Model Specification Tests of Heterogenous Agent Models with Aggregate Shocks under Partial Information*

Zongwu Cai† Hongwei Mei‡ and Rui Wang§

February 28, 2024

Abstract

For a heterogeneous agent model with aggregate shocks, the seminal paper by Krusell and Smith (1998) provides an equilibrium framework depending only on the (conditional) mean wealth rather than the wealth distribution of all agents, which is referred to as *approximate aggregation* for their prototype model. Their result can be obtained through the analysis of a forward-backward system consisting of the Hamilton-Jacobi-Bellman equation, the Fokker-Planck equation, and some constraint. Different from the existing literature, this paper proposes a statistical method to verify whether a heterogeneous agent model features approximate aggregation in the scenario that only one agent's wealth together with the aggregate shocks is observable over time. Our main approach lies in studying a model specification testing problem for the evolution of the wealth (i.e. the Fokker-Planck equation) in some appropriate parametric family featuring approximate aggregation. The key challenge stems from the partially observed information where the wealth distribution of all agents is infeasible. To overcome this difficulty, first, a novel two-step estimate is proposed for estimating the parameter in the parametric family. Then, several testing statistics are constructed and their asymptotic properties are established, which in turn provides several testing rules. Finally, some Monte Carlo simulations are conducted to illustrate the finite sample performance of the proposed tests.

Keywords: Heterogeneous agent model with aggregate shocks, Approximate aggregation; Model specification test; Equilibrium estimator; Partial observation.

JEL Classification: C12, C13, E20

*This work is partially supported by Simons Foundation's Travel Support for Mathematicians Program under Grant Number 00002835 and the Natural Science Foundation of China under Grant Numbers 72133002 and 72033008.

†Department of Economics, The University of Kansas, Lawrence, KS 66045, caiz@ku.edu,

‡Department of Mathematics and Statistics, Texas Tech University, Lubbock, TX 79409, hongwei.mei@ttu.edu.

§Department of Economics, The University of Kansas, Lawrence, KS 66045. rui.wang@ku.edu.

1 Introduction

The heterogeneous agent model (HAM) has been one of the key developments in macroeconomics in the last four decades since Bewley (1986), Huggett (1993), and Aiyagari (1994), which is called Aiyagari-Bewley-Huggett (ABH) model in the macroeconomics literature. Compared to the general equilibrium macroeconomic theory, where it is assumed that the economy behaves *as if* it is inhabited by a single (type of) consumer, the HAM is introduced to extend the standard macroeconomic model to include substantial heterogeneity in income and wealth. Such a model allows macroeconomists to empirically discipline macroeconomic theories from micro-data. Generally speaking, in the HAM, each agent aims to find a strategy to maximize a reward subject to the corresponding wealth process under some constraints. More specifically, the HAM can be formulated as a forward-back system consisting of three perspectives: (1) the backward Hamilton-Jacobi-Bellman (HJB) equation to determine the (optimal) decision rule; (2) the forward Fokker-Planck (FP) equation describing the evolution of agent's wealth; (3) the constraint restricting the admissible actions. If the HJB-FP system is solved, a corresponding decision rule, namely an equilibrium, can be obtained simultaneously.

As a benchmark model, Aiyagari (1994) initiated a computational scheme that the infinite-dimensionality of the heterogeneity can be reduced to some finite moments only so that its dynamic can be computed numerically. Based on this spirit, Krusell and Smith (1998) generalized the model to the case that aggregate shocks exist. Also, Krusell and Smith (1998) pointed out that "our main finding is that, in the stationary stochastic equilibrium, the behavior of the macroeconomic aggregates can be almost perfectly described by using only the mean of the wealth distribution", and "by approximate aggregation, we mean that, in equilibrium, all aggregate variables' consumption, the capital stock, and relative prices can be almost perfectly described as a function of two simple statistics: the mean of the wealth distribution and the aggregate productivity shock". To summarize, the key of their results is that the equilibrium depends on the total wealth distribution through its (conditional) mean only, which is referred to as approximate aggregation. It is worth mentioning that Krusell and Smith (1998) obtained this result for their prototype model, which is essentially a forward-backward HJB-FP system as mentioned above. The current paper is motivated by answering the following question: Is it possible to verify the approximate aggregation property for the HAM statistically if only one agent's wealth together with the shocks is observable over time (i.e. partial information)? More explicitly, in this paper, the HAM is investigated in the following scenario: The HAM has attained its stationary equilibrium over an infinite time horizon, and only the wealth of one particular agent together with the aggregate shocks is observable to us along time. To tackle the problem, we study a model specification testing problem whether the wealth evolution (i.e. the FP equation) falls into some parametric family $\mathcal{P}$ featuring the approximate aggregation. Once the $\mathcal{P}$ is specified, the approximate aggregation is verified. Moreover, with the specified model, one can further study the dynamic behavior of the total wealth distribution.

A comprehensive review on the early developments for the HAM can be found in the papers by, for instance, Rios-Rul and Rios-Rull (1995), Ríos-Rull and José-Víctor (2001), Hommes (2006), Heathcote et al. (2009), and references therein. More recently, there have been several new developments in monetary and fiscal policies and their distributional implications, see, for example, Hedlund et al. (2017) and Kaplan et al. (2018). Moreover, Achdou et al. (2022) provided a continuous-time approach for heterogeneous agent models by focusing on income and wealth distribution. On the other hand, several econometric methods are used to examine the quantitative properties of the HAM. For example, several efforts are made by Benhabib et al. (2019) and Abbott et al. (2019) with their focus on HAM without aggregate shocks. For the HAM with aggregate shocks using full information, the reader is referred to the recent papers by Reiter (2009), Winberry (2018), Mongey and Williams (2017), Williams (2017), Han et al. (2021), Parra-Alvarez et al. (2023), Liu and Plagborg-Moller (2023) and references therein. Furthermore, it is well known that the HAM can be cast in terms of "Mean Field Games" as in Achdou et al. (2022), which has been intensively studied in mathematics since the seminal works were completed independently by Huang et al. (2003, 2007), and Lasry and Lions (2006a,b).

Here, it is worth to mention that the recent paper by Parra-Alvarez et al. (2023) provides the maximum likelihood (ML) estimator for the benchmark ABH model without involving aggregate shocks. Under their framework, the mean wealth is constant in the stationary equilibrium and the maximum likelihood estimator can be computed with the full information on the income (see equation (18) therein). While for the HAM with aggregate shocks, the mean wealth turns out to be a conditional mean (on the aggregate shocks), which is not constant and is unobservable. Therefore, their method is not directly applicable to our framework. Instead, a two-step method is proposed to estimating the parameters, where the ML estimator is one step essentially. Moreover, we do not restrict the prototype model to ABH model in our paper either. Therefore, our newly proposed method generalizes the maximum likelihood method proposed in Parra-Alvarez et al. (2023), when the aggregate shocks are involved and the prototype model is possibly latent.

Suppose that a sequence of observations $\{(X_t, Y_t) : t = 1, \cdots, T\}$ are observed from in the HAM with aggregate shocks where X_t is an agent's wealth process over time and Y_t is the shock process. It is also assumed that the system has attained the stationary equilibrium over an infinite horizon (for more about the definition of the stationary equilibrium, one is referred to the papers by Achdou et al. (2022) and Parra-Alvarez et al. (2023)). In this paper, to tackle the 'approximate aggregation', our goal is to test the following parametric family

$$\mathcal{P} := \Big\{\mathbf{b}(x, y, u) = \bar{b}(x, y, u; \theta),\ \boldsymbol{\sigma}(x, y, u) = \bar{\sigma}(x, y, u; \sigma) : \theta \in \mathbf{\Theta}, \sigma \in \mathbf{\Sigma}\Big\},$$

where both $\bar{b}(\cdot)$ and $\bar{\sigma}(\cdot)$ are known functions with unknown parameters θ and σ, such that the observations $\{(X_t, Y_t) : t = 1, \cdots, T\}$ satisfy the following Euler equation (discrete-time version)

$$X_{t+1} - X_t = \mathbf{b}(X_t, Y_t, U_t)\Delta + \boldsymbol{\sigma}(X_t, Y_t, U_t)\sqrt{\Delta}w_t, \tag{1}$$

with a step-constant $\Delta > 0$, where w_t is the unobservable (stationary) income process, and $(\mathbf{\Theta}, \mathbf{\Sigma})$ is the collection of all admissible parameters (θ, σ) to be specified later. The (conditional) mean wealth

$$U_t = \mathbb{E}[X_t | Y_1, \cdots, Y_{t-1}]$$

for $t \geq 1$ which is the average wealth of all agents, is unobservable to us too. In fact, (1) is parallel to the nonlinear system to characterize the model's aggregate equilibrium by adopting the distributional approximation (see (35) in Reiter (2009) or (1) in Liu and Plagborg-Moller (2023)). For a special linear case when $\bar{b}(x, y, u; \theta) = A_0(\theta)x + A_1(\theta)u + B(\theta)Y_t$, it follows that

$$U_{t+1} = (A_0(\theta) + A_1(\theta) + 1)U_t + B(\theta)Y_t,$$

which corresponds to the linear transition equation for the aggregate state in heterogeneous household model (see (37) in Reiter (2009) or (2) in Liu and Plagborg-Moller (2023)). The key feature of (1) is that the evolution of one agent's wealth process depends on the wealth distribution through its (conditional) mean only, which is consistent to the approximate aggregation property. Once the parametric family is specified, the approximate aggregation property is verified.

It is also worth to note that (1) can be generalized to a continuous-time stochastic differential equation (SDE),

$$d\,X_t = \mathbf{b}(X_t, Y_t, U_t)dt + \boldsymbol{\sigma}(X_t, Y_t, U_t)d\,W_t,$$

which is an extension of the classical Black-Scholes model to include the shock process Y_t and the mean wealth U_t in the dynamic system. Throughout the paper, the true parameters are denoted by (θ_*, σ_*) if H_0 is true even though their values may not be given.

The model specification problem for conventional stochastic diffusions has been well studied in the literature since the pioneer work by Ait-Sahalia (1996). There are some extensions to the method proposed in Ait-Sahalia et al. (2009), Hong and Li (2005), Chen et al. (2008), Ait-Sahalia et al. (2009), and others. There also exist several works on goodness-of-fit testing problems for continuous-time stochastic diffusions. For example, one may refer to the papers by Dachian and Kutoyants (2008), Negri and Nishiyama (2009), Kleptsyna and Kutoyants (2014), López-Pérez et al. (2022), and references therein. Different from the above mentioned literature, the system in our paper is partially observed and those approaches proposed cannot be directly applied to our setting due to the lack of observations on the mean-wealth process U.

To overcome this challenge, a novel two-step verification procedure is proposed to obtaining an appropriate estimator for θ in the parametric family $\mathcal{P}$. Such an estimator is called by *equilibrium estimator* in our paper. To the best of our knowledge, the equilibrium estimator is new in the literature, which constitutes of the paper's main contribution in part. With the help of the equilibrium estimator, we can then construct several testing statistics and establish their asymptotic properties, which yield some appropriate testing procedures directly.

The rest of the paper is arranged as follows. Section 2 introduces the equilibrium estimator for general cases and its asymptotic property is investigated. With the help of the equilibrium estimator, the model specification testing is constructed in Section 3, together with a Bootstrap procedure for estimating the critical value. Some simulation studies are conducted in Section 4 to investigate the finite sample performance of our proposed testing procedure. Moreover, Section 5 concludes the paper. Finally, the mathematical proofs together with necessary lemmas are provided in Section 6.

2 Equilibrium Estimator

2.1 Estimation Procedure

In this subsection, we first introduce the method of the equilibrium estimator for a general setting. Then, we focus on some special cases in the asymptotic properties. As mentioned earlier, the key difficulty lies in the missing information on U_t. Our solution is to construct an auxiliary process (denoted by $\widetilde{U}(\vartheta, \varsigma)$ later) to approximate U_t. It is worth to mention that (X, Y, U) is not a Markov process in general, and U_t can not be constructed from all the information $\{(X_s, Y_s, U_s) : s \leq t-1\}$.

To gain the Markovian property, define Ξ_t by the distribution of X_t conditional on the aggregate shocks up to time $t-1$, i.e. $\{Y_1, \cdots, Y_{t-1}\}$, in the sense that for any bounded continuous function $h(\cdot)$,

$$\mathbb{E}\big[h(X_t)|Y_1, \cdots, Y_{t-1}\big] = \int_{\mathbb{R}} h(x)\Xi_t(dx).$$

It is straightforward to see that (X, Y, Ξ) is a Markov process with

$$U_t = \int_{\mathbb{R}} x\Xi_t(dx).$$

Moreover, Ξ_{t+1} satisfies the following equation

$$\int_{\mathbb{R}} h(x)\Xi_{t+1}(dx) = \int_{\mathbb{R}} h\Big(x + \bar{b}(x, Y_t, U_t; \theta_*)\Delta + \bar{\sigma}(x, Y_t, U_t; \sigma_*)\sqrt{\Delta}w\Big)\phi(dw)\Xi_t(dx) \tag{2}$$

for any bounded continuous function $h(\cdot)$, where $\phi(\cdot)$ is the distribution function of w_t given in (1). From (2), Ξ_t can be calculated explicitly, depending on the observable $(Y_1, \cdots, Y_{t-1})$, if Ξ_1 and (θ_*, σ_*) are given. In fact, (2) is parallel to the FP equation.

Inspired from above, given the observation of $\{Y_t\}$, we introduce the following auxiliary process $(\widetilde{\Xi}_t(\vartheta, \varsigma), \widetilde{U}_t(\vartheta, \varsigma))$ such that

$$\int_{\mathbb{R}} h(x)\widetilde{\Xi}_{t+1}(dx; \vartheta, \varsigma) = \int_{\mathbb{R}} h\Big(x + \bar{b}(x, Y_t, \widetilde{U}_t(\vartheta, \varsigma); \vartheta)\Delta + \bar{\sigma}(x, Y_t, \widetilde{U}_t(\vartheta, \varsigma); \varsigma)\sqrt{\Delta}w\Big)\phi(dw)\widetilde{\Xi}_t(dx; \vartheta, \varsigma)$$

for any bounded continuous function $h(\cdot)$, and

$$\widetilde{U}_t(\vartheta, \varsigma) = \int_{\mathbb{R}} x\widetilde{\Xi}_t(dx; \vartheta, \varsigma).$$

Such a construction of $(\widetilde{\Xi}_t(\vartheta, \varsigma), \widetilde{U}_t(\vartheta, \varsigma))$ depends on the parameter (ϑ, ς) and the selection of the initial value $\widetilde{\Xi}_1$. Due to the ergodicity, it is natural to see that the construction is independent of $\widetilde{\Xi}_1$ if t is large and thus, such dependence in the notations is omitted for the sake of convenience. Now, it is ready to present our equilibrium estimator $(\hat{\theta}_T, \hat{\sigma}_T)$, which consists of two steps, described as follows.

Step 1: Maximum likelihood method. Given $\mathcal{P}$, let

$$q(x, y, u, \tilde{x}; \theta, \sigma)d\tilde{x} := \mathbb{P}\Big(X_2 \in d\tilde{x}\Big| X_1 = x, Y_1 = y, U_1 = u; \theta, \sigma\Big).$$

If U_t was observable, it is straightforward to derive a maximum likelihood estimator

$$(\hat{\theta}_{\text{MLE}}, \hat{\sigma}_{\text{MLE}}) = \operatorname*{argmax}_{\theta \in \mathbf{\Theta}, \sigma \in \mathbf{\Sigma}} \sum_{t=1}^{T} \log q(X_t, Y_t, U_t, X_{t+1}; \theta, \sigma), \tag{3}$$

if it exists. In the sequel, for simplicity, it is assumed that w_t is a normal random variable such that the MLE in (3) admits an explicit form. Here, it should be emphasized that the idea of equilibrium estimator is applicable to other income processes w. If the distribution is not close to normal, then, the above MLE is the pseudo-MLE.

Step 2: Fixed-point procedure. Due to the lack of U_t in the observation, replacing U_t by $\widetilde{U}_t(\vartheta,\varsigma)$ in (3) leads to a random function $(\widehat{\Theta}_T(\cdot),\widehat{\Sigma}(\cdot)) : \mathbf{\Theta}\times\mathbf{\Sigma}\mapsto\mathbf{\Theta}\times\mathbf{\Sigma}$ defined by

$$(\widehat{\Theta}_T(\vartheta,\varsigma),\widehat{\Sigma}(\vartheta,\varsigma)) = \operatorname*{argmax}_{\theta\in\mathbf{\Theta},\sigma\in\mathbf{\Sigma}}\sum_{i=1}^{T}\log q(X_t,Y_t,\widetilde{U}_t(\vartheta,\varsigma),X_{t+1};\theta,\sigma).$$

Observe that $\widehat{\Theta}_T(\theta_*,\sigma_*)$ converges to θ_* and $\widehat{\Sigma}_T(\theta_*,\sigma_*)$ converges to σ_* as $T\to\infty$. Therefore, it is natural to expect that the estimator $(\hat{\theta}_T,\hat{\sigma}_T)$ can be defined as the fixed point of $(\widehat{\Theta}_T(\cdot),\widehat{\Sigma}_T(\cdot))$ in $\mathbf{\Theta}\times\mathbf{\Sigma}$. As the fixed-point may not exist, define the equilibrium estimator $\hat{\theta}_T$ by

$$(\hat{\theta}_T,\hat{\sigma}_T) := \operatorname*{argmin}_{\vartheta\in\mathbf{\Theta},\varsigma\in\mathbf{\Sigma}}\Big(|\widehat{\Theta}_T(\vartheta,\varsigma)-\vartheta| + |\widehat{\Sigma}_T(\vartheta,\varsigma)-\varsigma|\Big). \tag{4}$$

At the same time, the minimum error is denoted by $\hat{\gamma}_T$. If $\mathcal{P}$ is true, it should hold that $\hat{\gamma}_T = 0$ when T is large, i.e. $(\hat{\theta}_T,\hat{\sigma}_T)$ is the true fixed point. Such a result is critical when deriving the asymptotic normality of $\hat{\theta}_T$ under the null hypothesis. While if $\mathcal{P}$ is not true, it is natural to expect that the minimum point is not a fixed point (i.e. $\hat{\gamma}_T$ is far away from 0). In this scenario, $\mathcal{P}$ should be rejected. Therefore, it is reasonable that our testing procedure also involves $\hat{\gamma}_T$ in the future.

Until now, the equilibrium estimator is defined for general cases. Our next goal is to establish its asymptotic properties, which are necessary for the model specification tests. To achieve this, the main focus in this paper is on the following special case.

Assumption 1. (1) The income processes $\{w_t\}$ are i.i.d standard normal random variables, and Y_t is a Markov process. The parametric family $\mathcal{P}$ satisfies

$$\mathcal{P} := \Big\{\mathbf{b}(x,y,u) = b_0(x,y,u) + \langle\theta, b(x,y,u)\rangle,\ \boldsymbol{\sigma}(x,y,u) = \sigma * \sigma_0(x,y) : (\theta,\sigma)\in\mathbf{\Theta}\times\mathbb{R}_+\Big\},$$

where $\theta = (\theta_1,\cdots,\theta_m)$ lies in a compact subset $\mathbf{\Theta}$ of $\mathbb{R}^m$ and both functions $b_0(\cdot)$ and $\sigma_0(\cdot)$ are known.

(2) The drift function $\mathbf{b}(x,y,u)$ satisfies $b_i(x,y,u) = A_i(y)x + B_i(y,u)$ for $i = 0,1,\cdots,m$.

(3) The true θ_* is an interior point in $\mathbf{\Theta}$.

Some remarks are needed for the above assumption. First, if the prototype model is known to us, the drift coefficients b can be chosen from the prototype model directly. If the prototype model is latent, the drift coefficient vector b can be seen as a basis such that the true drift coefficient can be approximated by $\mathbf{b}(x, y, u)$. Second, we assume that $b_i(x, y, u)$ is linear in x so that (X_t, Y_t, U_t) is a Markovian system, which is consistent with the approximate aggregation feature. Finally, the assumption θ_* being an interior point in $\mathbf{\Theta}$ is a general assumption in the model specification problems. In fact, our results can be extended to the case that $\mathbf{\Theta} = \{\theta_*\}$ without any essential difficulties.

Under the above assumption, it is easy see that the definition of $\widetilde{U}_t(\vartheta, \varsigma)$ is independent of ς and the two-step verification procedure applies to $\hat{\theta}_T$ only. Therefore, we only write $\widetilde{U}_t(\vartheta)$ instead of $\widetilde{U}_t(\vartheta, \varsigma)$, in what follows. After $\hat{\theta}_T$ is achieved, $\hat{\sigma}_T$ can be achieved by the MLE step directly. Moreover, the equilibrium estimator admits a closed form, which will be given explicitly later.

In addition to Assumption 1, another technical assumption (Assumption 2) is assumed as well in Section 6 concerning the ergodicity of the HAM system. With all those assumptions, it is ready to study the asymptotic properties of the proposed equilibrium estimator in the next subsection.

2.2 Asymptotic Theory

In this subsection, our aim is to study the asymptotic theory for our equilibrium estimator $(\hat{\theta}_T, \hat{\sigma}_T)$: consistency and asymptotic normality.

Because of the linearity of $b_i(\cdot)$ for all i in x assumed in Assumption 1, it is easy to observe that $\widetilde{U}_t(\vartheta)$ can be computed through the following recursion directly without introducing the conditional distribution $\widetilde{\Xi}_t(\vartheta)$ and ς:

$$\widetilde{U}_{t+1}(\vartheta) = \widetilde{U}_t(\vartheta) + b_0(\widetilde{U}_t(\vartheta), Y_t, \widetilde{U}_t(\vartheta))\Delta + \langle \vartheta, b(\widetilde{U}_t(\vartheta), Y_t, \widetilde{U}_t(\vartheta))\rangle \Delta \text{ with } \widetilde{U}_1(\vartheta) = X_1. \quad (5)$$

Here, we take the feasible initial value $\widetilde{U}_1(\vartheta) = X_1$ because the true initial U_1 is not observable. We also want to emphasize that such a construction is independent of the parameter σ. Therefore, ς omitted in the function $\widehat{\Theta}_T(\vartheta, \varsigma)$ and the function $\widehat{\Sigma}_T(\vartheta, \varsigma)$ is not needed. More specifically, the equilibrium estimator can be rewritten as

$$\hat{\theta}_T = \underset{\vartheta \in \mathbf{\Theta}}{\operatorname{argmin}} \, |\widehat{\Theta}_T(\vartheta) - \vartheta|,$$

and

$$\hat{\sigma}_T^2 = \frac{1}{T}\sum_{t=1}^{T} \frac{[X_{t+1} - X_t - \Delta b_0(X_t, Y_t, \widetilde{U}_t(\hat{\theta}_T)) - \Delta\langle \hat{\theta}_T, b(X_t, Y_t, \widetilde{U}_t(\hat{\theta}_T))\rangle]^2}{\sigma_0^2(X_t, Y_t)};$$

where

$$\widehat{\Theta}_T(\cdot) = \widehat{M}_T^{-1}(\vartheta)\widehat{\Phi}_T(\vartheta), \quad \widehat{M}_T(\vartheta) = \frac{1}{T}\sum_{t=1}^{T} \frac{b(X_t, Y_t, \widetilde{U}_t(\vartheta))b^\top(X_t, Y_t, \widetilde{U}_t(\vartheta))}{\sigma_0^2(X_t, Y_t)}, \tag{6}$$

and

$$\widehat{\Phi}_T(\vartheta) = \frac{1}{T\Delta}\sum_{t=1}^{T} \Big[\frac{X_{t+1} - X_t - \Delta b_0(X_t, Y_t, \widetilde{U}_t(\vartheta))}{\sigma_0^2(X_t, Y_t)} b(X_t, Y_t, \widetilde{U}_t(\vartheta))\Big]. \tag{7}$$

To study the asymptotic properties of our equilibrium estimator, by the ergodicity of $(X_t, Y_t, U_t, \widetilde{U}(\vartheta))$ (see Assumption 2 listed in Section 6), it is expected that there exists a triple $(\Theta(\vartheta), M(\vartheta), \Phi(\vartheta))$ such that by (4), (6) and (7), as $T \to \infty$,

$$(\widehat{\Theta}_T(\vartheta), \widehat{M}_T(\vartheta), \widehat{\Phi}_T(\vartheta)) \;\; \to \;\; (\Theta(\vartheta), M(\vartheta), \Phi(\vartheta))$$

almost surely. Write $\mu(dx, dy, du, d\tilde{u}; \vartheta)$ as the ergodic measure of $(X, Y, U, \widetilde{U}(\vartheta))$. Then, one has the following

$$\Theta(\vartheta) = M(\vartheta)^{-1}\Phi(\vartheta)$$

with

$$M(\vartheta) = \int \frac{b(x, y, \tilde{u})b^\top(x, y, \tilde{u})}{\sigma^2(x, y)} \mu(dx, dy, du, d\tilde{u}; \vartheta),$$

and

$$\Phi(\vartheta) = \int \frac{b(x, y, \tilde{u})\big[b_0(x, y, u) - b_0(x, y, \tilde{u}) + \langle \theta_*, b(x, y, u)\rangle\big]}{\sigma^2(x, y)} \mu(dx, dy, du, d\tilde{u}; \vartheta).$$

Define $\partial\Theta(\cdot)$ by the gradient field of $\Theta(\cdot)$. Write $H_* = H(\theta_*)$ for $H = \Theta, M, \Phi, \partial\Theta$. Now, it is ready to present the asymptotic properties of our equilibrium estimators with their proofs provided in Section 6.

Theorem 1 (Consistency)**.** *Suppose that Assumptions 1 and 2 hold, and $\mathcal{P}$ is true. As $T \to \infty$, it follows that*

$$(\hat{\theta}_T, \hat{\sigma}_T^2) \;\; \longrightarrow \;\; (\theta_*, \sigma_*^2),$$

almost surely.

Theorem 2 (Asymptotic Normality). *Suppose that Assumptions 1 and 2 hold, and $\mathcal{P}$ is true. As $T \to \infty$, it follows that*

$$\sqrt{T\Delta}\begin{pmatrix} \hat{\theta}_T - \theta_* \\ \hat{\sigma}_T^2 - \sigma_*^2 \end{pmatrix} \implies N\left(0, \sigma_*^2 \begin{pmatrix} [\partial\Theta_* - I]^{-1} M_*^{-1} \big[[\partial\Theta_* - I]^{-1}\big]^\top & 0 \\ 0 & 5\Delta\sigma_*^2 \end{pmatrix}\right),$$

where "$\implies$" denotes the convergence in distribution.

With the asymptotic properties of our proposed equilibrium estimator, we are able to develop several statistical methods for our model specification test problems in the next section.

3 Model Specification Tests

3.1 Test Statistic

In this subsection, we construct a testing method using the generator of $\mathcal{A}$ and a function $f(\cdot)$, i.e.

$$\mathcal{A}f(x, y, u; \theta, \sigma^2) := \mathbb{E}\big[f(X_2)\big|X_1 = x, Y_1 = y, U_1 = u; \theta, \sigma^2\big] - f(x).$$

Define the test statistic $\widehat{\mathbb{S}}_T(f)$ by

$$\widehat{\mathbb{S}}_T(f) := \frac{1}{\sqrt{T\Delta}} \sum_{t=1}^{T} \mathcal{A}f(X_t, Y_t, \widetilde{U}_t(\hat{\theta}_T); \hat{\theta}_T, \hat{\sigma}_T^2).$$

We have the following asymptotic normality of $\widehat{\mathbb{S}}_T(f)$.

Theorem 3. *Let $f(\cdot)$ be a smooth function such that $\mathcal{A}f(x, y, u; \theta, \sigma^2)$ are second-order continuously differentiable with respect to u, θ, σ^2 with bounded second-order derivatives and*

$$|\partial_z(\mathcal{A}f)(x, y, u; \theta, \sigma^2)| \le K(1 + |x|^2 + |u|^2)$$

for $z = u, \theta, \sigma^2$, where $\partial_z(\mathcal{A}f)$ is the derivative of $\mathcal{A}f$ with respect to the variable z for $z = u, \theta, \sigma^2$. If $\mathcal{P}$ is true,

$$\widehat{\mathbb{S}}_T(f) \implies N(0, \varsigma^2(\theta_*, \sigma_*^2; f))$$

for some $\varsigma^2(\theta_*, \sigma_*^2; f) \ge 0$ being a proper variance depending on $f(\cdot)$ and (θ_*, σ_*^2).

Note that above theorem guarantees the asymptotic normality of $\widehat{\mathbb{S}}_T(f)$, while the explicit form of its asymptotic variance $\varsigma^2(\theta_*, \sigma_*^2; f)$ (or its approximation $\varsigma^2(\hat{\theta}_T, \hat{\sigma}_T^2; f)$) is nearly impossible to obtain for general $f(\cdot)$ and Δ. Therefore, a Bootstrap method is needed, introduced in the next subsection. The key in the Bootstrap simulation is to find the critical value $\hat{s}_{\mathrm{b},\alpha/2}(\hat{\theta}_T, \hat{\sigma}_T^2; f)$ to approximate $z_{\alpha/2} \cdot \varsigma(\theta_*, \sigma_*^2; f)$, where $z_{\alpha/2}$ is the $100(1-\alpha/2)$ percentile of the standard normal distribution.

3.2 Bootstrap Method and Testing Procedure

In this subsection, a Bootstrap method is proposed to finding the critical value of rejection. After the Bootstrap approximation is obtained, the corresponding testing method is concluded directly.

Given a sequence of observations $\{(X_t, Y_t)\}_{t=1}^T$, following the two-step procedure described as before, the equilibrium estimator $(\hat{\theta}_T, \hat{\sigma}_T^2)$ can be obtained. Using the parameter $(\theta, \sigma^2) = (\hat{\theta}_T, \hat{\sigma}_T^2)$ in (1), we can construct $\{(\tilde{X}_{i,t}, \tilde{Y}_{i,t}) : t = 1, \cdots, T_{\mathrm{b}}\}$ for $i = 1, \cdots, N_{\mathrm{b}}$. Here, T_{b} is the number of observations, commonly $T_{\mathrm{b}} = T$, and N_{b} is the number of the Bootstrap replications. For each i, we can derive a Bootstrap statistic from the simulated observations $\{(\tilde{X}_{i,t}, \tilde{Y}_{i,t}) : t = 1, \cdots, T_{\mathrm{b}}\}$ and the set of all N_{b} Bootstrap statistics is denoted by

$$\left\{\widehat{\mathbb{S}}_{i,\mathrm{b}}(\hat{\theta}_T, \hat{\sigma}_T^2; f) : i = 1, \cdots, N_b\right\}.$$

Finally, define

$$\hat{s}_{\mathrm{b},\alpha/2}(\hat{\theta}_T, \hat{\sigma}_T^2; f) = \sup\left\{s \geq 0 : \frac{\sharp\{i : |\widehat{\mathbb{S}}_{i,\mathrm{b}}(\hat{\theta}_T, \hat{\sigma}_T^2; f)| \geq s\}}{N_{\mathrm{b}}} \geq \alpha\right\}.$$

It is natural to have the following testing procedure:

Testing Procedure: For those $f(\cdot)$ with $\varsigma(\theta_*, \sigma_*^2; f) \neq 0$, the parametric family $\mathcal{P}$ is rejected if

$$\hat{\gamma}_T \geq \gamma_0 \quad \text{or} \quad |\widehat{\mathbb{S}}_T(f)| \geq \hat{s}_{\mathrm{b},\alpha/2}(\hat{\theta}_T, \hat{\sigma}_T^2; f)$$

for some given threshold $\gamma_0 \geq 0$. Our asymptotic normality established above together with Proposition 1 concludes that the probability of falsely rejecting H_0 is asymptotically α as $T_b \to \infty$, $N_b \to \infty$ and $T \to \infty$ sequentially. Note that if $\mathbf{\Theta} = \{\theta^*\}$ (consequently

$\hat{\theta}_T = \theta_*$), the testing procedure as above is also true. Some remarks are needed here for the applicability of our method.

Remark 1: The additional rejection rule $\hat{\gamma}_T \geq \gamma_0$ for some threshold $\gamma_0 \geq 0$ is to make our test powerful. It is with probability 0 asymptotically for larger T, when $\mathcal{P}$ is true. If $\mathcal{P}$ is not true, $\mathcal{P}$ can be rejected in two scenarios: (i) the equilibrium estimator as a minimum exists, while there does not exist a fixed point; (ii) the equilibrium estimator exists as a fixed point, while the test statistic $|\widehat{\mathbb{S}}_T(f)| \geq \hat{s}_{\mathrm{b},\alpha/2}(\hat{\theta}_T, \hat{\sigma}_T^2; f)$. Indeed, one can see that for some alternative hypotheses, scenario (i) is dominant if $\mathcal{P}$ is not true in our simulation study. To make the proposed test powerful, the additional rejection rule $\hat{\gamma}_T \geq \gamma_0$ is necessary. Because we need to approximate U using ergodicity, it is natural to expect that the method does not behave as well as the case with full information. Especially, the additional rejection rule $\hat{\gamma}_T \geq \gamma_0$ brings some significant errors for the test sizes when T is small. We will see this in our simulation study later.

Remark 2: The threshold γ_0 is free to select to balance between the power and the accuracy of test sizes. Note that the larger value of γ_0 leads to a more accurate test size but a small power. Because $\hat{\gamma}_T = 0$ if T is large and $\mathcal{P}$ is true, a small γ_0 is recommended to increase the test power. From Lemma 2, we see that the probability of $\gamma_T = 0$ is asymptotic 1 and thus, γ_T does not admit an asymptotic normality. To select an appropriate γ_0 in practice, one can simulate several replications of the estimation error $\hat{\gamma}_T$ and let γ_0 be its high (such as 99%) percentile such that the additional rejection rule has little effect on the empirical test size. Especially, it is expected that the method leads to $\gamma_0 = 0$, when T is large enough.

Remark 3: We emphasize that the Bootstrap method is adopted to approximate the $100(1-\alpha/2)$ percentile instead of the asymptotic variance $\varsigma^2(\theta_*, \sigma_*^2; f)$. This is because there might exist some extreme cases in the Bootstrap simulations with a large error γ, when finding the equilibrium estimator (the fixed point does not exist). The large error might provide an extremely inaccurate replication of the statistic which dominates the Bootstrap variance. In contrast, the percentile estimation $\hat{s}_{\mathrm{b},\alpha/2}(\hat{\theta}_T, \hat{\sigma}_T^2; f)$ is stable from those extreme cases.

Remark 4: Of importance it is to select the testing function $f(\cdot)$ such that the following two requirements should be satisfied.

- The asymptotic variance $\varsigma^2(\theta_*, \sigma_*^2; f)$ is not small such that the asymptotic normality holds. This can be verified using Bootstrap method too by excluding the extreme cases as mentioned.

- Both of $|f(X_T)|/\sqrt{T\Delta}$ and $|f(X_1)|/\sqrt{T\Delta}$ are negligible compared to $\widehat{\mathbb{S}}_T(f)$. Those two errors come from the testing function $f(\cdot)$, which can be checked explicitly (see (17) for details).

Finally, it would be interesting to explain the importance of aggregate shocks to understand the effect of mean wealth in the wealth evolution. Suppose that the HAM does not involve aggregate shocks, i.e., Y_t is a constant process. In this case, the mean wealth U_t becomes a mean instead of a conditional mean, which turns out to be a constant m in a long time range due to the ergodicity. Consequently, the mean term m is not distinguishable from the parameters in the HAM. The verification of approximate aggregation is impossible because a conventional recursion perfectly describes the evolution of the wealth process. This explains why the observations on aggregate shocks Y_t are critical here.

3.3 Extension to Continuous-Time Case

Now, we want to show how our proposed procedure can be applied to another important case: the continuous-time HAM. In this scenario, it is to test whether the continuous-time observation $\{(X_t, Y_t) : t \geq 0\}$ follows from the following mean-field SDE

$$dX_t = [b_0(X_t, Y_t, U_t) + \langle \theta, b_0(X_t, Y_t, U_t)\rangle]dt + \sigma * \sigma_0(X_t, Y_t)dW_t$$

with $U_t = \mathbb{E}[X_t|, \mathcal{F}_{t^-}^Y]$, where $\mathcal{F}_\cdot^Y$ is the natural filtration of Y and W_t is a standard Brownian motion. Then, the equilibrium estimators are

$$\hat{\theta}_T = \operatorname*{argmin}_{\vartheta \in \mathbf{\Theta}} |\widehat{\Theta}_T(\vartheta) - \vartheta| \quad \text{and} \quad \hat{\sigma}_T^2 = \frac{1}{T}\int_0^T \frac{(dX_t)^2}{\sigma_0^2(X_t, Y_t)},$$

where $\widehat{\Theta}_T(\vartheta) = \widehat{M}_T^{-1}(\vartheta)\widehat{\Phi}_T(\vartheta)$,

$$\widehat{M}_T(\vartheta) = \frac{1}{T}\int_0^T \frac{b(X_t, Y_t, \widetilde{U}_t(\vartheta))b^\top(X_t, Y_t, \widetilde{U}_t(\vartheta))}{\sigma_0^2(X_t, Y_t)}dt,$$

$$\widehat{\Phi}_T(\vartheta) = \frac{1}{T}\int_0^T \frac{b(X_t, Y_t, \widetilde{U}_t(\vartheta)}{\sigma_0^2(X_t, Y_t)}\Big(dX_t - b_0(X_t, Y_t, \widetilde{U}_t(\vartheta))dt\Big),$$

and

$$d\widetilde{U}_t(\vartheta) = [b_0(\widetilde{U}_t(\vartheta), Y_t, \widetilde{U}_t(\vartheta)) + \langle \vartheta, b_0(\widetilde{U}_t(\vartheta), Y_t, \widetilde{U}_t(\vartheta)) \rangle] dt.$$

Here, the linearity in x for the drift coefficients is used when calculating $\widetilde{U}_t(\vartheta)$. Applying a testing function $f(\cdot)$, the test statistic is defined as

$$\widehat{\mathbb{S}}_T(f) = \frac{1}{\sqrt{T}} \int_0^T \mathcal{A} f(X_t, Y_t, \widetilde{U}_t(\hat{\theta}_T); \hat{\theta}_T, \hat{\sigma}_T^2) dt.$$

Similar to the discrete-time case, one can verify the asymptotic normality through Taylor's expansion. Under some regularity conditions, as expected, the asymptotic variance of $\widehat{\mathbb{S}}_T(f)$ can be derived, given by

$$\varsigma^2(\theta, \sigma^2; f) = \sigma^2 \int \left[f' \sigma_0 + p^\top [\partial \Theta(\theta) - I]^{-1} M^{-1}(\theta) \frac{b(x,y,u)}{\sigma_0(x,y)} \right]^2 \nu_\theta(dx, dy, du, dv)$$

where $\nu_\theta(dx, dy, du, dv)$ is the ergodic measure of (X, Y, U, V) for $V = \partial_\vartheta U$ and $p = \int f'[b + v(\partial_u b_0 + \langle \theta, \partial_u b \rangle)] \nu_\theta(dx, dy, du, dv)$. With all the information above, a testing procedure can be constructed in the same manner.

4 Monte Carlo Simulation Study

In this section, we conduct a Monte Carlo simulation to illustrate the finite sample performance of our proposed tests. The model used can be seen as a generalized version of the discrete-time Vasicek model with mean-field interaction in a switching environment.

Let Y be a Markov Chain with state space $\{0, 1\}$ and transition matrix Q for which the different values of Y represent the state of the environment in practice. Consider the following linear recursion:

$$X_{t+1} - X_t = \Big[\theta_1 + \theta_2 Y_t + X_t(\theta_3 + \theta_4 Y_t) + U_t(\theta_5 + \theta_6 Y_t) \Big] \Delta + \sigma \sqrt{\Delta} w_t$$

with $U_t = \mathbb{E}[X_t | Y_1, \cdots, Y_{t-1}]$. Now, write $b(x, y, u) = (1, y, x, xy, u, uy)^\top$, $\sigma \in \mathbb{R}_+$, and $\theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6)^\top \in \mathbf{\Theta} \subset \mathbb{R}^6$. By (5), $\widetilde{U}_t(\vartheta)$ can be constructed through

$$\widetilde{U}_{t+1}(\vartheta) = \widetilde{U}_t(\vartheta) + \Big(\vartheta_1 + \vartheta_2 Y_t + \widetilde{U}_t(\vartheta)[\vartheta_3 + \vartheta_5 + (\vartheta_4 + \vartheta_6) Y_t] \Big) \Delta$$

with $\widetilde{U}_1(\vartheta) = X_1$.

Next, let us present a concrete example to justify our theory. Suppose

$$\Delta = 1 \text{ and } Q = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

Here, we take a fixed Q because the model specification test for a Markov chain is well-known and our focus is on that of X only. The test function is taken by $f(x) = x^2$ and the test statistic becomes

$$\widehat{\mathbb{S}}_T(f) = \frac{1}{\sqrt{T\Delta}} \sum_{t=1}^{T} \Big[\Big(X_t + \Delta \langle \hat{\theta}_T, b(X_t, Y_t, \widetilde{U}(\hat{\theta}_T)) \rangle \Big)^2 - X_t^2 + \Delta \hat{\sigma}_T^2 \Big].$$

Recall that $\mathcal{P}$ is rejected if

$$\hat{\gamma}_T(\hat{\theta}_T) \geq \gamma_0 \text{ or } |\widehat{\mathbb{S}}_T(f)| \geq \hat{s}_{\mathrm{b},\alpha/2}(\hat{\theta}_T, \hat{\sigma}_T^2; f),$$

where we take $\gamma_0 = 0.01$ in the simulation. For each simulation, 500 replications are performed. A replication is called a "failure" if the equilibrium estimator is such that $\hat{\gamma}_T = |\widehat{\Theta}_T(\hat{\theta}_T) - \hat{\theta}_T| > \gamma_0 = 0.01$, i.e. the additional rejection rule is applied. The proportion of failures is listed to justify the necessity of the additional rejection rule.

Tables 1 and 2 report the test sizes for different numbers of sample size (i.e. $T = 100$ and 500) for $(\theta_*, \sigma_*) = (0.5, 0.3, -0.8, 0.1, 0.1, 0.1, 0.3)$ and $(0.5, 0.3, -0.8, 0.1, 0, 0, 0.3)$, respectively. Note that the latter selection of the true parameter corresponds to the conventional stochastic systems independent of U_t. From Tables 1 and 2, it is clear that the test sizes converge to the nominal size as the sample size becomes large. Such a simulation result justifies that our tests perform well for not only mean-field systems but also for conventional stochastic systems. We also see that the number of failures decreases when T gets larger, which is in line with our asymptotic theory. Moreover, we see that $\gamma_0 = 0.01$ is a good choice for the case of $T = 500$ because the probability of falsely rejecting H_0 due to the additional rejection rule $\hat{\gamma}_T \geq \gamma_0$ is at most 0.2%. For the case $T = 100$, the false rejection from the additional rejection rule has a proportion 3% and 5.2%, respectively, which is significant to the nominal size α. Therefore, the choice of $\gamma_0 = 0.01$ is not good for the case $T = 100$ considering the test sizes. Note that the way to select an appropriate threshold γ_0 in practice is presented in Section 3.2.

Table 1: The test sizes for different significance levels α and sample size T with $\theta_* = (0.5, 0.3, -0.8, 0.1, 0.1, 0.1)$ and $\sigma_* = 0.3$.

α	0.10	0.05	0.01	♯ of failures/500
T=100	0.0920	0.0640	0.0380	0.0300
T=500	0.1040	0.0580	0.0080	0.0020

Table 2: The test sizes for different significance levels α and sample size T with $\theta_* = (0.5, 0.3, -0.8, 0.1, 0, 0)$ and $\sigma_* = 0.3$.

α	0.10	0.05	0.01	♯ of failures/500
T=100	0.1140	0.0900	0.0520	0.0520
T=500	0.0940	0.0580	0.0180	0.0000

In Tables 3 and Table 4, we consider two hypotheses, $H_0 : \mathbf{b}(x, y, u) = \langle \theta, b(x, y, u) \rangle$ and $\boldsymbol{\sigma}(x, y, u) = \sigma$ versus $H_1 : \mathbf{b}(x, y, u) = \langle \theta, b(x, y, u) \rangle$ and $\boldsymbol{\sigma}(x, y, u) = \sigma(1 + \lambda\sqrt{|x|})$, and $H_0 : \mathbf{b}(x, y, u) = \langle \theta, b(x, y, u) \rangle$ and $\boldsymbol{\sigma}(x, y, u) = \sigma$ versus $H_1 : \mathbf{b}(x, y, u) = \langle \theta, b(x, y, u) \rangle + \lambda h(x)$ and $\boldsymbol{\sigma}(x, y, u) = \sigma$, respectively, where $\lambda \geq 0$ is a varying constant and $h(x) = \sin(x)$. Tables 3 and 4 summarize the test powers for $\theta_* = (0.5, 0.3, -0.8, 0.1, 0.1, 0.1)$ and $\sigma_* = 0.3$. From Tables 3 and 4, one can see clearly that when λ departures from 0, the test power tends to one quickly because the number of failures increases rapidly which dominates the power. This justifies the power of our tests and the necessity of the additional rejection rule $\hat{\gamma}_T \geq \gamma_0$ in our testing procedure.

Table 3: The test powers for different values of λ under $H_0 : \mathbf{b}(x, y, u) = \langle \theta, b(x, y, u)$ and $\boldsymbol{\sigma}(x, y, u) = \sigma$ versus $H_1 : \mathbf{b}(x, y, u) = \langle \theta, b(x, y, u) \rangle$ and $\boldsymbol{\sigma}(x, y, u) = \sigma(1 + \lambda\sqrt{|x|})$ with $\theta_* = (0.5, 0.3, -0.8, 0.1, 0.1, 0.1)$, $\sigma_* = 0.3$, $\alpha = 0.05$ and $T = 500$.

λ	0	1	3	5
Power	0.0580	0.0573	0.5480	0.9440
♯ of failures/500	0.0020	0.0220	0.5120	0.9440

5 Conclusion

To test whether the HAM features *approximate aggregation*, we propose a model-specification testing problem for the evolution of one agent's wealth under partial infor-

Table 4: The test powers for different values of λ under $H_0 : \mathbf{b}(x, y, u) = \langle \theta, b(x, y, u) \rangle$ and $\boldsymbol{\sigma}(x, y, u) = \sigma$ versus $H_1 : \mathbf{b}(x, y, u) = \langle \theta, b(x, y, u) \rangle + \lambda h(x)$ and $\boldsymbol{\sigma}(x, y, u) = \sigma$ with $\theta_* = (0.5, 0.3, -0.8, 0.1, 0.1, 0.1)$, $\sigma_* = 0.3$, $\alpha = 0.05$ and $T = 500$.

λ	0	1	2	3
Power	0.0580	0.0460	0.1000	0.8760
$\sharp$ of failures/500	0.0020	0.0000	0.0500	0.8620

mation. The main difficulty of our problem lies in the lack of observations on the total wealth of all agents. To overcome this challenge, we propose a novel two-step verification to derive an equilibrium estimator if the null hypothesis is true. We see that our method generalizes the ML method proposed in Parra-Alvarez et al. (2023) for the HAM without shocks. Also, we establish the consistency and asymptotic normality of the proposed equilibrium estimator. Then, through the generator of the observations, we construct a testing statistic and establish its asymptotic theory. Consequently, a testing rule is obtained. Different from the testing rules for conventional diffusion models as in Ait-Sahalia (1996), Hong and Li (2005), Chen et al. (2008), and Ait-Sahalia et al. (2009), our testing rule consists of two parts: the estimating error $\hat{\gamma}_T$ and the test statistic. To the best of our knowledge, the first part involving $\hat{\gamma}_T$ is new in the literature which is necessary to make our tests powerful. The simulation study shows that indeed, our tests have good test sizes and are powerful.

This paper is the first attempt to study the model specification testing problems for partially observed HAMs with aggregate shocks. The idea introduced has great potential for those generalized model specification testing problems for mean-dependent stochastic processes. We admit that the theory developed here is just an infant. For example, it is assumed that the income processes are normally distributed (they may be allowed other distributions such as Bernoulli distribution which is used in Krusell and Smith (1998) and Parra-Alvarez et al. (2023)) such that the equilibrium estimator has an explicit form. Also, it is assumed that the diffusion coefficient involves a one-dimensional parameter only such that the two-step verification only applies to those parameters in the drifts. Those assumptions might restrict the application of the developed theory here to more general HAMs. Despite those restrictive assumptions in the paper, there is no doubt that the idea of an equilibrium estimator is applicable to similar problems with more general $\mathcal{P}$. Hopefully, more results on these issues can be obtained and reported in the future.

6 Mathematical Proofs

In this section, we present the mathematical proofs of the main results. For the sake of convenience, we only deal with $m = 1$ in the proof, because the proof for other m's is similar. First, note that $O_p(1)$ stands for a term which is bounded in probability, and $o_p(1)$ means that it converges to 0 in probability. Now, recall the parametric family $\mathcal{P}$ satisfies

$$\mathbf{b}(x, y, u) = A_0(y)x + B_0(y, u) + \theta\big[A(y)x + B(y, u)\big].$$

Define the auxiliary processes $\{(\widetilde{U}_t(\vartheta), \widetilde{V}_t(\vartheta), \widetilde{W}_t(\vartheta))\}_{t=1}^{T}$ by

$$\widetilde{U}_{t+1}(\vartheta) = \widetilde{U}_t(\vartheta) + \Delta\big[A_0(Y_t)\widetilde{U}_t(\vartheta) + B_0(Y_t, \widetilde{U}_t(\vartheta))\big] + \Delta\vartheta\big[A(Y_t)\widetilde{U}_t(\vartheta) + B(Y_t, \widetilde{U}_t(\vartheta))\big]$$

with $\widetilde{U}_1(\vartheta) = X_1$,

$$\begin{aligned}\widetilde{V}_{t+1}(\vartheta) = \widetilde{V}_t(\vartheta)\Big[1 + \Delta\Big(A_0(Y_t) + \partial_u B_0(Y_t, \widetilde{U}_t(\vartheta)) + \vartheta\big[A(Y_t) + \partial_u B(Y_t, \widetilde{U}_t(\vartheta))\big]\Big)\Big] \\ + \Delta\big[A(Y_t)\widetilde{U}_t(\vartheta) + B(Y_t, \widetilde{U}_t(\vartheta))\big]\end{aligned} \tag{8}$$

with $V_1(\vartheta) = 0$, and

$$\begin{aligned}\widetilde{W}_{t+1}(\vartheta) = \widetilde{W}_t(\vartheta)\Big[1 + \Delta\Big(A_0(Y_t) + \partial_u B_0(Y_t, \widetilde{U}_t(\vartheta)) + \vartheta\big[A(Y_t) + \partial_u B(Y_t, \widetilde{U}_t(\vartheta))\big]\Big)\Big] \\ + \widetilde{V}_t(\vartheta)\Big[\widetilde{V}_t(\vartheta)\Big(\partial_u^2 B_0(Y_t, \widetilde{U}_t(\vartheta)) + \vartheta\partial_u^2 B(Y_t, \widetilde{U}_t(\vartheta))\Big) + A(Y_t) + \partial_u B(Y_t, \widetilde{U}_t(\vartheta))\Big]\Delta, \\ + \Delta\big[A(Y_t) + \partial_u B(Y_t, \widetilde{U}_t(\vartheta))\big]\widetilde{V}_t(\vartheta)\end{aligned}$$

with $\widetilde{W}_1(\vartheta) = 0$. In fact, $\widetilde{V}(\vartheta)$ and $\widetilde{W}(\vartheta)$ are defined such that $\widetilde{V}_t(\vartheta) = \partial_\vartheta \widetilde{U}_t(\vartheta)$and $\widetilde{W}_t(\vartheta) = \partial_\vartheta \widetilde{V}_t(\vartheta)$.

The following is one of our main assumptions.

Assumption 2. (A1) *The coefficients $B(y, u)$ are second-order differentiable with respect to u with bounded Lipschitz derivatives and satisfy*

$$|B_i(y, u)| + |\sigma(x, y)| \le L(1 + |x| + |u|), \quad |A_i(y)| \le L, \quad \text{and} \quad |\sigma(x, y)| > \varepsilon$$

for some $L, \varepsilon > 0$.

(A2) *The stochastic process* $(X(\theta), Y(\theta), U(\theta), \widetilde{U}(\vartheta))$ *is ergodic and satisfies*

$$\sup_{\theta\in\mathbf{\Theta}}\sup_{t}\mathbb{E}|X_t(\theta)|^4<\infty,\quad \mathbb{E}|X_t^{(1)}(\theta)-X_t^{(2)}(\theta)|^2\le Le^{\lambda_0 t}\mathbb{E}|X_0^{(1)}-X_0^{(2)}|^2, \tag{9}$$

and

$$\sup_{\vartheta}\mathbb{E}|\widetilde{U}_t^{(1)}(\vartheta)-\widetilde{U}_t^{(2)}(\vartheta)|^2\le Le^{-\lambda_0 t}\mathbb{E}|\widetilde{U}_0^{(1)}-\widetilde{U}_0^{(2)}|^2 \tag{10}$$

for some $L, \lambda_0 > 0$ *independent of* $\vartheta \in \mathbf{\Theta}$*, where* $(X_t^{(i)}(\theta), U_t^{(i)}(\theta), \widetilde{U}_t^{(i)}(\vartheta))$ *satisfies* (1) *with initial value* $(X_0^{(i)}, U_0^{(i)}, \widetilde{U}_0^{(i)})$ *and parameter* $(\theta, \vartheta, \sigma_*)$*. Moreover, the ergodic measure* $\nu_\theta(dx, dy, du)$ *of* $(X(\theta), Y(\theta), U(\theta))$ *is continuous in Wasserstein-2 metric (see, for example, Villani (2009) for details) with respect to* θ*. We also write* $\mu_\theta(dx, dy, du, d\tilde{u}; \vartheta)$ *by the ergodic measure of* $(X(\theta), Y(\theta), U(\theta), \widetilde{U}(\vartheta))$*.*

(A3) *The true parameter* θ_* *is the unique solution to* $\Theta(\vartheta) = \vartheta$ *in* $\mathbf{\Theta}$ *and* $\inf_{\vartheta\in\mathbf{\Theta}} \det[M(\vartheta)] > 0$*.*

(A4) *It follows that*

$$\limsup_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\big[\sup_{\vartheta\in\mathbf{\Theta}}|\widetilde{U}_t(\vartheta)|^2+\sup_{\vartheta\in\mathbf{\Theta}}|\widetilde{V}_t(\vartheta)|^2+\sup_{\vartheta\in\mathbf{\Theta}}|\widetilde{W}_t(\vartheta)|^2\big]<\infty. \tag{11}$$

(A5) $\Theta(\cdot)$ *is continuously differentiable in* $\mathbf{\Theta}$ *with* $\det[I - \partial\Theta(\theta_*)] \neq 0$ *where* $\partial\Theta(\cdot)$ *is the gradient field of* $\Theta(\cdot)$*.*

Now, let us discuss the aforementioned assumptions. (A1) distinguishes our concern from the trivial case when $\mathbf{\Theta} = \{\theta^*\}$. The exponential ergodicity in (A2) and the uniform boundedness in (A4) can be verified through the well-known Lyapunov method if some appropriate conditions are imposed on the coefficients. The readers are referred to Mao (2007) for more details on such a method. (A3) is a necessary assumption such that $\Theta(\cdot;\cdot)$ is well defined on $\mathbf{\Theta}$ and θ^* is the unique a fixed point. Note that θ_* is always a solution if $M(\theta^*)$ is invertible. The uniqueness is the essence here. From above, it seems that (A1)-(A3) are mild assumptions for model specification problems on stochastic systems. (A4) is separately verified later (see Proposition 2) under some regularity conditions and (A5) is a necessary technical assumption needed for the two-step procedure. In the sequel, we always assume Assumption 2 holds. Then, we proceed with the following several lemmas.

Lemma 1. *It follows that under Assumptions 1 and 2,*

$$\lim_{T\to\infty}\sup_{\vartheta\in\boldsymbol{\Theta}}|\widehat{\Theta}_T(\vartheta)-\Theta(\vartheta)|=0,\ \ \lim_{T\to\infty}\sup_{\vartheta\in\boldsymbol{\Theta}}|\partial\widehat{\Theta}_T(\vartheta)-\partial\Theta(\vartheta)|=0,\ \ \text{and}\ \ \sup_{\vartheta\in\boldsymbol{\Theta}}|\partial_{\vartheta_i}\partial_{\vartheta_j}\widehat{\Theta}_T(\vartheta)|\le L_0 \tag{12}$$

almost surely, for some finite random variable $L_0>0$.

Proof. Recall the definition of $\widehat{M_T}(\vartheta)$ in (6), we have

$$\begin{aligned}|\partial_\vartheta\widehat{M_T}(\vartheta)|&=\Big|\frac{1}{T}\sum_{t=1}^{T}\frac{2[A(Y_t)+B(Y_t,\widetilde{U}_t(\vartheta)]\partial_u B(Y_t,\widetilde{U}_t(\vartheta)V_t(\vartheta)}{\sigma_0^2(X_t,Y_t)}\Big|\\ &\le\frac{K}{T\varepsilon}\sum_{t=1}^{T}(|U_t(\vartheta)|^2+|V_t(\vartheta)|^2)\to L_0\end{aligned}$$

for some $L_0>0$. This implies that $\partial_\vartheta\widehat{M_T}(\vartheta)$ is uniformly bounded in T on almost all sample paths. Therefore with probability 1, $\widehat{M_T}(\vartheta)\to M(\vartheta)$ uniformly for $\theta\in\boldsymbol{\Theta}$. Similarly, one can show that $\widehat{\Phi}_T(\vartheta)$ uniformly converges to $\Phi(\vartheta)$ uniformly for $\theta\in\boldsymbol{\Theta}$ for almost all sample paths. By (A4) in Assumption 2, we have

$$\inf_{\vartheta\in\boldsymbol{\Theta}}\det\left[\widehat{M_T}(\vartheta)\right]\ge\inf_{\vartheta\in\boldsymbol{\Theta}}\det\left[M(\vartheta)\right]-\sup_{\vartheta\in\boldsymbol{\Theta}}\Big|\det\left[\widehat{M_T}(\vartheta)\right]-\det\left[M(\vartheta)\right]\Big|\to\inf_{\vartheta\in\boldsymbol{\Theta}}\det\left[M(\vartheta)\right]>0$$

almost surely. By writing the adjugate matrix of $\widehat{M_T}(\cdot)$ by $\widehat{M_T^\dagger}(\cdot)$, then,

$$\widehat{\Theta}_T(\vartheta)=\Big(\det\left[\widehat{M_T}(\vartheta)\right]\Big)^{-1}\widehat{M_T^\dagger}(\vartheta)\widehat{\Phi}_T(\vartheta),$$

which yields the first assertion in (12). The proofs for the second and third results in (12) are similar, omitted. □

Next, it is ready to prove the consistency result in Theorem 1.

Proof of Theorem 1 . Note that

$$\begin{aligned}|\Theta(\hat{\theta}_T)-\hat{\theta}_T|&\le\sup_{\vartheta\in\boldsymbol{\Theta}}|\widehat{\Theta}_T(\vartheta)-\Theta(\vartheta)|+|\widehat{\Theta}_T(\hat{\theta}_T)-\hat{\theta}_T|\\ &=\sup_{\vartheta\in\boldsymbol{\Theta}}|\widehat{\Theta}_T(\vartheta)-\Theta(\vartheta)|+\inf_{\vartheta\in\boldsymbol{\Theta}}|\widehat{\Theta}_T(\vartheta)-\vartheta|\\ &\le 2\sup_{\vartheta\in\boldsymbol{\Theta}}|\widehat{\Theta}_T(\vartheta)-\Theta(\vartheta)|+\inf_{\vartheta\in\boldsymbol{\Theta}}|\Theta(\vartheta)-\vartheta|=2\sup_{\vartheta\in\boldsymbol{\Theta}}|\widehat{\Theta}_T(\vartheta)-\Theta(\vartheta)|\to 0,\end{aligned}$$

almost surely. As θ_* is the unique solution to $\Theta(\vartheta) = \vartheta$ in $\mathbf{\Theta}$, therefore, $\hat{\theta}_T \to \theta_*$ almost surely. Let $\widetilde{U}_t(\theta_*)$ satisfy (5) with $\widetilde{U}_1(\theta_*) = X_1$. Note that $\widetilde{U}_t(\theta_*)$ and the true U_t are both calculated from (5) with the same parameter θ_* while with different initial values. Then, it follows that

$$\begin{aligned}
\hat{\sigma}_T^2 &= \frac{1}{T}\sum_{t=1}^{T} \frac{[X_{t+1} - X_t - \Delta b_0(X_t, Y_t, \widetilde{U}_t(\hat{\theta}_T)) - \Delta\langle \hat{\theta}_T, b(X_t, Y_t, \widetilde{U}_t(\hat{\theta}_T))\rangle]^2}{\Delta \sigma_0^2(X_t, Y_t)} = \frac{{\sigma_*}^2}{T}\sum_{t=1}^{T} w_t^2 \\
&+ \frac{\Delta}{T}\sum_{t=1}^{T} \frac{\Big(b_0(X_t, Y_t, U_t) + \langle \theta_*, b(X_t, Y_t, U_t)\rangle - b_0(X_t, Y_t, \widetilde{U}_t(\hat{\theta}_T)) - \langle \hat{\theta}_T, b(X_t, Y_t, \widetilde{U}_t(\hat{\theta}_T))\rangle\Big)^2}{\sigma_0^2(X_t, Y_t)} \\
&+ \frac{2\sigma_*\sqrt{\Delta}}{T}\sum_{t=1}^{T} \frac{b_0(X_t, Y_t, U_t) + \langle \theta_*, b(X_t, Y_t, U_t)\rangle - b_0(X_t, Y_t, \widetilde{U}_t(\hat{\theta}_T)) - \langle \hat{\theta}_T, b(X_t, Y_t, \widetilde{U}_t(\hat{\theta}_T))\rangle}{\sigma_0(X_t, Y_t)} w_t \\
&= \frac{{\sigma_*}^2}{T}\sum_{t=1}^{T} w_t^2 + O_p(1)\frac{\Delta}{T}\sum_{t=1}^{T}\Big(|\widetilde{U}_t(\hat{\theta}_T) - U_t|^2 + |\hat{\theta}_T - \theta_*|^2(1 + |X_t|^2 + \sup_{\vartheta}|\widetilde{U}(\vartheta)|^2)\Big) \\
&+ O_p(1)\frac{2\sigma_*\sqrt{\Delta}}{T}|\hat{\theta}_T - \theta_*|\sum_{t=1}^{T}|w_t|\sup_{\vartheta}|V_t(\vartheta)|. \qquad (13)
\end{aligned}$$

Note that for some uniform $K > 0$, one has

$$\mathbb{E}(|X_t|^2 + \sup_{\vartheta}|\widetilde{U}_t(\vartheta)|^2) < K, \quad \mathbb{E}|\widetilde{U}_t(\hat{\theta}_T) - \widetilde{U}_t(\theta_*)|^2 \le K\mathbb{E}|\hat{\theta}_T - \theta_*|,$$

and $\mathbb{E}|\widetilde{U}_t(\theta_*) - U_t|^2 \le K\mathbb{E}|X_1 - U_1|^2 e^{-\lambda_0 t}$. By Lemma 1 and the consistency of $\hat{\theta}_T$ (in the compact set $\mathbf{\Theta}$), so that $\hat{\sigma}_T^2 = {\sigma_*}^2 + o_p(1)$. The proof is complete. □

To prove Theorem 2 , we need the following two lemmas.

Lemma 2. *It follows that $\hat{\gamma}_T = 0$ when T is large.*

Proof. Let $h(\vartheta) = \Theta(\vartheta) - \vartheta$ and $h_T(\vartheta) = \widehat{\Theta}_T(\vartheta) - \vartheta$. Without loss of generality, it is assumed that $\partial h(\theta_*) = I$. Otherwise, we perform a local linear transformation on ϑ. Then, around θ_*, $h(\vartheta) = \vartheta - \theta_* + o(|\vartheta - \theta_*|)$. Therefore, it can be easily seen that the conditions for the well-known Poincaré-Miranda Theorem hold. By the uniform convergence of h_T to h on each sample path, the Poincaré-Miranda theorem is applicable for h_T, which yields h_T admits a solution in $\mathbf{\Theta}$, i.e. $\hat{\gamma}_T = 0$. The proof is complete. □

Lemma 3. *If $\xi_t \to 0$ in probability and $\sup_t \mathbb{E}|\eta_t| < \infty$, then $\xi_t\eta_t \to 0$ in probability.*

Proof. For any $\varepsilon > 0$ and $\delta > 0$, it follows that

$$\limsup_{t\to\infty} \mathbb{P}(|\xi_t \eta_t| \geq \varepsilon) \leq \limsup_{t\to\infty} \Big(\mathbb{P}(|\xi_t| \geq \delta) + \mathbb{P}(|\eta_t| \geq \varepsilon/\delta)\Big) \leq \delta/\varepsilon \sup_t \mathbb{E}|\eta_t|.$$

By the arbitrariness of $\delta > 0$, then, $\limsup_{t\to\infty} \mathbb{P}(|\xi_t\eta_t| \geq \varepsilon) = 0$. Therefore, this concludes the lemma. □

Proof of Theorem 2. By the definitions of $\widehat{M}_T(\vartheta)$ and $\widehat{\Theta}_T(\vartheta)$ and the Lipschitz continuity of $B(y,u)$ with respect u, it follows that

$$\begin{aligned}
&\widehat{M}_T(\theta_*)(\widehat{\Theta}_T(\theta_*) - \theta_*) = \widehat{\Phi}_T(\theta_*) - \widehat{M}_T(\theta_*)\theta_* \\
&= \frac{1}{T\Delta}\sum_{t=1}^T \frac{b(X_t, Y_t, \widetilde{U}_t(\theta_*))}{\sigma_0^2(X_t, Y_t)}\big[X_{t+1} - X_t - b_0(X_t, Y_t, \widetilde{U}_t(\theta_*))\Delta - \theta_* b(X_t, Y_t, \widetilde{U}_t(\theta_*))\Delta\big] \\
&= \frac{1}{T\sqrt{\Delta}}\sum_{t=1}^T \frac{b(X_t, Y_t, \widetilde{U}_t(\theta_*))}{\sigma_0(X_t, Y_t)} w_t + O(1) \times \frac{\Delta}{T}\sum_{t=1}^T |U_t - \widetilde{U}_t(\theta_*)|,
\end{aligned}$$

which concludes that

$$\begin{aligned}
&\sqrt{T\Delta}[\widehat{\Theta}_T(\theta_*) - \theta_*] = \frac{1}{\sqrt{T}} M^{-1}(\theta_*)\sum_{t=1}^T \frac{b(X_t, Y_t, \widetilde{U}_t(\theta_*))}{\sigma_0(X_t, Y_t)} w_t \\
&+ (\widehat{M}_T^{-1}(\theta_*) - M^{-1}(\theta_*)) \cdot \frac{1}{\sqrt{T}}\sum_{t=1}^T \frac{b(X_t, Y_t, \widetilde{U}_t(\theta_*))}{\sigma_0(X_t, Y_t)} w_t + O(1) \times \widehat{M}_T^{-1}(\theta_*)\frac{1}{T}\sum_{t=1}^T |U_t - \widetilde{U}_t(\theta_*)|.
\end{aligned} \tag{14}$$

This together with Lemma 3 yields that

$$(\widehat{M}_T^{-1}(\theta_*) - M^{-1}(\theta_*)) \cdot \frac{1}{\sqrt{T}}\sum_{t=1}^T \frac{b(X_t, Y_t, \widetilde{U}_t(\theta_*))}{\sigma_0(X_t, Y_t)} w_t = o_p(1). \tag{15}$$

An application of (14) and (15) implies that

$$\sqrt{T\Delta}[\widehat{\Theta}_T(\theta_*) - \theta_*] = \frac{1}{\sqrt{T}} M^{-1}(\theta_*)\sum_{t=1}^T \frac{b(X_t, Y_t, \widetilde{U}_t(\theta_*))}{\sigma_0(X_t, Y_t)} w_t + o_p(1).$$

By (12), using the Taylor expansion, one has

$$\begin{aligned}
\sqrt{T\Delta}[\theta_* - \widehat{\Theta}_T(\theta_*)] &= \sqrt{T\Delta}[\widehat{\Theta}_T(\hat{\theta}_T) - \widehat{\Theta}_T(\theta_*)] + \sqrt{T\Delta}\hat{\gamma}_T - \sqrt{T\Delta}[\hat{\theta}_T - \theta_*] \\
&= [\partial\widehat{\Theta}(\theta_*) - I]\sqrt{T\Delta}(\hat{\theta}_T - \theta_*) + \frac{1}{2}O_p(1)\sqrt{T}(\hat{\theta}_T - \theta_*)^2 + o_p(1) + \sqrt{T\Delta}\hat{\gamma}_T.
\end{aligned}$$

As $\partial\widehat{\Theta}(\theta_*) \to \partial\Theta(\theta_*)$ almost surely and $\sqrt{T\Delta}\hat{\gamma}_T = o_p(1)$ by Lemma 2, we have

$$\sqrt{T\Delta}(\hat{\theta}_T - \theta_*) = o_p(1) + \frac{1}{\sqrt{T}}[I - \partial\Theta(\theta_*)]^{-1} M^{-1}(\theta_*) \sum_{t=1}^{T} \Big[\frac{b(X_t, Y_t, \widetilde{U}_t(\theta_*))}{\sigma_0(X_t, Y_t)} w_t\Big]. \tag{16}$$

Together with (13), the asymptotic normality holds. The proof is complete. □

Proof of Theorem 3. Note that

$$\begin{aligned}
\widehat{\mathbb{S}}(f) &= \frac{1}{\sqrt{T\Delta}} \sum_{t=1}^{T} \mathcal{A}f(X_t, Y_t, \widetilde{U}_t(\hat{\theta}_T); \hat{\theta}_T, \hat{\sigma}_T^2) - \frac{1}{\sqrt{T\Delta}} \sum_{t=1}^{T} \mathcal{A}f(X_t, Y_t, \widetilde{U}_t(\theta^*); \theta_*, \sigma_*^2) \qquad (17) \\
&\quad + \frac{1}{\sqrt{T\Delta}} \sum_{t=1}^{T} \mathcal{A}f(X_t, Y_t, \widetilde{U}_t(\theta^*); \theta_*, \sigma_*^2) - \frac{1}{\sqrt{T\Delta}} \sum_{t=1}^{T} \mathcal{A}f(X_t, Y_t, U_t; \theta_*, \sigma_*^2) \\
&\quad + \frac{1}{\sqrt{T\Delta}} \sum_{t=1}^{T} \mathcal{A}f(X_t, Y_t, U_t; \theta_*, \sigma_*^2) \\
&= O_p(1) \frac{1}{\sqrt{T\Delta}} \sum_{t=1}^{T} |\widetilde{U}_t(\hat{\theta}_T) - \widetilde{U}_t(\theta^*)| \\
&\quad + \sqrt{T\Delta}(\hat{\theta}_T - \theta_*) \frac{1}{T\Delta} \sum_{t=1}^{T} \Big(\partial_u \mathcal{A}f(X_t, Y_t, U_t; \theta_*, \sigma_*^2) V_t + \partial_\theta \mathcal{A}f(X_t, Y_t, U_t; \theta_*, \sigma_*^2)\Big) \\
&\quad + \sqrt{T\Delta}(\hat{\sigma}_T^2 - \sigma_*^2) \frac{1}{T\Delta} \sum_{t=1}^{T} \partial_{\sigma^2} \mathcal{A}f(X_t, Y_t, U_t; \theta_*, \sigma_*^2) + \frac{O(1)}{\sqrt{T\Delta}} \sum_{t=1}^{T} \Big(|\hat{\theta}_T - \theta_*|^2 + |\hat{\sigma}^2 - \sigma_*^2|^2\Big) \\
&\quad + \frac{1}{\sqrt{T\Delta}} \Big(f(X_{T+1}) - f(X_1)\Big) - \frac{1}{\sqrt{T\Delta}} \sum_{t=1}^{T} \Big[f(X_{t+1}) - f(X_t) - \mathcal{A}f(X_t, Y_t, U_t; \theta_*, \sigma_*^2)\Big].
\end{aligned}$$

By (9) and (10) and the representation of $\hat{\theta}_T, \hat{\sigma}_T^2$ in (16) and (13), the CLT for martingales yields that $\widehat{\mathbb{S}}_T(f)$ converges to a normal distribution whose variance is written by $\varsigma^2(\theta_*, \sigma_*^2; f)$.

□

Proposition 1. *Suppose that $\varsigma^2(\theta, \sigma^2; f)$ is continuous of $(\theta, \sigma^2) \in \mathbf{\Theta} \times \mathbb{R}_+$. As $T_b \to \infty$, $N_b \to \infty$ and $T \to \infty$ sequentially, it follows that*

$$\lim_{T\to\infty} \lim_{N_b\to\infty} \lim_{T_b\to\infty} \mathbb{P}\Big(|\widehat{\mathbb{S}}_T(f)| \geq \hat{s}(\hat{\theta}_T, \hat{\sigma}_T^2; f)\Big) = \alpha/2.$$

Proof. Given $\hat{\theta}_T, \hat{\sigma}_T^2$, by the CLT of the test statistics in Theorem 3, we know that

$$\lim_{N_b\to\infty} \lim_{T_b\to\infty} \hat{s}(f; \hat{\theta}_T, \hat{\sigma}_T^2) = z_{\alpha/2} \cdot \varsigma(\hat{\theta}_T, \hat{\sigma}_T^2; f)$$

almost surely. Therefore, by the continuity of $\sigma(\cdot; f)$ and the almost sure consistency of $(\hat{\theta}_T, \sigma_T^2)$, we have

$$\lim_{T\to\infty}\lim_{N_b\to\infty}\lim_{T_b\to\infty}\mathbb{P}\Big(|\widehat{\mathbb{S}}_T(f)| \geq \hat{s}(\hat{\theta}_T, \hat{\sigma}_T^2; f)\Big) \geq \lim_{T\to\infty}\mathbb{P}\Big(|\widehat{\mathbb{S}}_T(f)| > z_{\alpha/2}\cdot\varsigma(\hat{\theta}_T, \hat{\sigma}_T^2; f)\Big) = \alpha/2,$$

and

$$\lim_{T\to\infty}\lim_{N_b\to\infty}\lim_{T_b\to\infty}\mathbb{P}\Big(|\widehat{\mathbb{S}}_T(f)| \geq \hat{s}(\hat{\theta}_T, \hat{\sigma}_T^2; f)\Big) \leq \lim_{T\to\infty}\mathbb{P}\Big(|\widehat{\mathbb{S}}_T(f)| \geq z_{\alpha/2}\cdot\varsigma(\hat{\theta}_T, \hat{\sigma}_T^2; f)\Big) = \alpha/2.$$

The proof is complete. □

Finally, let us present a proposition to verify (A4) in Assumption 2.

Proposition 2. *Suppose that either of the following hold:*

- $B_i(y, 0)$ *are uniformly bounded and*

$$-\lambda_1 < A_0(y) + \partial_u B_0(y, u) + \sum_{i=1}^m \vartheta_i[A_i(y) + \partial_u B_i(y, u)] < -\lambda_0,$$

  or all (y, u, ϑ) *and some constants* $\lambda_1 > \lambda_0 > 0$.

- Y *is a finite-state Markov Chain with state space* $\mathcal{Y}$ *and* $B_i(y, u) = C_i(y) + D_i(y)u$. *Let*

$$K(y) = 1 + \Delta\Big(A_0(y) + D_0(y) + \sup_{\vartheta\in\mathbf{\Theta}}\sum_{i=1}^m \vartheta_i[A_i(y) + D_i(y)]\Big).$$

  There exists a positive function G *on* $\mathcal{Y}$ *such that*

$$\sum_{\bar{y}\in\mathcal{Y}} q_{y,\bar{y}} G(\bar{y}) K^2(\bar{y}) < \lambda_0 G(y) \tag{18}$$

  for some $0 < \lambda_0 < 1$ *where* $(q_{y,\bar{y}})$ *is the transition matrix of* Y.

Then, when Δ *is small, the tuple* $(\sup_{\vartheta\in\mathbf{\Theta}} \widetilde{U}_t(\vartheta), \sup_{\vartheta\in\mathbf{\Theta}} \widetilde{V}_t(\vartheta), \sup_{\vartheta\in\mathbf{\Theta}} \widetilde{W}_t(\vartheta))$ *satisfies*

$$\limsup_{T\to\infty}\frac{1}{T}\sum_{t=1}^T \big[\sup_{\vartheta\in\mathbf{\Theta}}|\widetilde{U}_t(\vartheta)|^2 + \sup_{\vartheta\in\mathbf{\Theta}}|\widetilde{V}_t(\vartheta)|^2 + \sup_{\vartheta\in\mathbf{\Theta}}|\widetilde{W}_t(\vartheta)|^2\big] < \infty.$$

Proof. Let us prove (11) holds for those two cases.

(1) By the recursion of $\widetilde{U}(\vartheta)$ in (8), using mean-value theorem, we have

$$\sup_{\vartheta}|\widetilde{U}_{t+1}(\vartheta)|$$

$$\le \sup_{\vartheta} |\widetilde{U}_t(\vartheta)| \cdot \sup_{\vartheta} \Big[1 + \Delta\Big(A_0(Y_t) + \sum_{i=1}^m \vartheta_i A_i(Y_t) + \partial_u B_0(Y_t, \cdot) + \sum_{i=1}^m \vartheta_i \partial_u B_i(Y_t, \cdot)\Big)\Big]$$
$$+ |B(Y_t, 0)| + \sup_{\vartheta} \sum_{i=1}^m |\vartheta_i B_i(Y_t, 0)|$$
$$\le \lambda_0 \sup_{\vartheta} |\widetilde{U}_t(\vartheta)| + |B(Y_t, 0)| + \sup_{\vartheta} \sum_{i=1}^m |\vartheta_i B_i(Y_t, 0)|$$

for some $\lambda_0 \in (0, 1)$, when Δ is small. This says that $\widetilde{U}_t(\vartheta)$ are uniformly bounded. Similarly, one can prove that $\widetilde{V}_t(\vartheta)$ and $\widetilde{W}_t(\vartheta)$ are uniformly bounded as well.

(2) In this case, we have

$$\begin{aligned}
&\sup_{\vartheta} |\widetilde{U}_{t+1}(\vartheta)| \\
&= \sup_{\vartheta} \left|\widetilde{U}_t(\vartheta)\Big[1 + \Delta\Big(A_0(Y_t) + D_0(Y_t) + \sum_{i=1}^m \vartheta_i [A_i(Y_t) + D_i(Y_t)\Big)\Big] + \Delta\Big(C_0(Y_t) + \sum_{i=1}^m \vartheta_i C_i(Y_t)\Big)\right| \\
&\le K(Y_t) \sup_{\vartheta} |\widetilde{U}_t(\vartheta)| + L(Y_t), \qquad (19)
\end{aligned}$$

where $L(y) = \sup_\vartheta |C_0(y) + \sum_{i=1}^m \vartheta_i C_i(y)|$. This motivates us to consider $\mathcal{U}_t$ satisfying $\mathcal{U}_{t+1} = K(Y_t)\mathcal{U}_t + L(Y_t)$. For any $\delta > 0$, it follows that

$$\begin{aligned}
\mathbb{E}\Big[G(Y_t)\mathcal{U}_{t+1}^2\Big] &\le \mathbb{E}\Big[(1+\delta)[G(Y_t)K^2(Y_t)\mathcal{U}_t^2] + (1+\delta^{-1})G(Y_t)L^2(Y_t)\Big] \\
&= (1+\delta)\lambda_0 \mathbb{E}\big[G(Y_{t-1})\mathcal{U}_t^2\big] + (1+\delta^{-1})\mathbb{E}[G(Y_t)L^2(Y_t)]. \qquad (20)
\end{aligned}$$

Picking δ small such that $(1+\delta)\lambda_0 < 1$, by Grownwall's inequality, it follows that $\mathbb{E}[G(Y_t)\mathcal{U}_t^2] < L_0$ for some $L_0 > 0$. Since $G(y) > \varepsilon$ for some $\varepsilon > 0$, then, $\mathbb{E}\mathcal{U}_{t+1}^2 \le \varepsilon^{-1}\mathbb{E}[G(Y_t)\mathcal{U}_t^2] < L_0\varepsilon^{-1}$. Because $(\mathcal{U}_t, Y_t)$ is a Markov process and $\mathcal{U}_t$ satisfies a linear recursion with the uniform bounded second moments, it is ergodic with

$$\lim_{T\to\infty} T^{-1} \sum_{t=1}^T \mathcal{U}_t^2 < \infty$$

almost surely. Similar to (19) and (20), we have

$$\sup_{\vartheta} |\widetilde{V}_{t+1}(\vartheta)|^2 \le \lambda \sup_{\vartheta} |\widetilde{V}_t(\vartheta)|^2 + L\mathcal{U}_t^2.$$

Grownwall's inequality implies that

$$\lim_{T\to\infty} T^{-1} \sum_{t=1}^T \sup_{\vartheta} |\widetilde{V}_{t+1}(\vartheta)| \le \lim_{T\to\infty} LT^{-1} \sum_{t=1}^T (T-t)e^{-\lambda(T-t)}\mathcal{U}_t^2 \le L \lim_{T\to\infty} T^{-1} \sum_{t=1}^T \mathcal{U}_t^2 < \infty.$$

The proof for $\widetilde{W}_t(\vartheta)$ is same given the linear structure assumed. $\square$

Now, let us discuss the two cases dealt with in Proposition 2. The first case requires the system to be uniform dissipative for all of Y. Such a condition is fairly strong on the dissipativity but weak in the form of stochastic process Y. To weaken the uniform dissipativity condition, (18) is assumed, when Y is a finite-state Markov chain and the drift coefficients are linear with respect u. Such a condition is parallel to the Lyapunov condition for stochastic switching systems (see Yin and Zhu (2010) for more details). We admit that those conditions are far from necessary for (12), while how to get a better condition is beyond the scope of this paper and thus is omitted here.

Disclosure Statements

The authors claim that there is no conflict of interest in this manuscript. Also, the authors declare that they do not use any generative AI and AI-assisted technologies in the writing process, to analyze and draw insights from data as part of the research process.

References

Abbott, B., G. Gallipoli, C. Meghir, and G. L. Violante (2019). Education policy and intergenerational transfers in equilibrium. *Journal of Political Economy 127*(6), 2569–2624.

Achdou, Y., J. Han, J.-M. Lasry, P.-L. Lions, and B. Moll (2022). Income and wealth distribution in macroeconomics: A continuous-time approach. *Review of Economic Studies 89*(1), 45–86.

Ait-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *Review of Financial Studies 9*(2), 385–426.

Ait-Sahalia, Y., J. Fan, and H. Peng (2009). Nonparametric transition-based tests for jump diffusions. *Journal of the American Statistical Association 104*(487), 1102–1116.

Aiyagari, S. R. (1994). Uninsured idiosyncratic risk and aggregate saving. *Quarterly Journal of Economics 109*(3), 659–684.

Benhabib, J., A. Bisin, and M. Luo (2019). Wealth distribution and social mobility in the us: A quantitative approach. *American Economic Review 109*(5), 1623–1647.

Bewley, T. (1986). Stationary monetary equilibrium with a continuum of independently fluctuating consumers. *Contributions to Mathematical Economics In Honor of Gérard Debreu 79*.

Chen, S., J. Gao, and C. Tang (2008). A test for model specification of diffusion processes. *Annals of Statistics 36*(1), 167–198.

Dachian, S. and Y. A. Kutoyants (2008). On the goodness-of-fit tests for some continuous time processes. *Statistical models and methods for biomedical and technical systems*, 385–403.

Han, J., Y. Yang, et al. (2021). Deepham: A global solution method for heterogeneous agent models with aggregate shocks. *arXiv preprint arXiv:2112.14377*.

Heathcote, J., K. Storesletten, and G. L. Violante (2009). Quantitative macroeconomics with heterogeneous households. *Annual Review of Economics 1*(1), 319–354.

Hedlund, A., F. Karahan, K. Mitman, and S. Ozkan (2017). Monetary policy, heterogeneity, and the housing channel. In *2017 Meeting Papers*, Volume 1610.

Hommes, C. H. (2006). Heterogeneous agent models in economics and finance. *Handbook of Computational Economics 2*, 1109–1186.

Hong, Y. and H. Li (2005). Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *Review of Financial Studies 18*(1), 37–84.

Huang, M., P. E. Caines, and R. P. Malhamé (2003). Individual and mass behaviour in large population stochastic wireless power control problems: Centralized and nash equilibrium solutions. In *42nd IEEE International Conference on Decision and Control (IEEE Cat. No. 03CH37475)*, Volume 1, pp. 98–103. IEEE.

Huang, M., P. E. Caines, and R. P. Malhamé (2007). Large-population cost-coupled lqg problems with nonuniform agents: Individual-mass behavior and decentralized ε-nash equilibria. *IEEE Transactions on Automatic Control 52*(9), 1560–1571.

Huggett, M. (1993). The risk-free rate in heterogeneous-agent incomplete-insurance economies. *Journal of Economic Dynamics and Control 17*(5-6), 953–969.

Kaplan, G., B. Moll, and G. L. Violante (2018). Monetary policy according to hank. *American Economic Review 108*(3), 697–743.

Kleptsyna, M. and Y. A. Kutoyants (2014). On asymptotically distribution free tests with parametric hypothesis for ergodic diffusion processes. *Statistical Inference for Stochastic Processes 17*, 295–319.

Krusell, P. and A. A. Smith, Jr (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy 106*(5), 867–896.

Lasry, J.-M. and P.-L. Lions (2006a). Jeux à champ moyen. i–le cas stationnaire. *Comptes Rendus Mathématique 343*(9), 619–625.

Lasry, J.-M. and P.-L. Lions (2006b). Jeux à champ moyen. ii–horizon fini et contrôle optimal. *Comptes Rendus. Mathématique 343*(10), 679–684.

Liu, L. and M. Plagborg-Moller (2023). Full-information estimation of heterogeneous agent models using macro and micro data. *Quantitative Economics 14*(1), 1–35.

López-Pérez, A., M. Febrero-Bande, and W. González-Manteiga (2022). A comparative review of specification tests for diffusion models. *arXiv preprint arXiv:2208.08420*.

Mao, X. (2007). *Stochastic Differential Equations and Applications*. Elsevier.

Mongey, S. and J. Williams (2017). Firm dispersion and business cycles: Estimating aggregate shocks using panel data. *Manuscript, New York University*.

Negri, I. and Y. Nishiyama (2009). Goodness of fit test for ergodic diffusion processes. *Annals of the Institute of Statistical Mathematics 61*, 919–928.

Parra-Alvarez, J. C., O. Posch, and M.-C. Wang (2023). Estimation of heterogeneous agent models: A likelihood approach. *Oxford Bulletin of Economics and Statistics 85*(2), 304–330.

Reiter, M. (2009). Solving heterogeneous-agent models by projection and perturbation. *Journal of Economic Dynamics and Control 33*(3), 649–665.

Rios-Rul, J. V. and V. Rios-Rull (1995). Models with heterogeneous agents. In *Frontiers of Business Cycles Research*. Princeton University Press.

Ríos-Rull and José-Víctor (2001). Computation of equilibria in heterogeneous-agent models. In *Computational Methods for the Study of Dynamic Economies*. Oxford University Press, New York.

Villani, C. (2009). *Optimal Transport: Old and New*, Volume 338. Springer, Berlin.

Williams, J. (2017). Bayesian estimation of DSGE models with heterogeneous agents. pp. Unpublished Mamuscript.

Winberry, T. (2018). A toolbox for solving and estimating heterogeneous agent macro models. *Quantitative Economics 9*(3), 1123–1151.

Yin, G. G. and C. Zhu (2010). *Hybrid Switching Diffusions*. Springer, New York.