}{N} L_N\left(\boldsymbol{\theta}\right) = -\frac{1}{2} \bar{g}\left(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}\right)' \lambda_N' V_N^{-1}\left(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}\right) \lambda_N \bar{g}\left(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}\right) \xrightarrow{p} M\left(\boldsymbol{\theta}\right).$$

Further, in the framework, we implies that the matrix $V_N\left(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}\right)$ and $V\left(\boldsymbol{\theta}\right)$ are positive definite for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Thus, the $\boldsymbol{a}$s $W_N\left(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}\right) = V_N^{-1}\left(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}\right)$ and $W\left(\boldsymbol{\theta}\right) = V^{-1}\left(\boldsymbol{\theta}\right)$.(**what is the meaning of the above two sentences?**) Due to $W\left(\boldsymbol{\theta}\right) > 0$ and $M\left(\boldsymbol{\theta}_0\right) = 0$, by Assumption 3, for any $\delta > 0$, $\boldsymbol{\theta} \in \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \geq \delta\} \subset \boldsymbol{\Theta}$, we have $M\left(\boldsymbol{\theta}\right) < 0$, so that $M\left(\boldsymbol{\theta}\right) - M\left(\boldsymbol{\theta}_0\right) < 0$. Therefore, the Lemma 1 in Chernozukov and Hong (2003) is satisfied.

Since $\{g_{i,t}(\boldsymbol{\theta};\boldsymbol{\chi})\}$ are independent across i, we have

$$\sqrt{N}\lambda_N\bar{g}_N(\boldsymbol{\theta}_0;\boldsymbol{\chi}_0)\xrightarrow{d}N\left(\mathbf{0},\lambda\Sigma_g\lambda'\right),$$

where $\Sigma_g=E\left[g_i(\boldsymbol{\theta}_0;\boldsymbol{\chi}_0)\,g_i(\boldsymbol{\theta}_0;\boldsymbol{\chi}_0)'\right]$. If we use the GMM method to estimate the parameter $\boldsymbol{\chi}_0$, for $\sqrt{N}\lambda_N\bar{g}(\boldsymbol{\theta}_0;\widehat{\boldsymbol{\chi}})$, expanding it around $\boldsymbol{\chi}_0$,

$$\begin{aligned}\sqrt{N}\lambda_N\bar{g}_N(\boldsymbol{\theta}_0;\widehat{\boldsymbol{\chi}})=&\sqrt{N}\lambda_N\left[\bar{g}_N(\boldsymbol{\theta}_0;\boldsymbol{\chi}_0)+\bar{g}_\chi(\boldsymbol{\theta}_0;\tilde{\boldsymbol{\chi}})'(\widehat{\boldsymbol{\chi}}-\boldsymbol{\chi}_0)+o_p\left(\frac{1}{\sqrt{J}}\right)\right]\\=&\sqrt{N}\lambda_N\bar{g}_N(\boldsymbol{\theta}_0;\boldsymbol{\chi}_0)+\sqrt{\frac{N}{J}}\lambda_N\bar{g}_{N,\chi}(\boldsymbol{\theta}_0;\tilde{\boldsymbol{\chi}})'\sqrt{J}(\widehat{\boldsymbol{\chi}}-\boldsymbol{\chi}_0)+o_p\left(\sqrt{\frac{N}{J}}\right).\end{aligned}$$

By Assumption 2, from the first-stage estimation,

$$\sqrt{J}(\widehat{\boldsymbol{\chi}}-\boldsymbol{\chi}_0)\xrightarrow{d}N(\mathbf{0},\Sigma_\chi).$$

Following GP, since the first-stage estimator is obtained conditional on exogenous structural models and mostly different data, then we can have

$$\sqrt{N}\lambda_N\bar{g}_N(\boldsymbol{\theta}_0;\widehat{\boldsymbol{\chi}})\xrightarrow{d}N\left(\mathbf{0},\lambda\Sigma_g\lambda'+\gamma\lambda G_\chi'\Sigma_\chi G_\chi\lambda'\right),\tag{A.1}$$

where $G_\chi=E[\nabla_\chi g_i(\boldsymbol{\theta}_0;\boldsymbol{\chi}_0)]$, $\gamma=\lim_{N\to\infty}\frac{N}{J}$, $\lambda=\lim_{N\to\infty}\lambda_N$, $\Sigma_g=E\left[g_i(\boldsymbol{\theta}_0;\boldsymbol{\chi}_0)\,g_i(\boldsymbol{\theta}_0;\boldsymbol{\chi}_0)'\right]$.

We can rewrite the criterion function as

$$\begin{aligned}L_N(\boldsymbol{\theta})=&-\frac{N}{2}[\lambda_N\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})]'W_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\lambda_N\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\\=&-\frac{N}{2}[\lambda_N\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})]'\left[\zeta_N\sum_{i=1}^{N^{obs}}\lambda_N\tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\,\tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N'\zeta_N'\right.\\&\left.+\frac{N}{J}\lambda_N\bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\,\widehat{\Sigma}_\chi\bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N'\right]^{-1}\lambda_N\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\\=&-\frac{N}{2}tr\left\{\lambda_N\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\,\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N'\left[\zeta_N\sum_{i=1}^{N^{obs}}\lambda_N\tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\,\tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N'\zeta_N'\right.\right.\\&\left.\left.\times+\frac{N}{J}\lambda_N\bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\,\widehat{\Sigma}_\chi\bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N'\right]^{-1}\right\}\\=&-\frac{N}{2}tr\left[C(\boldsymbol{\theta})\,D^{-1}(\boldsymbol{\theta})\right],\end{aligned}$$

where $C(\boldsymbol{\theta})$ and $D(\boldsymbol{\theta})$ are symmetric. Then following Magnus and Neudecker (1995), we have

$$\begin{aligned}d\left\{tr\left[C(\boldsymbol{\theta})\,D^{-1}(\boldsymbol{\theta})\right]\right\}=&tr\left\{dC(\boldsymbol{\theta})\,D^{-1}(\boldsymbol{\theta})+C(\boldsymbol{\theta})\,dD^{-1}(\boldsymbol{\theta})\right\}\\=&tr\left\{D^{-1}(\boldsymbol{\theta})\,dC(\boldsymbol{\theta})+C(\boldsymbol{\theta})\,D^{-1}(\boldsymbol{\theta})\,dD(\boldsymbol{\theta})\,D^{-1}(\boldsymbol{\theta})\right\}\\=&tr\left\{D^{-1}(\boldsymbol{\theta})\,dC(\boldsymbol{\theta})-D^{-1}(\boldsymbol{\theta})\,C(\boldsymbol{\theta})\,D^{-1}(\boldsymbol{\theta})\,dD(\boldsymbol{\theta})\right\}.\end{aligned}$$

Before we derive the first-order and second-order differentiation of $L_N(\boldsymbol{\theta})$, we consider the following formula,

$$
\begin{aligned}
& tr\left\{K_1(\boldsymbol{\theta})\, dD(\boldsymbol{\theta})\, K_2(\boldsymbol{\theta})\right\} \\
=& tr\left\{K_1(\boldsymbol{\theta})\,\zeta_N\lambda_N d\left[\sum_{i=1}^{N^{obs}} \tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\,\tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\right]\lambda_N'\zeta_N' K_2(\boldsymbol{\theta})\right\} \\
& + \frac{N}{J} tr\left\{K_1(\boldsymbol{\theta})\,\lambda_N d\left[\bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\,\widehat{\Sigma}_\chi \bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\right]\lambda_N' K_2(\boldsymbol{\theta})\right\} \\
=& \sum_{i=1}^{N^{obs}} tr\left\{K_1(\boldsymbol{\theta})\,\zeta_N\lambda_N\left[\nabla_\theta\tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\, d\boldsymbol{\theta}\tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})' + \tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\, d\boldsymbol{\theta}'\nabla_\theta\tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\right]\lambda_N'\zeta_N' K_2(\boldsymbol{\theta})\right\} + \\
& \frac{N}{J} tr\left\{K_1(\boldsymbol{\theta})\,\lambda_N\left[\nabla_\theta\bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\, d\boldsymbol{\theta}\widehat{\Sigma}_\chi\bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})' + \bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\,\widehat{\Sigma}_\chi d\boldsymbol{\theta}'\nabla_\theta\bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\right]\lambda_N' K_2(\boldsymbol{\theta})\right\} \\
=& \sum_{i=1}^{N^{obs}} tr\left\{\tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N'\zeta_N' K_2(\boldsymbol{\theta})\, K_1(\boldsymbol{\theta})\,\zeta_N\lambda_N\nabla_\theta\tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\, d\boldsymbol{\theta}\right\} + \\
& \sum_{i=1}^{N^{obs}} tr\left\{\tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N'\zeta_N' K_1(\boldsymbol{\theta})' K_2(\boldsymbol{\theta})'\zeta_N\lambda_N\nabla_\theta\tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\, d\boldsymbol{\theta}\right\} + \\
& \frac{N}{J} tr\left\{\widehat{\Sigma}_\chi\bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N' K_2(\boldsymbol{\theta})\, K_1(\boldsymbol{\theta})\,\lambda_N\nabla_\theta\bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\, d\boldsymbol{\theta}\right\} + \\
& \frac{N}{J} tr\left\{\widehat{\Sigma}_\chi\bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N' K_1(\boldsymbol{\theta})' K_2(\boldsymbol{\theta})'\lambda_N\nabla_\theta\bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\, d\boldsymbol{\theta}\right\}.
\end{aligned}
\tag{A.2}
$$

Then, for the first term $tr\left[D^{-1}(\boldsymbol{\theta})\, dC(\boldsymbol{\theta})\right]$,

$$
\begin{aligned}
& tr\left[D^{-1}(\boldsymbol{\theta})\, dC(\boldsymbol{\theta})\right] \\
=& tr\left\{D^{-1}(\boldsymbol{\theta})\,\lambda_N d\left[\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\right]\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N' + D^{-1}(\boldsymbol{\theta})\,\lambda_N\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\left[d\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\right]'\lambda_N'\right\} \\
=& tr\left\{D^{-1}(\boldsymbol{\theta})\,\lambda_N\nabla_\theta\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\, d\boldsymbol{\theta}\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N' + D^{-1}(\boldsymbol{\theta})\,\lambda_N\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\left[\nabla_\theta\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\, d\boldsymbol{\theta}\right]'\lambda_N'\right\} \\
=& tr\left\{\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N' D^{-1}(\boldsymbol{\theta})\,\lambda_N\nabla_\theta\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\, d\boldsymbol{\theta} + \lambda_N\nabla_\theta\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\, d\boldsymbol{\theta}\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N' D^{-1}(\boldsymbol{\theta})\right\} \\
=& 2tr\left\{\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N' D^{-1}(\boldsymbol{\theta})\,\lambda_N\nabla_\theta\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\, d\boldsymbol{\theta}\right\} \\
=& 2tr\left\{\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N' W_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\,\lambda_N\nabla_\theta\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\, d\boldsymbol{\theta}\right\}.
\end{aligned}
$$

By formula (A.2),

$$
\begin{aligned}
& tr\left\{D^{-1}(\boldsymbol{\theta})\, C(\boldsymbol{\theta})\, D^{-1}(\boldsymbol{\theta})\, dD(\boldsymbol{\theta})\right\} \\
=& 2\sum_{i=1}^{N^{obs}} tr\left\{\tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N'\zeta_N' D^{-1}(\boldsymbol{\theta})\, C(\boldsymbol{\theta})\, D^{-1}(\boldsymbol{\theta})\,\zeta_N\lambda_N\nabla_\theta\tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\, d\boldsymbol{\theta}\right\} + \\
& \frac{2N}{J} tr\left\{\widehat{\Sigma}_\chi\bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N' D^{-1}(\boldsymbol{\theta})\, C(\boldsymbol{\theta})\, D^{-1}(\boldsymbol{\theta})\,\lambda_N\nabla_\theta\bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\, d\boldsymbol{\theta}\right\} \\
=& 2\sum_{i=1}^{N^{obs}} tr\left\{\tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N'\zeta_N' W_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\,\lambda_N\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\,\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N' W_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\,\zeta_N\lambda_N\nabla_\theta\tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\, d\boldsymbol{\theta}\right\} + \\
& \frac{2N}{J} tr\left\{\widehat{\Sigma}_\chi\bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N' W_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\,\lambda_N\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\,\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N' W_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\,\lambda_N\nabla_\theta\bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\, d\boldsymbol{\theta}\right\}
\end{aligned}
$$

Therefore,

$$\begin{aligned}
& dL_N(\boldsymbol{\theta}) \\
=& -Ntr\left\{\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N' W_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\lambda_N \nabla_\theta \bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})d\boldsymbol{\theta}\right\}+ \\
& N\sum_{i=1}^{N^{obs}} tr\left\{\tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N'\zeta_N' W_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\lambda_N\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N' W_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\zeta_N\lambda_N\nabla_\theta\tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})d\boldsymbol{\theta}\right\}+ \\
& \frac{N^2}{J}tr\left\{\widehat{\Sigma}_\chi\bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N' W_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\lambda_N\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N' W_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\lambda_N\nabla_\theta\bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})d\boldsymbol{\theta}\right\},
\end{aligned}$$

which implies,

$$\begin{aligned}
& \nabla_\theta L_N(\boldsymbol{\theta}) \\
=& -N\nabla_\theta\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N' W_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\lambda_N\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})+ \\
& N\sum_{i=1}^{N^{obs}}\nabla_\theta\tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N'\zeta_N' W_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\lambda_N\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N' W_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\zeta_N\lambda_N\tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})+ \\
& \frac{N^2}{J}\nabla_\theta\bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N' W_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\lambda_N\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\lambda_N' W_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\lambda_N\bar{g}_{N,\chi}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\widehat{\Sigma}_\chi.
\end{aligned}$$

By (A.1),

$$\bar{g}_N(\boldsymbol{\theta}_0;\widehat{\boldsymbol{\chi}}) = o_p\left(\frac{1}{\sqrt{n}}\right), W_N(\boldsymbol{\theta}_0;\widehat{\boldsymbol{\chi}}) = O_p(1), \tag{A.3}$$

it is obvious that

$$\begin{aligned}
& N\sum_{i=1}^{N^{obs}}\nabla_\theta\tilde{g}_i(\boldsymbol{\theta}_0;\widehat{\boldsymbol{\chi}})'\lambda_N'\zeta_N' W_N(\boldsymbol{\theta}_0;\widehat{\boldsymbol{\chi}})\lambda_N\bar{g}_N(\boldsymbol{\theta}_0;\widehat{\boldsymbol{\chi}})\bar{g}_N(\boldsymbol{\theta}_0;\widehat{\boldsymbol{\chi}})'\lambda_N' W_N(\boldsymbol{\theta}_0;\widehat{\boldsymbol{\chi}})\zeta_N\lambda_N\tilde{g}_i(\boldsymbol{\theta}_0;\widehat{\boldsymbol{\chi}}) \\
=& N^2 O_p\left(\frac{1}{\sqrt{N}}\right)o_p\left(\frac{1}{\sqrt{N}}\right)o_p\left(\frac{1}{\sqrt{N}}\right)O_p\left(\frac{1}{\sqrt{N}}\right) \\
=& o_p(1).
\end{aligned}$$

$$\begin{aligned}
& \frac{N^2}{J}\nabla_\theta\bar{g}_{N,\chi}(\boldsymbol{\theta}_0;\widehat{\boldsymbol{\chi}})'\lambda_N' W_N(\boldsymbol{\theta}_0;\widehat{\boldsymbol{\chi}})\lambda_N\bar{g}_N(\boldsymbol{\theta}_0;\widehat{\boldsymbol{\chi}})\bar{g}_N(\boldsymbol{\theta}_0;\widehat{\boldsymbol{\chi}})'\lambda_N' W_N(\boldsymbol{\theta}_0;\widehat{\boldsymbol{\chi}})\lambda_N\bar{g}_{N,\chi}(\boldsymbol{\theta}_0;\widehat{\boldsymbol{\chi}})\widehat{\Sigma}_\chi \\
=& NO_p(1)\,o_p\left(\frac{1}{\sqrt{N}}\right)o_p\left(\frac{1}{\sqrt{N}}\right)O_p(1) = o_p(1).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{\nabla_\theta L_N(\boldsymbol{\theta}_0)}{\sqrt{N}} =& -\nabla_\theta\bar{g}_N(\boldsymbol{\theta}_0;\widehat{\boldsymbol{\chi}})'\lambda_N' W_N(\boldsymbol{\theta}_0;\widehat{\boldsymbol{\chi}})\sqrt{N}\lambda_N\bar{g}_N(\boldsymbol{\theta}_0;\widehat{\boldsymbol{\chi}}) + o_p(1) \\
& \xrightarrow{d} N\left(0, G_\theta'\lambda' V^{-1}(\boldsymbol{\theta}_0)\lambda G_\theta'\right),
\end{aligned}$$

where $V^{-1}(\boldsymbol{\theta}_0) = \left(\lambda\Sigma_g\lambda' + \gamma\lambda G_\chi'\Sigma_\chi G_\chi\lambda'\right)^{-1}$ and $G_\theta = \nabla_\theta E\left[g_{i,t}(\boldsymbol{\theta}_0;\boldsymbol{\chi}_0)\right]$. This is because from (A.1),

$$\sqrt{N}\lambda_N\bar{g}(\boldsymbol{\theta}_0;\widehat{\boldsymbol{\chi}}) \xrightarrow{d} N\left(\mathbf{0}, \lambda\Sigma_g\lambda' + \gamma\lambda G_\chi'\Sigma_\chi G_\chi\lambda'\right),$$

where $G_\chi = E[\nabla_\chi \bar{g}(\boldsymbol{\theta}_0; \boldsymbol{\chi}_0)]$, $\gamma = \lim_{N\to\infty} \frac{N}{J}$, $\lambda = \lim_{N\to\infty} \lambda_N$, $\Sigma_g = E\left[g_i(\boldsymbol{\theta}_0; \boldsymbol{\chi}_0) g_i(\boldsymbol{\theta}_0; \boldsymbol{\chi}_0)'\right]$ and

$$\nabla_\theta \bar{g}_N(\boldsymbol{\theta}_0; \widehat{\boldsymbol{\chi}})' \xrightarrow{p} \nabla_\theta E[g_{i,t}(\boldsymbol{\theta}_0; \boldsymbol{\chi}_0)] = G_\theta,$$

$$\begin{aligned}
W_N(\boldsymbol{\theta}_0; \widehat{\boldsymbol{\chi}}) \xrightarrow{p} & V^{-1}(\boldsymbol{\theta}_0) \\
&= \left\{\lambda E\left[g_i(\boldsymbol{\theta}_0; \boldsymbol{\chi}_0) g_i(\boldsymbol{\theta}_0; \boldsymbol{\chi}_0)'\right]\lambda + \gamma\lambda E[g_{i,\chi}(\boldsymbol{\theta}_0; \boldsymbol{\chi}_0)] \Sigma_\chi E\left[g_{i,\chi}(\boldsymbol{\theta}_0; \boldsymbol{\chi}_0)'\right]\lambda'\right\}^{-1} \\
&= \left(\lambda\Sigma_g\lambda' + \gamma\lambda G_\chi' \Sigma_\chi G_\chi \lambda'\right)^{-1}.
\end{aligned}$$

Now turn to the second derivative of the criterion function, which is the Hessian matrix of $L_n(\boldsymbol{\theta})$. The second order differentiation,

$$\begin{aligned}
& d^2\left\{tr\left[A(\boldsymbol{\theta}) B^{-1}(\boldsymbol{\theta})\right]\right\} \\
=& d\left\{-tr\left\{\bar{g}_N(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}})' \lambda_N' W_N(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) \lambda_N \nabla_\theta \bar{g}_N(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) d\boldsymbol{\theta}\right\} + \right. \\
& \sum_{i=1}^{N^{obs}} tr\left\{\tilde{g}_i(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}})' \lambda_N' \zeta_N' W_N(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) \lambda_N \bar{g}_N(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) \bar{g}_N(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}})' \lambda_N' W_N(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) \zeta_N \lambda_N \nabla_\theta \tilde{g}_i(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) d\boldsymbol{\theta}\right\} + \\
& \left.\frac{N^2}{J} tr\left\{\widehat{\Sigma}_\chi \bar{g}_{N,\chi}(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}})' \lambda_N' W_N(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) \lambda_N \bar{g}_N(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) \bar{g}_N(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}})' \lambda_N' W_N(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) \lambda_N \nabla_\theta \bar{g}_{N,\chi}(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) d\boldsymbol{\theta}\right\}\right\}.
\end{aligned}$$

Following the preceding procedure to derive the first-order differentiation,, we can obtain the form of $\nabla_{\theta\theta'} L_n(\boldsymbol{\theta})$. Due to Assumptions 5-9, for any $\delta > 0$, $\nabla_{\theta\theta'} L_n(\boldsymbol{\theta})$ is continuous when $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \delta$ and we can have

$$\frac{\nabla_{\theta\theta'} L_N(\boldsymbol{\theta}_0)}{N} = -\nabla_\theta \bar{g}_N(\boldsymbol{\theta}_0; \widehat{\boldsymbol{\chi}})' \lambda_N' W_N(\boldsymbol{\theta}_0; \widehat{\boldsymbol{\chi}}) \lambda_N \nabla_\theta \bar{g}_N(\boldsymbol{\theta}_0; \widehat{\boldsymbol{\chi}}) + o_p(1).$$

Meanwhile, we have

$$M(\boldsymbol{\theta}) = -\frac{1}{2} E[g_i(\boldsymbol{\theta}; \boldsymbol{\chi}_0)]' \lambda' W(\boldsymbol{\theta}) \lambda E[g_i(\boldsymbol{\theta}; \boldsymbol{\chi}_0)],$$

where $W(\boldsymbol{\theta}) = V^{-1}(\boldsymbol{\theta}) = \left\{\lambda E\left[g_i(\boldsymbol{\theta}; \boldsymbol{\chi}_0) g_i(\boldsymbol{\theta}; \boldsymbol{\chi}_0)'\right]\lambda + \gamma\lambda E[g_{i,\chi}(\boldsymbol{\theta}; \boldsymbol{\chi}_0)] \Sigma_\chi E\left[g_{i,\chi}(\boldsymbol{\theta}; \boldsymbol{\chi}_0)'\right]\lambda'\right\}^{-1}$. Then,

$$\begin{aligned}
\nabla_{\theta\theta'} M(\boldsymbol{\theta}) = & -E[\nabla_\theta g_i(\boldsymbol{\theta}; \boldsymbol{\chi}_0)]' \lambda' W(\boldsymbol{\theta}) \lambda E[\nabla_\theta g_i(\boldsymbol{\theta}; \boldsymbol{\chi}_0)] - \\
& \{W(\boldsymbol{\theta}) E[g_i(\boldsymbol{\theta}; \boldsymbol{\chi}_0)] \otimes I_d\} E[\nabla_{\theta\theta'} g_i(\boldsymbol{\theta}; \boldsymbol{\chi}_0)] - \\
& -\frac{1}{2} E[g_i(\boldsymbol{\theta}; \boldsymbol{\chi}_0)]' \lambda' \nabla_{\theta\theta'} W(\boldsymbol{\theta}) \lambda E[g_i(\boldsymbol{\theta}; \boldsymbol{\chi}_0)]
\end{aligned}$$

$$\nabla_{\theta\theta'} M(\boldsymbol{\theta}_0) = -E[g_i(\boldsymbol{\theta}_0, \boldsymbol{\chi}_0)]' V^{-1}(\boldsymbol{\theta}_0) E[g_i(\boldsymbol{\theta}_0, \boldsymbol{\chi}_0)] + o_p(1).$$

And thus,

$$\frac{\nabla_{\theta\theta'} L_N(\boldsymbol{\theta}_0)}{N} - \nabla_{\theta\theta'} M(\boldsymbol{\theta}_0) \xrightarrow{p} 0.$$

Then for $\epsilon > 0$, $N > 0$, $\exists \delta_1(\epsilon, N) > 0$, $\forall \boldsymbol{\theta} \in \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \delta_1(\epsilon, N)\}$, due to the continuity,

$$\sup_{\boldsymbol{\theta}} \left\|\frac{\nabla_{\theta\theta'} L_N(\boldsymbol{\theta})}{N} - \frac{\nabla_{\theta\theta'} L_N(\boldsymbol{\theta}_0)}{N}\right\| < \frac{1}{3}\epsilon.$$

$\exists\delta_2(\epsilon)>0$, $\forall\boldsymbol{\theta}\in\{\boldsymbol{\theta}:\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|<\delta_2(\epsilon)\}$, due to continuity,

$$\sup_{\boldsymbol{\theta}}\|\nabla_{\theta\theta'}M(\boldsymbol{\theta})-\nabla_{\theta\theta'}M(\boldsymbol{\theta}_0)\|<\frac{1}{3}\epsilon.$$

And for $\varepsilon>0$, $\exists N(\varepsilon,\epsilon)>0$, $\forall N>N(\varepsilon,\epsilon)$,

$$P\left\{\left\|\frac{\nabla_{\theta\theta'}L_N(\boldsymbol{\theta}_0)}{N}-\nabla_{\theta\theta'}M(\boldsymbol{\theta}_0)\right\|<\frac{1}{3}\epsilon\right\}\geq 1-\varepsilon.$$

Therefore, for any $\varepsilon>0$, $\forall N>N(\epsilon,\varepsilon)$, let $\delta(\epsilon,N)=\min\{\delta_1(\epsilon,N),\delta_2(\epsilon)\}$, $\forall\boldsymbol{\theta}\in\{\boldsymbol{\theta}:\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|<\delta(\varepsilon,N)\}$,

$$\begin{aligned}\sup_{\boldsymbol{\theta}}\left\|\frac{\nabla_{\theta\theta'}L_N(\boldsymbol{\theta})}{N}-\nabla_{\theta\theta'}M(\boldsymbol{\theta})\right\|\leq&\sup_{\boldsymbol{\theta}}\left\|\frac{\nabla_{\theta\theta'}L_N(\boldsymbol{\theta})}{N}-\frac{\nabla_{\theta\theta'}L_N(\boldsymbol{\theta}_0)}{N}\right\|+\sup_{\boldsymbol{\theta}}\|\nabla_{\theta\theta'}M(\boldsymbol{\theta})-\nabla_{\theta\theta'}M(\boldsymbol{\theta}_0)\|\\&+\left\|\frac{\nabla_{\theta\theta'}L_n(\boldsymbol{\theta}_0)}{n}-\nabla_{\theta\theta'}M(\boldsymbol{\theta}_0)\right\|\\&<\frac{2}{3}\epsilon+\left\|\frac{\nabla_{\theta\theta'}L_N(\boldsymbol{\theta}_0)}{N}-\nabla_{\theta\theta'}M(\boldsymbol{\theta}_0)\right\|.\end{aligned}$$

Then

$$\left\{\sup_{\boldsymbol{\theta}}\left\|\frac{\nabla_{\theta\theta'}L_N(\boldsymbol{\theta}_0)}{N}-\nabla_{\theta\theta'}M(\boldsymbol{\theta}_0)\right\|<\frac{1}{3}\epsilon\right\}\subset\left\{\sup_{\boldsymbol{\theta}}\left\|\frac{\nabla_{\theta\theta'}L_N(\boldsymbol{\theta})}{N}-\nabla_{\theta\theta'}M(\boldsymbol{\theta})\right\|<\epsilon\right\},$$

which implies

$$P\left\{\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|<\delta(\varepsilon)}\left\|\frac{\nabla_{\theta\theta'}L_N(\boldsymbol{\theta})}{N}-\nabla_{\theta\theta'}M(\boldsymbol{\theta})\right\|<\epsilon\right\}\geq 1-\varepsilon,$$

in other words, for $\epsilon>0$,

$$\limsup_{n\to\infty}P\left\{\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|<\delta(\varepsilon)}\left\|\frac{\nabla_{\theta\theta'}L_N(\boldsymbol{\theta})}{N}-\nabla_{\theta\theta'}M(\boldsymbol{\theta})\right\|>\epsilon\right\}=0$$

Therefore, the Lemma 2 in CH (2003) is satisfied. By the Theorem 2 in CH (2003), for the etimator $\widehat{\boldsymbol{\theta}}$ defined in (15), we can have

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\xrightarrow{d}\tau+N(\mathbf{0},\Sigma_\theta),$$

where

$$\Sigma_\theta=\left[G_\theta'\lambda'\left(\lambda\Sigma_g\lambda'+\gamma\lambda G_\chi'\Sigma_\chi G_\chi\lambda'\right)^{-1}\lambda G_\theta\right]^{-1},$$

$G_\theta=\nabla_\theta E[g_i(\boldsymbol{\theta}_0;\boldsymbol{\chi}_0)]$, $G_\chi=E[\nabla_\chi g_i(\boldsymbol{\theta}_0;\boldsymbol{\chi}_0)]$, $\gamma=\lim_{N\to\infty}\frac{N}{J}$, $\lambda=\lim_{N\to\infty}\lambda_N$, $\Sigma_g=E\left[g_i(\boldsymbol{\theta}_0;\boldsymbol{\chi}_0)\,g_i(\boldsymbol{\theta}_0;\boldsymbol{\chi}_0)'\right]$, $\tau=\arg\inf_{z\in R^d}\left\{\int_{R^d}\rho(z-u)\,f(u;\mathbf{0},G_\theta'\lambda'W(\boldsymbol{\theta}_0)\lambda G_\theta)\,du\right\}$.

A.3 The proof of Theorem 4.3

Lemma A.1. *By the definition of Δ_j in (24), $\forall \boldsymbol{\theta} \in \boldsymbol{\Theta}$,*

$$\bar{g}_N(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) - \bar{g}_N^j(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) = O_p(\Delta_j).$$

$$\bar{g}_{N,\chi}(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) - \bar{g}_{N,\chi}^j(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) = O_p(\Delta_j),$$

$$V_N^j(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) - V_N(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) = O_p(\Delta_j).$$

Proof: By definition, for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$\bar{g}_N(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) = (\bar{g}_{t_{min}}(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}), \ldots, \bar{g}_{T_r}(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}))' = \left(\frac{1}{N_{t_{min}}} \sum_{i=1}^{N_{t_{min}}} g_{i,t_{min}}(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}), \ldots, \frac{1}{N_{T_r}} \sum_{i=1}^{N_{T_r}} g_{i,T_r}(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) \right)'.$$

Let $t \in [t_{min}, T_r]$, for $g_{i,t}(\boldsymbol{\theta}; \boldsymbol{\chi}) = C_{i,t}^d - C_t\left(M_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \boldsymbol{\chi}\right)$, $g_{i,t}^j(\boldsymbol{\theta}; \boldsymbol{\chi}) = C_{i,t}^d - C_t^j\left(M_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \boldsymbol{\chi}\right)$,

$$\begin{aligned}
\bar{g}_t(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) - \bar{g}_t^j(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) =& \frac{1}{N_t} \sum_{i=1}^{N_t} \left[g_{i,t}(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) - g_{i,t}^j(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) \right] \\
=& \frac{1}{N_t} \sum_{i=1}^{N_t} \left[C_t\left(M_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \widehat{\boldsymbol{\chi}}\right) - C_t^j\left(M_{i,t}^d, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \widehat{\boldsymbol{\chi}}\right) \right] \\
=& \frac{1}{N_t} \sum_{i=1}^{N_t} E_{P_{i,t}} \left\{ \left[c_t\left(\frac{M_{i,t}^d}{P_{i,t}}, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \widehat{\boldsymbol{\chi}}\right) - c_t^j\left(\frac{M_{i,t}^d}{P_{i,t}}, \boldsymbol{z}_{i,t}^d; \boldsymbol{\theta}, \widehat{\boldsymbol{\chi}}\right) \right] P_{i,t} \right\} \\
\leq& \Delta_j \frac{1}{N_t} \sum_{i=1}^{N_t} E_{P_{i,t}}(P_{i,t}) = O_p(\Delta_j),
\end{aligned}$$

which implies

$$\bar{g}_N(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) - \bar{g}_N^j(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) = O_p(\Delta_j).$$

And similarly, we can also have

$$\bar{g}_{N,\chi}(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) - \bar{g}_{N,\chi}^j(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) = O_p(\Delta_j).$$

And thus for,

$$\begin{aligned}
V_N^j(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) =& \zeta_N \sum_{i=1}^{N^{obs}} \lambda_N \tilde{g}_i^j(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) \tilde{g}_i^j(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}})' \lambda_N' \zeta_N' \\
&+ \frac{N}{J} \lambda_N \bar{g}_{N,\chi}^j(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}}) \widehat{\Sigma}_\chi \bar{g}_{N,\chi}^j(\boldsymbol{\theta}; \widehat{\boldsymbol{\chi}})' \lambda_N',
\end{aligned}$$

the first term, since $\tilde{g}_i^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})$ and $\tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})$ are continuous and $\boldsymbol{\Theta}$ is compact by assumptions,

$$
\begin{aligned}
&\zeta_N \sum_{i=1}^{N^{obs}} \lambda_N \tilde{g}_i^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) \tilde{g}_i^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})' \lambda_N' \zeta_N' - \zeta_N \sum_{i=1}^{N^{obs}} \lambda_N \tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) \tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})' \lambda_N' \zeta_N' \\
=&\zeta_N \lambda_N \sum_{i=1}^{N^{obs}} \left[\tilde{g}_i^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) \tilde{g}_i^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})' - \tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) \tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\right] \lambda_N' \zeta_N' \\
=&\zeta_N \lambda_N \sum_{i=1}^{N^{obs}} \left[\tilde{g}_i^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) \tilde{g}_i^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})' - \tilde{g}_i^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) \tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\right] \lambda_N' \zeta_N' + \\
&\zeta_N \lambda_N \sum_{i=1}^{N^{obs}} \left[\tilde{g}_i^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) \tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})' - \tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) \tilde{g}_i(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})'\right] \lambda_N' \zeta_N' \\
=&O_p(\Delta_j).
\end{aligned}
$$

And the second term is similar, which means

$$V_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) - V_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) = O_p(\Delta_j).$$

The Proof of Theorem 4.3: The criterion function for the case using analytical solution and the one approximated by numerical methods are

$$L_N(\boldsymbol{\theta}) = -\frac{N}{2} \bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})' \lambda_N' W_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) \lambda_N \bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}),$$

and

$$L_N^j(\boldsymbol{\theta}) = -\frac{N}{2} \bar{g}_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})' \lambda_N' W_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) \lambda_N \bar{g}_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}),$$

respectively. By Lemma $A.1$, if $N\Delta_j \to 0$, as $N \to \infty$, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$V_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) - V_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) = O_p(\Delta_j),$$

so that

$$
\begin{aligned}
\left[V_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\right]^{-1} \left[V_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) - V_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\right] V_N^{-1}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) =& V_N^{-1}(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) - \left[V_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\right]^{-1} \\
=&O_p(1)\, O(\Delta_j)\, O_p(1) \\
=&O_p(\Delta_j).
\end{aligned}
$$

So that,

$$
\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\{L_N^j(\boldsymbol{\theta}) - \tilde{L}_N^j(\boldsymbol{\theta})\right\} \\
=&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\{-\frac{N}{2} \bar{g}_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})' \lambda_N' W_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) \lambda_N \bar{g}_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) + \frac{N}{2} \bar{g}_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})' \lambda_N' W_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) \lambda_N \bar{g}_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\right\} \\
=&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\{-\frac{N}{2} \bar{g}_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})' \lambda_N' \left[W_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) - W_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\right] \lambda_N \bar{g}_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\right\} \\
=&N O_p(\Delta_j) = O_p(N\Delta_j).
\end{aligned}
$$

Therefore, denote $\tilde{W}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) = \lambda_N' W_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})\lambda_N$,

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| L_N(\boldsymbol{\theta}) - L_N^j(\boldsymbol{\theta}) \right\| \\
\leq &\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| L_N(\boldsymbol{\theta}) - \tilde{L}_N^j(\boldsymbol{\theta}) \right\| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| L_N^j(\boldsymbol{\theta}) - \tilde{L}_N^j(\boldsymbol{\theta}) \right\| \\
\leq &\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{N}{2}\bar{g}_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})' \tilde{W}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) \bar{g}_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) - \frac{N}{2}\bar{g}_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})' \tilde{W}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) \bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) \right\| + \\
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{N}{2}\bar{g}_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})' \tilde{W}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) \bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) - \frac{N}{2}\bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})' \tilde{W}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) \bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) \right\| + O_p(N\Delta_j) \\
\leq &\frac{N}{2}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \bar{g}_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})' \tilde{W}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) \right\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \bar{g}_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) - \bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) \right\| + \\
&\frac{N}{2}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}})' \tilde{W}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) \right\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \bar{g}_N^j(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) - \bar{g}_N(\boldsymbol{\theta};\widehat{\boldsymbol{\chi}}) \right\| + O_p(N\Delta_j) \\
= &O_p(N\Delta_j).
\end{aligned}$$

Therefore, when $N\Delta_j \to 0$, as $N \to \infty$, $L_N(\boldsymbol{\theta}) - L_N^j(\boldsymbol{\theta}) \xrightarrow{p} 0$ over $\boldsymbol{\Theta}$. Further, due to the compactness of $\boldsymbol{\Theta}$ and the Taylor expansion,

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \exp[L_N(\boldsymbol{\theta})] - \exp\left[L_N^j(\boldsymbol{\theta})\right] \right\| \\
= &\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\exp[L_N(\boldsymbol{\theta})]\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \exp\left[L_N(\boldsymbol{\theta}) - L_N^j(\boldsymbol{\theta})\right] - 1 \right\| \\
\leq &C_1 \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \exp\left(L_N\left(\tilde{\boldsymbol{\theta}}\right) - L_N^j\left(\tilde{\boldsymbol{\theta}}\right)\right)\left[L_N(\boldsymbol{\theta}) - L_N^j(\boldsymbol{\theta})\right] \right\| \\
\leq &C_1 \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \exp\left(L_N\left(\tilde{\boldsymbol{\theta}}\right) - L_N^j\left(\tilde{\boldsymbol{\theta}}\right)\right) \right\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \left[L_N(\boldsymbol{\theta}) - L_N^j(\boldsymbol{\theta})\right] \right\| \\
\approx &C_1(1 + O_p(N\Delta_j))\,O_p(N\Delta_j) \\
= &O_p(N\Delta_j),
\end{aligned} \tag{A.4}$$

where $\tilde{\boldsymbol{\theta}}$ is between $\mathbf{0}$ and $\boldsymbol{\theta}$.

$$\begin{aligned}
&\int_{\boldsymbol{\Theta}} \exp\left[L_N^j(\boldsymbol{\theta})\right]\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta} - \int_{\boldsymbol{\Theta}} \exp[L_N(\boldsymbol{\theta})]\,\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta} \\
= &\int_{\boldsymbol{\Theta}} \exp\left[L_N^j(\boldsymbol{\theta}) - L_N(\boldsymbol{\theta})\right]\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta} \\
\leq &\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \exp[L_N(\boldsymbol{\theta})] - \exp\left[L_N^j(\boldsymbol{\theta})\right] \right\| \int_{\boldsymbol{\Theta}} \pi(\boldsymbol{\theta})\,d\boldsymbol{\theta} \\
= &O_p(N\Delta_j).
\end{aligned} \tag{A.5}$$

Following the proof of Theorem 4.1, we define

$$J(\boldsymbol{\theta}_0) = -E\left[\nabla_\theta g_i(\boldsymbol{\theta}_0, \boldsymbol{\chi}_0)\right]' V^{-1}(\boldsymbol{\theta}_0)\, E\left[\nabla_\theta g_i(\boldsymbol{\theta}_0, \boldsymbol{\chi}_0)\right],$$

and

$$h \equiv \sqrt{N}\left(\boldsymbol{\theta} - T_N\right), T_N = \boldsymbol{\theta}_0 + \frac{1}{\sqrt{N}} U_N, U_N = \frac{1}{\sqrt{N}} J^{-1}\left(\boldsymbol{\theta}_0\right) \nabla_\theta L_N\left(\boldsymbol{\theta}_0\right),$$

so that, let $H_N = \left\{\sqrt{N}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right) - U_N : \boldsymbol{\theta} \in \boldsymbol{\Theta}\right\}$, $p_N\left(\boldsymbol{\theta}\right)$ and $p_N^j\left(\boldsymbol{\theta}\right)$ can be transformed into $\frac{1}{\sqrt{N}} p_N^*\left(h\right)$ and $\frac{1}{\sqrt{N}} p_N^{*j}\left(h\right)$, respectively, where,

$$p_N^{*j}\left(h\right) = \frac{\pi\left(T_N + \frac{h}{\sqrt{N}}\right)\exp\left[L_N^j\left(T_N + \frac{h}{\sqrt{N}}\right)\right]}{\int_{H_N}\pi\left(T_N + \frac{h}{\sqrt{N}}\right)\exp\left[L_N^j\left(T_N + \frac{h}{\sqrt{N}}\right)\right]dh} = \frac{\pi\left(T_N + \frac{h}{\sqrt{N}}\right)\exp\left[L_N^j\left(T_N + \frac{h}{\sqrt{N}}\right)\right]}{C^j},$$

$$p_N^{*}\left(h\right) = \frac{\pi\left(T_N + \frac{h}{\sqrt{N}}\right)\exp\left[L_N\left(T_N + \frac{h}{\sqrt{N}}\right)\right]}{\int_{H_N}\pi\left(T_N + \frac{h}{\sqrt{N}}\right)\exp\left[L_N\left(T_N + \frac{h}{\sqrt{N}}\right)\right]dh} = \frac{\pi\left(T_N + \frac{h}{\sqrt{N}}\right)\exp\left[L_N\left(T_N + \frac{h}{\sqrt{N}}\right)\right]}{C}.$$

The corresponding transformed risk functions of $R_N^j\left(\boldsymbol{\xi}\right)$ and $R_N\left(\boldsymbol{\xi}\right)$ are denoted as $Q_N^j\left(\zeta\right)$ and $Q_N\left(\zeta\right)$, respectively, where

$$Q_N^j\left(\zeta\right) = \int_{H_N}\rho\left(h + U_N - \zeta\right)p_N^{*j}\left(h\right)dh,$$

$$Q_N\left(\zeta\right) = \int_{H_N}\rho\left(h + U_N - \zeta\right)p_N^{*}\left(h\right)dh.$$

As in Theorem 4.1, the Lemma 1 and Lemma 2 in CH (2003) are satisfied, which implies that the Theorem 1 and Theorem 2 in their paper hold. So that we have for any $0 \le \alpha < \infty$,

$$\int_{H_N}\|h\|^{\alpha}\left|p_N^*\left(h\right) - p_\infty\left(h\right)\right|dh \xrightarrow{p} 0,$$

where

$$p_\infty\left(h\right) = \sqrt{\frac{\left|J\left(\boldsymbol{\theta}_0\right)\right|}{\left(2\pi\right)^d}}\exp\left(-\frac{1}{2}h' J\left(\boldsymbol{\theta}_0\right)h\right),$$

and

$$\lim_{N\to\infty}\int_{H_N}\|h\|^{\alpha}p_\infty\left(h\right)dh = C_\alpha < \infty.$$

$$Q_\infty\left(\zeta\right) = \int_{R^d}\rho\left(h + U_N - \zeta\right)p_\infty\left(h\right)dh.$$

Therefore,

$$
\begin{aligned}
&\int_{H_N} \|h\|^{\alpha} \left|p_N^{*j}(h) - p_{\infty}(h)\right| dh \\
\leq &\int_{H_N} \|h\|^{\alpha} \left|p_N^{*j}(h) - p_N^{*}(h)\right| dh + \int_{H_N} \|h\|^{\alpha} \left|p_N^{*}(h) - p_{\infty}(h)\right| dh \\
= &\int_{H_N} \|h\|^{\alpha} \pi\left(T_N + \frac{h}{\sqrt{N}}\right) \left| \frac{\exp\left[L_N^j\left(T_N + \frac{h}{\sqrt{N}}\right)\right]}{C^j} - \frac{\exp\left[L_N\left(T_N + \frac{h}{\sqrt{N}}\right)\right]}{C} \right| dh + o_p(1) \\
\leq &\int_{H_N} \|h\|^{\alpha} \pi\left(T_N + \frac{h}{\sqrt{N}}\right) \left| \frac{\exp\left[L_N^j\left(T_N + \frac{h}{\sqrt{N}}\right)\right]}{C^j} - \frac{\exp\left[L_N\left(T_N + \frac{h}{\sqrt{N}}\right)\right]}{C^j} \right| dh + \\
&\int_{H_N} \|h\|^{\alpha} \pi\left(T_N + \frac{h}{\sqrt{N}}\right) \left| \frac{\exp\left[L_N\left(T_N + \frac{h}{\sqrt{N}}\right)\right]}{C^j} - \frac{\exp\left[L_N\left(T_N + \frac{h}{\sqrt{N}}\right)\right]}{C} \right| dh + o_p(1).
\end{aligned}
$$

For the second term, it is obvious that

$$
C^j = \int_{H_N} \pi\left(T_N + \frac{h}{\sqrt{N}}\right) \exp\left[L_N^j\left(T_N + \frac{h}{\sqrt{N}}\right)\right] dh = \int_{\boldsymbol{\Theta}} \exp\left[L_N^j(\boldsymbol{\theta})\right] \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta},
$$

$$
C = \int_{H_N} \pi\left(T_N + \frac{h}{\sqrt{N}}\right) \exp\left[L_N\left(T_N + \frac{h}{\sqrt{N}}\right)\right] dh = \int_{\boldsymbol{\Theta}} \exp\left[L_N(\boldsymbol{\theta})\right] \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta}.
$$

which implies $C^j - C = O_p(N\Delta_j)$ by (A.5) and then for the first term, since $N\Delta_j \to 0$,

$$
\begin{aligned}
&\int_{H_N} \|h\|^{\alpha} \pi\left(T_N + \frac{h}{\sqrt{N}}\right) \left| \frac{\exp\left[L_N\left(T_N + \frac{h}{\sqrt{N}}\right)\right]}{C^j} - \frac{\exp\left[L_N\left(T_N + \frac{h}{\sqrt{N}}\right)\right]}{C} \right| dh \\
= &\left|\frac{1}{C^j} - \frac{1}{C}\right| \int_{H_N} \|h\|^{\alpha} \pi\left(T_N + \frac{h}{\sqrt{N}}\right) \exp\left[L_N\left(T_N + \frac{h}{\sqrt{N}}\right)\right] dh \\
= &\left|\frac{1}{C^j} - \frac{1}{C}\right| \int_{H_N} \|h\|^{\alpha} p_{\infty}(h)\, dh + o_p(1) \\
= &C_{\alpha} \left|\frac{1}{C^j} - \frac{1}{C}\right| + o_p(1) = O_p(N\Delta_j).
\end{aligned}
$$

For the second term, by the Taylor expansion and (A.5)

$$\begin{aligned}
&\int_{H_N} \|h\|^\alpha \pi\left(T_N+\frac{h}{\sqrt{N}}\right)\left|\frac{\exp\left[L_N^j\left(T_N+\frac{h}{\sqrt{N}}\right)\right]}{C^j}-\frac{\exp\left[L_N\left(T_N+\frac{h}{\sqrt{N}}\right)\right]}{C^j}\right| dh \\
=&\frac{C}{C^j}\int_{H_N}\|h\|^\alpha \frac{1}{C}\pi\left(T_N+\frac{h}{\sqrt{N}}\right)\exp\left[L_N\left(T_N+\frac{h}{\sqrt{N}}\right)\right] \\
&\times\left|L_N^j\left(T_N+\frac{h}{\sqrt{N}}\right)-L_N\left(T_N+\frac{h}{\sqrt{N}}\right)+o_p\left(N\Delta_j\right)\right| dh \\
=&O_p(1)\,O_p\left(N\Delta_j\right)\int_{H_N}\|h\|^\alpha\frac{1}{C}\pi\left(T_N+\frac{h}{\sqrt{N}}\right)\exp\left[L_N\left(T_N+\frac{h}{\sqrt{N}}\right)\right] dh \\
=&O_p(1)\,O_p\left(N\Delta_j\right)C_\alpha \\
=&O_p\left(N\Delta_j\right).
\end{aligned}$$

Therefore,

$$\int_{H_N}\|h\|^\alpha\left|p_N^{*j}(h)-p_\infty(h)\right| dh = O_p\left(N\Delta_j\right).$$

By the Assumption 3, $\rho(u)\le 1+|u|^p$ and by $|a+b|^p\le 2^{p-1}|a|^p+2^{p-1}|b|^p$ for $p\ge 1$. For any fixed ζ,

$$\begin{aligned}
\left|Q_N^j(\zeta)-Q_\infty(\zeta)\right| \le& \int_{H_N}\left(1+\|h+U_N-\zeta\|^p\right)\left|p_N^{*j}(h)-p_\infty(h)\right| dh \\
&+\int_{R^d\backslash H_N}\left(1+\|h+U_N-\zeta\|^p\right)p_\infty(h)\,dh \\
\le&\int_{H_N}\left(1+2^{p-1}\|h\|^{p-1}+2^{p-1}\|U_N-\zeta\|^{p-1}\right)\left|p_N^{*j}(h)-p_\infty(h)\right| dh \\
&+\int_{R^d\backslash H_N}\left(1+2^{p-1}\|h\|^{p-1}+2^{p-1}\|U_N-\zeta\|^{p-1}\right)p_\infty(h)\,dh \\
=&\int_{H_N}\left(1+2^{p-1}\|h\|^{p-1}+O_p(1)\right)\left|p_N^{*j}(h)-p_\infty(h)\right| dh \\
&+\int_{R^d\backslash H_N}\left(1+2^{p-1}\|h\|^{p-1}+O_p(1)\right)p_\infty(h)\,dh.
\end{aligned}$$

From above discussions,

$$\int_{H_N}\left(1+2^{p-1}\|h\|^{p-1}+O_p(1)\right)\left|p_N^{*j}(h)-p_\infty(h)\right| dh = O_p\left(N\Delta_j\right),$$

and by the exponentially small tails of the normal density,

$$\int_{R^d\backslash H_N}\left(1+2^{p-1}\|h\|^{p-1}+O_p(1)\right)p_\infty(h)\,dh = o_p(1).$$

Hence, if $N\Delta_j\to 0$, given fixed ζ, $Q_N^j(\zeta)-Q_\infty(\zeta)\xrightarrow{p}0$.

Then, we show that both $Q_N^j(\zeta)$ and $Q_\infty(\zeta)$ are convex, for any given ζ and $\tilde{\zeta}$, and $\alpha \in [0,1]$,

$$
\begin{aligned}
Q_N^j\left(\alpha\zeta + (1-\alpha)\tilde{\zeta}\right) =& \int_{H_N} \rho\left[h + U_N - \alpha\zeta - (1-\alpha)\tilde{\zeta}\right] p_N^{j*}(h)\, dh \\
=& \int_{H_N} \rho\left[\alpha(h + U_N - \zeta) + (1-\alpha)\left(h + U_N - \tilde{\zeta}\right)\right] p_N^{j*}(h)\, dh \\
\leq& \alpha \int_{H_N} \rho(h + U_N - \zeta)\, p_N^{j*}(h)\, dh \\
&+ (1-\alpha)\int_{H_N} \rho\left(h + U_N - \tilde{\zeta}\right) p_N^{j*}(h)\, dh \\
=& \alpha Q_N^j(\zeta) + (1-\alpha) Q_N^j\left(\tilde{\zeta}\right).
\end{aligned}
$$

Hence $Q_N^j(\zeta)$ is convex. Similarly, $Q_\infty(\zeta)$ is also convex. Further,

$$
\begin{aligned}
Q_\infty(\zeta) &\leq \int_{H_N} \left(1 + 2^{p-1}\|h\|^{p-1} + 2^{p-1}\|U_N - \zeta\|^{p-1}\right) p_\infty(h)\, dh \\
&= 1 + 2^{p-1}\int_{H_N} \|h\|^{p-1} p_\infty(h)\, dh + 2^{p-1}\int_{H_N} \|U_N - \zeta\|^{p-1} p_\infty(h)\, dh \\
&= O_p(1).
\end{aligned}
$$

And by the same logic $Q_N^j(\zeta) = O_p(1)$.

If $N\Delta_j \to 0$, by the convexity lemma of Polard (1991), pointwise convergence entails the uniform convergence over the compact set $\mathfrak{B}$,

$$
\sup_{\zeta\in\mathfrak{B}} \left| Q_N^j(\zeta) - Q_\infty(\zeta) \right| \xrightarrow{p} 0.
$$

For $Q_\infty(\zeta) = \int_{R^d} \rho(h + U_N - \zeta)\, p_\infty(h)\, dh$, it is minimized at $\zeta^* = \tau + U_N = O_p(1)$. And $Q_N^j(\zeta)$ is minimized at $\sqrt{N}\left(\widehat{\boldsymbol{\theta}}^j - \boldsymbol{\theta}_0\right)$. Following CH, the uniform convergence property above as well as the convexity property imply that $\sqrt{N}\left(\widehat{\boldsymbol{\theta}}^j - \boldsymbol{\theta}_0\right) = U_N + \tau + o_p(1)$. Combined with the fact that

$$
U_N = \frac{1}{\sqrt{N}} J^{-1}(\boldsymbol{\theta}_0)\, \nabla_\theta L_N(\boldsymbol{\theta}_0) \xrightarrow{d} N(0, \Sigma_\theta),
$$

the results in the theorem follows.

A.4 The Proof of Corollary 4.4

The asymptotic theory is easily obtained from Theorem 4.3. For

$$
E^j\left[N\left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}^j\right)\left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}^j\right)' \middle| \cdot \right] = \int_{\boldsymbol{\Theta}} N\left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}^j\right)\left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}^j\right)' p_N^j(\boldsymbol{\theta})\, d\boldsymbol{\theta},
$$

we let

$$
h \equiv \sqrt{N}(\boldsymbol{\theta} - T_N), T_N = \boldsymbol{\theta}_0 + \frac{1}{\sqrt{N}} U_N, U_N = \frac{1}{\sqrt{N}} J^{-1}(\boldsymbol{\theta}_0)\, \nabla_\theta L_N(\boldsymbol{\theta}_0),
$$

then

$$\boldsymbol{\theta} = \frac{h}{\sqrt{N}} + T_N, \bar{\boldsymbol{\theta}}^j = \frac{\bar{h}^j}{\sqrt{N}} + T_N, \bar{h}^j = \int_{H_N} h p_N^{*j}(h)\, dh,$$

so that

$$\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}^j = \frac{1}{\sqrt{N}}\left(h - \bar{h}^j\right).$$

Therefore,

$$\begin{aligned}
&\int_{\boldsymbol{\Theta}} N\left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}^j\right)\left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}^j\right)' p_N^j(\boldsymbol{\theta})\, d\boldsymbol{\theta} \\
=& \int_{H_N} \left(h - \bar{h}^j\right)\left(h - \bar{h}^j\right)' p_N^{*j}(h)\, dh \\
=& \int_{H_N} h h' p_N^{*j}(h)\, dh - \bar{h}^j \bar{h}^{j\prime},
\end{aligned}$$

As in Theorem 4.3, if $N\Delta_j \to 0$, $\int_{H_N} \|h\|^{\alpha} \left| p_N^{*j}(h) - p_\infty(h) \right| dh = o_p(1)$, which implies

$$\begin{aligned}
\bar{h}^j \bar{h}^{j\prime} =& \int_{H_N} h p_N^{*j}(h)\, dh \int_{H_N} h' p_N^{*j}(h)\, dh \\
\xrightarrow{p}& \int_{R^d} h p_\infty(h)\, dh \int_{R^d} h' p_\infty(h)\, dh \\
=& \bar{h}\bar{h}',
\end{aligned}$$

and

$$\int_{H_N} h h' p_N^{*j}(h)\, dh \xrightarrow{p} \int_{R^d} h h' p_\infty(h)\, dh.$$

Therefore,

$$\begin{aligned}
\int_{H_N} \left(h - \bar{h}^j\right)\left(h - \bar{h}^j\right)' p_N^{*j}(h)\, dh \xrightarrow{p}& \int_{H_N} \left(h - \bar{h}\right)\left(h - \bar{h}\right)' p_N^{*}(h)\, dh \\
=& J^{-1}(\boldsymbol{\theta}_0) \\
=& -\nabla_{\theta\theta'} M(\boldsymbol{\theta}_0) \\
=& \Sigma_g.
\end{aligned}$$

That is, if $N\Delta_j \to 0$ as $N \to \infty$,

$$\int_{\boldsymbol{\Theta}} N\left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}^j\right)\left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}^j\right)' p_N^j(\boldsymbol{\theta})\, d\boldsymbol{\theta} = \Sigma_g + o_p(1).$$

B The Details of Estimation and Computation

B.1 The Computation of Bias and Root Mean Square Error

This subsection shows how to compute the bias and RMSE. Assume the true value of the target parameter x is x_0 and $\{\hat{x}^m\}_{m=1}^{M}$ is the set of estimates of x in M Monte Carlo replications. The bias is defined as

$$Bias(x) = \frac{1}{M}\sum_{m=1}^{M} \hat{x}^m - x_0.$$

The root mean square error is defined as

$$RMSE\left(x\right)=\sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(\hat{x}^{m}-x_{0}\right)^{2}}.$$

B.2 The Endogenous Grid Method for the Model (7)

The application of EGM for model (7) is documented in Algorithm B.1.

Algorithm B.1 The Endogenous Grid Method for Dynamic Model (7)

1: **Inputs:** Optimal consumption at period $t+1$, $c^{j}\left(\overrightarrow{m_{t+1}},\boldsymbol{z}_{t+1};\boldsymbol{\theta},\boldsymbol{\chi}\right)$ and the endogenous grid at period $t+1$, $\overrightarrow{m_{t+1}}$.

2: Form an exogenous ascending grid over end-of-period wealth at period t, denoted as $\overrightarrow{A_t}=\left\{A_t^k\right\}_{k=1}^{j}$, where $A_t^k>A_t^{k-1}$, $\forall k\in\{2,\ldots,j\}$.

3: **for** $k=1$ to j **do**

4: Compute $c_{i,t}^{k}=\left\{\beta_0 RE_{\varsigma_{t+1},\epsilon_{t+1},\boldsymbol{z}_{i,t+1}}\left[\frac{v(\boldsymbol{z}_{t+1};\boldsymbol{\eta}_0)}{v(\boldsymbol{z}_t;\boldsymbol{\eta}_0)}\left(G_{t+1}\varsigma_{t+1}\right)^{-\rho}c^{j}\left(m_{t+1}^{k},\boldsymbol{z}_{t+1};\boldsymbol{\theta},\boldsymbol{\chi}\right)\right]\right\}^{-\frac{1}{\rho}}$ with $m_{t+1}^{k}=\frac{RA_t^k}{G_{t+1}\varsigma_{t+1}}+\epsilon_{t+1}$.

5: Compute $m_t^k=c_t^k+A_t^k$.

6: **end for**

7: Store the endogenous grid. $\overrightarrow{m_t}=\left\{m_t^k\right\}_{k=1}^{j}$.

8: Store the corresponding optimal consumption at period t. $c^{j}\left(\overrightarrow{m_t},\boldsymbol{z}_t;\boldsymbol{\theta},\boldsymbol{\chi}\right)=\left\{c_t^k\right\}_{k=1}^{j}$

9: **Ouputs:** $c^{j}\left(\overrightarrow{m_t},\boldsymbol{z}_t;\boldsymbol{\theta},\boldsymbol{\chi}\right)$, $\overrightarrow{m_t}$.

Note:

(i) In Step 4, numerical method is used.

- $E_{\varsigma_{t+1},\epsilon_{t+1},\boldsymbol{z}_{i,t+1}}$ is the expectation with respect to $\varsigma_{t+1},\epsilon_{t+1}$ and $\boldsymbol{z}_{t+1}$. The expectation is numerically evaluated by using Gauss-Hermite quadrature method.
- The algorithm solves the model backwards, therefore $c^{j}\left(m_{t+1}^{k},\boldsymbol{z}_{t+1};\boldsymbol{\theta},\boldsymbol{\chi}\right)$ is the interpolated value of optimal consumption at period $t+1$ to approximate the income shocks.

(ii) During the EGM step, as in Carroll (2006), the credit constraints are dealt with by setting the smallest possible end-of-period resources A_t^1 equal 0. After operating the EGM, due to the monotonicity of saving, m_t^1 is the threshold value so that when $m_t<m_t^1$, the optimal consumption $c_t=m_t$.

B.3 The details of the estimation procedure for Section 5.1

During the estimation for the model (27), let $K_1=12800$, $K_2=3840$, $K=2560$, $\delta=0.5$ and the cutoff value $L=-10$. The number of grid to solve the model is 100. The perturbation variance is $\Sigma=diag\left(0.0001,0.04\right)$, where 0.0001 and 0.04 are for β and ρ, respectively. We use the case where $N^{obs}=3000$ for illustration.

Figure B.1: The particle points selected during the estimation

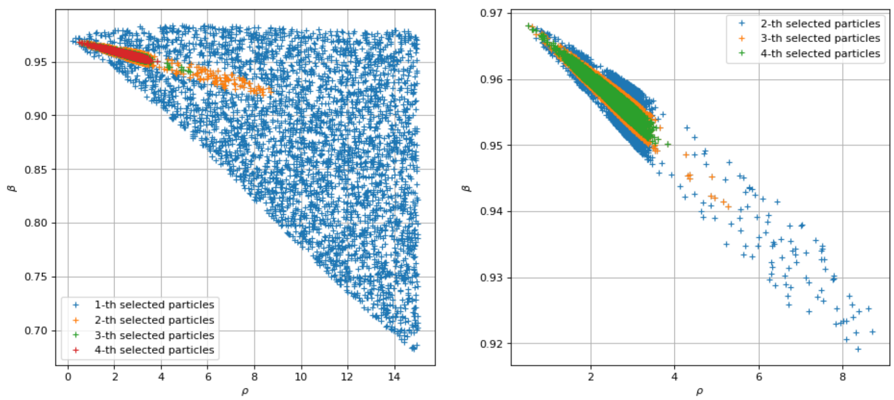


Figure B.2: The contour of the quasi-posterior density function and finally selected particle points

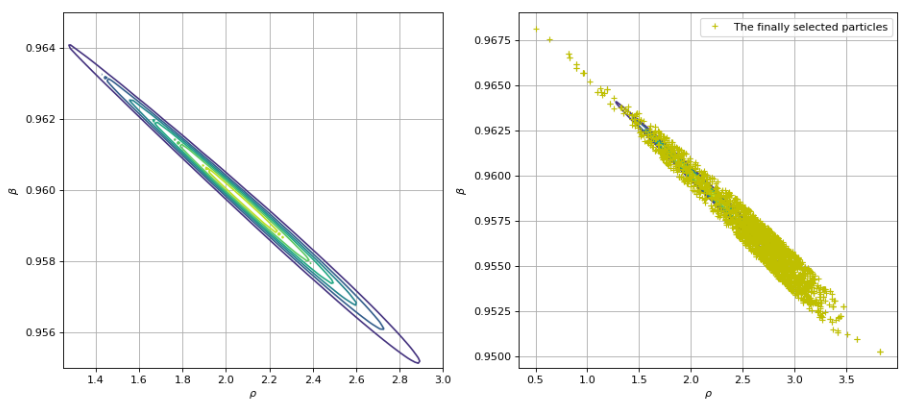
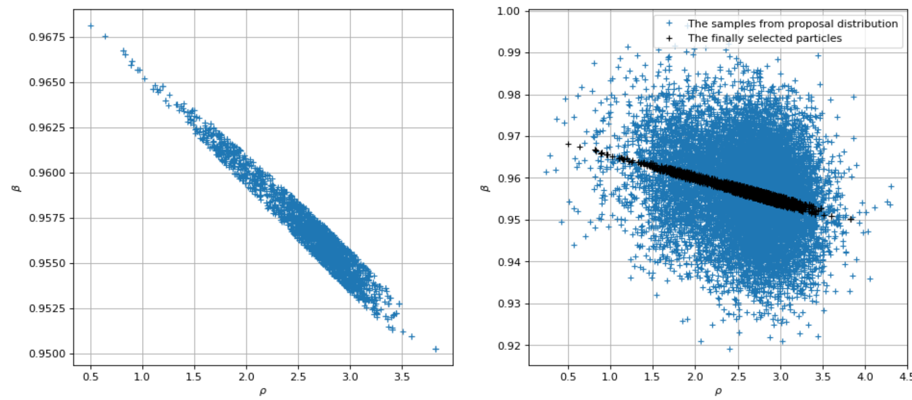


Figure B.1 plots the particle selected during the estimation procedure. As the process goes on, the area shrinks very quickly. The area of the first particle selection is wide but starting from the second selection, the area is very narrow. After the fourth particle points selection, we collect all the particles and select a subset of them based on the threshold value L. Afterwards, we uniformly choose K points from the subset. Based on these K selected particles, we construct a proposal distribution – a mixture normal distribution. At last, we draw K_3 samples from the proposal distribution.

The subset of particles and the contour of the quasi-posterior density are plotted in Figure B.2. The left panel is the contour plot and the right panel is the contour plot plus the subset of particles. We can readily find that the particles cover the area with significant density value quite well, which justifies that the proposal distribution is very close to the quasi-density function.

We can see from the left panel of Figure B.3. The area with significant weights is very narrow. The algorithm can identify the area quite accurately. After the final selection, we draw

Figure B.3: The finally selected particle points and samples from proposal distribution



K_3 samples from the proposal distribution. From the right panel of Figure B.3, we can find that the finally selected particles are almost covered by the samples from the proposal distribution.

B.4 The Kalman Filter for the Income Process

When there is not income shock, we have

$$z_{it} = \mathfrak{U} + \mathfrak{B}x_{it} + \mathfrak{y}_{it},$$

$$x_{it} = \mathfrak{C}_t + \mathfrak{D}x_{it-1} + \mathfrak{u}_{it},$$

where $\mathfrak{U} = 0$, $\mathfrak{B} = 1$, $z_{it} = \log Y_{it}$, $x_{it} = \log P_{it}$, $\mathfrak{y}_{it} = \log \varepsilon_{it} \sim N\left(0, \sigma_\varepsilon^2\right)$, $\mathfrak{C}_t = \log G_t$, $\mathfrak{D} = 1$, $\mathfrak{u}_{it} = \log \varsigma_{it} \sim N\left(0, \sigma_\varsigma^2\right)$. According to the dataset, Y_{it} is observed household income, G_t , σ_ε^2 and σ_ς^2 are known. The permanent income component P_{it} is the one that we want to recover. In the following, the subscripts i is suppressed.

The Kalman filter consists of following three steps. Since the error terms are all normal and the structure is linear, all the variables in the system are normal distributed. Thus we only need to filter the mean and variance. Initialize the mean and variance at the beginning, $\mu_{0|0} = E\left[x_0|F_0\right]$, $\Sigma_{0|0} = Var\left(x_0|F_0\right)$, where F_0 is the information set known at time 0. Later the details of initialization is discussed.

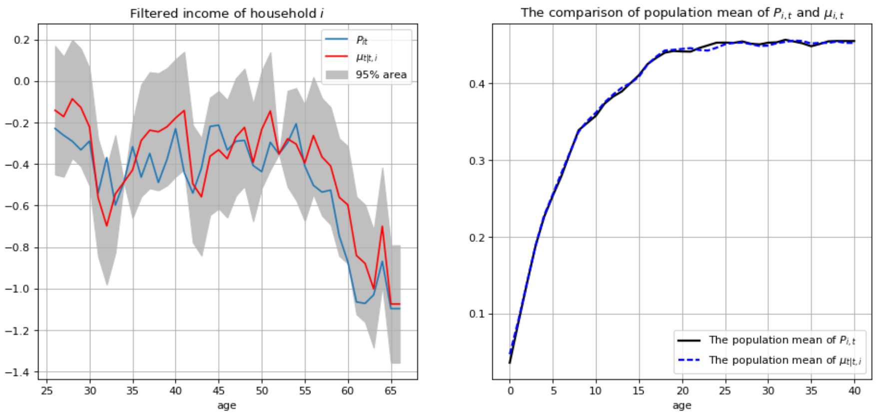
- Initialize $\mu_{0|0}$ and $\Sigma_{0|0}$. At the beginning of time t, we have $\mu_{t-1|t-1}$, $\Sigma_{t-1|t-1}$.
- One-step-ahead predictive distribution of $x_t|F_{t-1} \sim N\left(\mu_{t|t-1}, \Sigma_{t|t-1}\right)$:

$$\begin{aligned}\mu_{t|t-1} \equiv E\left[x_t|F_{t-1}\right] =& E\left[\mathfrak{C}_t + \mathfrak{D}x_{it-1}|F_{t-1}\right] \\ =& \mathfrak{C}_t + \mathfrak{D}\mu_{t-1|t-1},\end{aligned}$$

$$\begin{aligned}\Sigma_{t|t-1} \equiv Var\left[x_t|F_{t-1}\right] =& E\left[Var\left(x_t|F_{t-1}\right)|F_{t-1}\right] + Var\left[E\left(x_t|F_{t-1}\right)|F_{t-1}\right] \\ =& \sigma_\varsigma^2 + \mathfrak{D}^2\Sigma_{t-1|t-1},\end{aligned}$$

where F_t denotes the information known up to time t.

Figure B.4: The performance of income filter



- One-step-ahead predictive distribution of $z_t|F_{t-1} \sim N\left(f_{t|t-1}, Q_{t|t-1}\right)$:

$$\begin{aligned} f_{t|t-1} \equiv E\left[z_t|F_{t-1}\right] =& E\left\{E\left[z_t|x_t, F_{t-1}\right]|F_{t-1}\right\} \\ =& \mathfrak{U} + \mathfrak{B}\mu_{t|t-1}, \end{aligned}$$

$$\begin{aligned} Q_{t|t-1} \equiv Var\left[z_t|F_{t-1}\right] =& E\left[Var\left(z_t|F_{t-1}\right)|F_{t-1}\right] + Var\left[E\left(z_t|F_{t-1}\right)|F_{t-1}\right] \\ =& \sigma_\varepsilon^2 + \mathfrak{B}^2\Sigma_{t|t-1}. \end{aligned}$$

- The filtering distribution of x_t given F_t. $x_t|F_t \sim N\left(\mu_{t|t}, \Sigma_{t|t}\right)$:

$$\mu_{t|t} = \mu_{t|t-1} + \Sigma_{t|t-1}\mathfrak{B}Q_{t|t-1}^{-1}\left(z_t - f_{t|t-1}\right),$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1}\mathfrak{B}^2Q_{t|t-1}^{-1}\Sigma_{t|t-1}.$$

If $p > 0$, and μ is very closed to 0. We can use some threshold value to judge whether there is a shock or not. Once the shock is in presence at any time t, $\log P_t = \log Y_t - \log\mu$, in which case P_t can be directly recovered. Thus, we can set $\mu_{t|t} = \log Y_t - \log\mu$ and $\Sigma_{t|t} = 0$. Otherwise if $p > 0$, and $\mu = 0$, the income here can be treated as missing.

For the values of $\mu_{0|0} = E\left[x_0|F_0\right]$, $\Sigma_{0|0} = Var\left(x_0|F_0\right)$, since $\log Y_{it} = \log P_{it} + \log\varepsilon_{it}$, we simply assume for each household i, the initial value $\mu_{0|0} = \log Y_0 - p\mu$, where $\log Y_0$ is the population mean of income level at time 0, and accordingly $\Sigma_{0|0} = \sigma_\varepsilon^2$.

Figure B.4 reports the performance of the income filter where $G_{26:29} = 1.05$, $G_{30:35} = 1.03$, $G_{36:45} = 1.01$, $G_{46:65} = 1$, $T_r = 65$, $p = 0.03$, $\mu = 10^{-6}$, $\sigma_\varsigma^2 = 0.02$, $\sigma_\epsilon^2 = 0.04$. From the left panel, the 95% area centering at the filtered mean $\mu_{t|t,i}$ and bounded by $\pm 2\Sigma_{t|t}$ can cover $P_{i,t}$ at majority of the life time. Further, the right panel shows that the difference between the population means of $\mu_{t|t}$ and $P_{i,t}$ are quite small.

B.5 The Comparison of Different Computations for Optimal Consumption

Here the second example in the Monte Carlo study section is used with $T_r = 65$ to compare the performance of different computation methods for the optimal consumption level $C_{i,t}$ for household i at age t. One is to simulate numerous income sample paths and compute the optimal consumption at every path at every age. At each age we collect the consumptions of all households and compute sample mean. This is the approach proposed by GP. We call it as 'GP' and it can be expressed by

$$C_{i,t}^{GP} = E\left[c_t\left(\frac{M_{i,t}^d}{P_{i,t}}\right)P_{i,t}\right] = \frac{1}{G}\sum_{g=1}^{G} c_t\left(\frac{M_{i,t}^d}{P_{i,t}^{(g)}}\right)P_{i,t}^{(g)},\ \text{for each } i,t,$$

$$E_i\left\{E\left[c_t\left(\frac{M_{i,t}^d}{P_{i,t}}\right)P_{i,t}\right]\right\} = \frac{1}{N_t}\sum_{i=1}^{N_t} G_{i,t}^{GP},$$

where $\left\{P^{(g)}\right\}_{t=26}^{66}$ is the permanent income component from $t = 26$ to $t = 66$ at g^{th} simulated income path.

The other is to treat the filtered mean $\mu_{t|t,i}$ from the Kalman income filter, as $\log P_{i,t}$, which is used by Jørgensen (2017). We call this approach as 'J' and it can also expressed by

$$C_{i,t}^{J} = c_t\left(\frac{M_{i,t}^d}{\mu_{t|t,i}}\right)\mu_{t|t,i},\ \text{for each } i,t,$$

$$E_i\left[c_t\left(\frac{M_{i,t}^d}{\mu_{t|t,i}}\right)\mu_{t|t,i}\right] = \frac{1}{N_t}\sum_{i=1}^{N_t} G_{i,t}^{J},$$

The proposed approach in equation (9) is denoted as 'L'. Given $N^{obs} = 1500$, we compare these three computation approaches, which is reported in Figure B.5. The number of simulated paths for 'GP' is 1000. From the following figures, it is obvious 'GP' does not approximate the population mean of consumption profile quite well even when sample path is 1000. 'J' is close to the population mean, similar to 'L'.

For further comparison, we use the following statistics to compare the three approaches,

$$dist = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N} C_{i,t}^{d} - \frac{1}{N}\sum_{i=1}^{N} C_{i,t}^{\mathfrak{a}}\right)^2},\ \mathfrak{a} = GP, J, L.$$

The values of the statistics are reported in Table B.1. It is apparent that 'L' has the smallest distance from the population mean of consumption profile in all cases. As the sample size increases, the distance of 'L' decreases dramatically. But the other two approaches remains the same magnitudes.

Besides, we change the value of ρ into 0.5, which is the same as GP. Following Figure B.5, we draw the corresponding figures in Figure B.6 which shows that 'L' is better.

Figure B.5: The computed consumption profiles when $N^{obs} = 1500, \rho = 2$

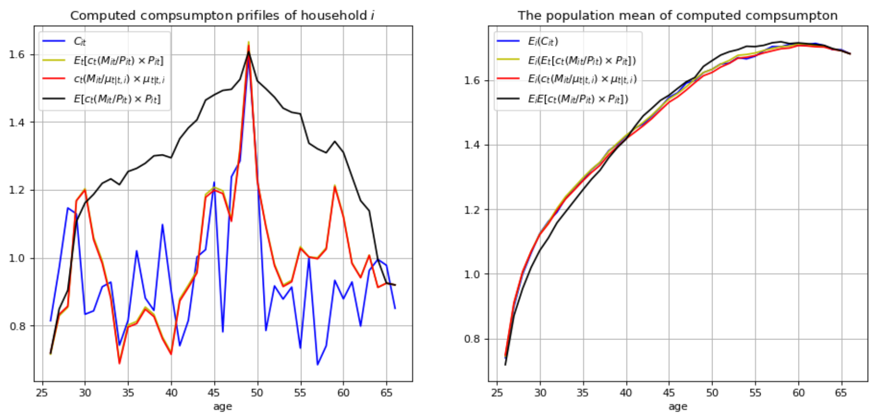


Table B.1: The values of the statistics for three approaches

	GP	J	L
$N^{obs} = 1500$	6.5382×10^{-4}	7.2569×10^{-5}	2.8096×10^{-5}
$N^{obs} = 3000$	6.8139×10^{-4}	7.1134×10^{-5}	1.3468×10^{-5}
$N^{obs} = 6000$	2.4233×10^{-3}	6.6381×10^{-5}	9.9386×10^{-6}

Figure B.6: The computed consumption profiles when $N^{obs} = 6000, \rho = 0.5$