LIKELIHOOD INFERENCE FOR LARGE FINANCIAL SYSTEMS
PRELIMINARY DRAFT

JUSTIN A. SIRIGNANO, GUSTAVO SCHWENKLER, AND KAY GIESECKE

Abstract. We consider the problem of parameter estimation for large interacting stochastic systems where data is available on the aggregate state of the system. Parameter inference is computationally challenging due to the scale and complexity of such systems. Weak convergence results, similar in spirit to a law of large numbers and a central limit theorem, can be used to approximate large systems in distribution. We exploit these weak convergence results in order to develop approximate maximum likelihood estimators for such systems. The approximate estimators are shown to converge to the true parameters and are asymptotically normal as the number of observations and the size of the system become large. Numerical studies demonstrate the computational efficiency and accuracy of the approximate MLEs. Although our approach is widely applicable to large systems in many fields, we are particularly motivated by examples arising in finance such as systemic risk in banking systems and large portfolios of loans.

1. Introduction

Large interacting stochastic systems are common in economics and finance. Examples include the banking system, an auction system, the stock market, and a pool of assets such as corporate, consumer, or mortgage loans held by a financial institution or backing a security with cash flows tied to the pool. While the modeling and analysis of financial systems has received considerable attention, parameter inference is less well developed, hindering practical applications. A significant issue is the scale and complexity of a typical system, which render conventional estimation methods computationally burdensome. For instance, a typical pool of residential mortgages for a mortgage-backed security may consist of tens of thousands of loans and a major US bank might easily have on the order of 20,000 wholesale loans, 50,000 to 100,000 mid-market and commercial loans, and derivatives trades with 10,000 to 20,000 different counterparties. Another major issue is data availability. In practice, granular data on individual system components are usually difficult and costly to collect for large systems. Often, data are available only for the aggregate state of the system, which means filtering is required for statistical estimation. Filtering is typically computationally intractable due to the size and complexity of the system, especially when observations are exact (i.e., no observational noise). This paper addresses the inference problem arising in this situation. It develops computationally tractable likelihood estimators for the parameters of a large financial system. The results include consistency, asymptotic normality, and a numerical algorithm for computing estimators.

We consider a broad class of dynamic interacting stochastic systems. The behavior of a system component is described by a jump-diffusion process addressing idiosyncratic, systematic, and interaction sources of randomness. The latter two sources cause correlation between the components' behavior. Components are exposed to a common exogenous diffusion (systematic source) and a mean-field term (interaction source). The mean field term depends upon the state of the entire system. A concrete example is a pool of mortgages. Each loan has idiosyncratic risk specific to their particular circumstances, such as income and credit score of the homeowner. Every mortgage is also exposed to systematic risk such as the state of the local, regional, and national economy, as described by factors such as unemployment, housing prices, GDP growth, and inflation. Finally, there is interaction between the mortgages: foreclosure on a home due to default of the home owner depresses the prices of homes in the same area, making additional defaults more likely. Our formulation is broad enough to include the models of financial systems recently proposed and analyzed in the literature, including portfolio credit risk models such as Bush, Hambly, Haworth, Jin & Reisinger (2011), Cvitanic, Ma & Zhang (2012), Dai Pra, Runggaldier, Sartori & Tolotti (2009), and Giesecke, Spiliopoulos

Date: November 11, 2014. The authors gratefully acknowledge support from the National Science Foundation through Methodology, Measurement, and Statistics Grant No. 1325031.

& Sowers (2012); systemic risk models such as Garnier, Papanicolaou & Yang (2013), and Fouque & Ichiba (2013); limit order book/market dynamic models such as Crisan, Kurtz & Lee (2012) and Lasry & Lions (2007); and mean field game models such as Guéant (2009), Carmona, Fouque & Sun (2013), and Chan & Sircar (2014).

We develop an asymptotic likelihood approach to address the inference problem for a large stochastic system. Estimation is based on a misspecified model of the system, which is obtained by an approximation of the system via weak convergence results. This approach exploits the laws of large numbers and central limit theorems which have been developed for many systems, including several financial systems.[1] An "approximate likelihood function" is constructed by computing the likelihood for the misspecified model and evaluating it at the actual data (generated by the true model). Our approximate likelihood estimator for the parameters is the maximizer of the approximate likelihood. The approximate likelihood computation is based on the conditionally Gaussian structure of the misspecified model. Taking advantage of this conditionally Gaussian structure, a computationally efficient numerical method is devised to calculate the approximate estimator. In many cases, the approximate likelihood can be evaluated semi-analytically and easily maximized using standard optimization methods. Under mild technical conditions, the approximate likelihood estimator is consistent and asymptotically normal as the number of observations and the size of the system become large. Numerical studies demonstrate the properties of the estimators for various systems and observation schemes. The estimators are very accurate even for systems of moderate size.

There exists very little previous literature on maximum likelihood estimation for large interacting systems. Kasonga (1990) considers a system of interacting diffusions and proves the consistency and asymptotic normality of the true maximum likelihood estimator. Bishwal (2011) considers a similar system where the parameter to be estimated is a function of time. He uses the Radon-Nikodym theorem to develop a maximum likelihood estimator in the continuous observation case. In the discrete observation case, he approximates the integrals that appear in the Radon-Nikodym derivative by a discrete sum with the timestep chosen to coincide with the discrete observation interval. Such an approach introduces error when the discrete observation interval is large. Both authors' frameworks assume one can observe each particle in the system and require that the system only has drift and diffusion (no jumps). Due to the jumps in the system we consider, the Radon-Nikodym derivative approach would not work. Furthermore, neither author addresses the case where only aggregate observations of the system are available.

Calculation of the true likelihood function for a typical large interacting system is computationally infeasible due to several challenges. Evaluating the likelihood function requires knowledge of the transition density. This is already computationally expensive due to the large size of the system and the lack of analytic or semi-analytic solutions. Generally, the transition density is only available via brute-force Monte Carlo simulation. Simulation becomes prohibitively expensive as the size of the system grows. Even in the relatively rare cases where the transition density can be calculated semi-analytically (such as when the system is affine, see Duffie, Pan & Singleton (2000)), the computational effort involved with evaluating the density becomes intractable as the size of the system grows. Furthermore, in order to calculate the true likelihood function one must also be able to filter the underlying state of the system from the observation process (since in our framework, one only observes some aggregate state of the system). There are currently no methods designed for this task since we assume the observation is exact (i.e., without noise). Even if there was observational noise and one applied particle filters, the computational expense of particle filters grows exponentially with the hidden state (in our case, the size of the system). Finally, direct estimation of the joint density of all the observations via Monte Carlo simulation and multivariate kernel density estimation becomes intractable for any more than a few observations (since the dimension is equal to the number of observations). Even for a few observations, the computations are costly due to simulating a very large system.

[1]Examples in financial settings include Bo & Capponi (2014*a*), Bo & Capponi (2014*b*), Bush et al. (2011), Carmona et al. (2013), Chan & Sircar (2014), Collet, Dai Pra & Sartori (2010), Crisan et al. (2012), Cvitanic et al. (2012), Dai Pra et al. (2009), Fouque & Ichiba (2013), Garnier et al. (2013), Giesecke et al. (2012), Giesecke, Spiliopoulos, Sowers & Sirignano (2013), Lasry & Lions (2007), and Spiliopoulos, Sirignano & Giesecke (2014). Examples from non-financial settings include Achdou & Capuzzo-Dolcetta (2010), Bossy, Fezoui & Piperno (2009), Bossy & Talay (1997), Carmona & Delarue (2013), Carmona, Delarue & Lachapelle (2012), Crimaldi, Pra & Minelli (2014), Dai Pra & den Hollander (1996), Dawson (1983), Delarue, Inglis, Rubenthaler & Tanre (2014), Fernandez & Meleard (2000), Fontbona (2006), Guéant (2009), Guéant (2012), Haynatzka, Gani & Rachev (2000), Kurtz & Xiong (2004), Meleard (1996), Morale, Capasso & Oelschlager (2005), Nagasawa & Tanaka (1987), and Sznitman (1984).

Large interacting stochastic systems are common in other fields besides economics and finance, including energy, reliability, transportation, and biology. For instance, Morale et al. (2005) model social interaction in a large population and Delarue et al. (2014) model a neuronal network in the brain. The estimators developed in this paper could also be used to perform inference for systems occurring in these other areas which model large interacting systems.

The paper is organized as follows. The general interacting stochastic system as well as the observational setting that we consider is presented in Section 2. We introduce and describe our approach for parameter estimation in Section 3. Conditions for consistency and asymptotic normality are stated in Section 4. We consider several different observation settings in Sections 5, 6, and 7. In these sections, we present numerical methods to compute the approximate estimators and theoretic properties such as consistency and asymptotic normality of the approximate estimators are proven. In addition, although the main portion of the paper considers a continuous-time setting, the methods can easily be applied in a discrete-time setting. The discrete-time case is addressed in Section 9. Finally, we demonstrate the efficiency and accuracy of the approximate estimators in practice through various numerical studies in Section 8. All proofs of propositions, lemmas, and theorems can be found in the Appendix.

2. Model Framework

2.1. **Interacting Stochastic System.** Let $(\Omega, \mathbb{F}, \mathbb{P})$ be a probability space equipped with a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ satisfying the usual conditions of right-continuity and completeness (see Protter (2004)). Let X be a one-dimensional diffusion process that takes values in a set $\mathscr{X} \subseteq \mathbb{R}$ and $Y^N = (Y^{N,n} : 1 \leq n \leq N)$ be a sequence of càdlàg processes which compose a system of size $N \subset \mathbb{N}$. Each $Y^{N,n}$ is valued in $\mathscr{Y} \subseteq \mathbb{R}^{d_Y}$. The collection $Y^N = (Y^{N,n})_{1\leq n\leq N}$ are the "agents", "components", or "particles" which compose the system.

The particles Y^N interact through their dependence on the "average" particle state. In addition, the particles' dynamics are correlated through their dependence on the exogenous factor X. More specifically, writing δ for the Dirac measure on $\mathscr{Y}$ and E for the set of probability measures on $\mathscr{Y}$, we define the empirical measure

$$\mu_t^N(\mathrm{d}y) = \frac{1}{N}\sum_{n=1}^{N} \delta_{Y_t^{N,n}}(\mathrm{d}y)$$

in the Skorohod space $D_E([0,T])$ equipped with the weak topology. We assume that the process (X, Y^N) satisfies the following system of stochastic differential equations:

$$\begin{aligned} dX_t &= a(X_t;\alpha)dt + dW_t, \\ dY_t^{N,n} &= \nu(X_t, Y_t^{N,n}, \mu_t^N;\beta)dt + \sigma(X_t, Y_t^{N,n}, \mu_t^N;\beta)^\top dW_t^{N,n} + \gamma(X_t, Y_t^{N,n}, \mu_t^N;\beta)^\top dJ_t^{N,n} \\ &\quad + \zeta(X_t, Y_t^{N,n}, \mu_t^N;\beta)^\top dX_t. \end{aligned} \tag{1}$$

Here, $\theta = (\alpha, \beta)$ is a vector of parameters valued in a Euclidean set Θ, $W^N = (W^{N,n} : 1 \leq n \leq N)$ is a sequence of independent d_Y-dimensional standard Brownian motions, W is a standard Brownian motion independent of W^N, $J^N = (J^{N,n} : 1 \leq n \leq N)$ is a system of non-explosive d_Y-dimensional counting processes so that $J^{N,n}$ has intensity $\Lambda^{N,n} = \big(\Lambda_1(X_t, Y_t^{N,n}, \mu_t^N;\beta), \ldots, \Lambda_{d_Y}(X_t, Y_t^{N,n}, \mu_t^N;\beta)\big)$, and $\nu : \mathscr{Y} \times \mathscr{X} \times E \times \Theta \to \mathbb{R}^{d_Y}$, $\sigma : \mathscr{Y} \times \mathscr{X} \times E \times \Theta \to \mathbb{R}^{d_Y}$, $\gamma : \mathscr{Y} \times \mathscr{X} \times E \times \Theta \to \mathbb{R}^{d_Y}$, $\zeta : \mathscr{Y} \times \mathscr{X} \times E \times \Theta \to \mathbb{R}^{d_Y}$, and $\Lambda : \mathscr{Y} \times \mathscr{X} \times E \times \Theta \to \mathbb{R}_+^{d_Y}$ are coefficient functions. The initial values $Y_0^{N,n}$ are assumed to be i.i.d. random variables with probability law π_0^N.

The model framework (1) describes a broad array of interacting systems across many fields and is characterized by several key ingredients. The system is composed of agents $Y^{N,n}$ whose dynamics are driven by three factors: an idiosyncratic factor $W^{N,n}$, a systematic process X, and a mean field factor μ_t^N. The idiosyncratic factor models behavior or events specific to a particular agent. The systematic factor X represents exogeneous processes that all agents are exposed to. For instance, in a model of the interbank system, the systematic process could be the general state of the economy (e.g., inflation, GDP growth, or the current interest rate). The mean field factor models the interaction of a particular agent with the other agents in the system. Consequently, the behavior of one agent affects the rest of the population and vice versa. In a model of the interbank system, e.g., the default of one bank may have an adverse effect on the creditworthiness of other banks through various financial, legal, and economic interconnections. The agents' dynamics are therefore correlated through the systematic factor common to all agents as well as interactions between agents via the mean field term.

Various extensions to the framework (1) are possible. In equation (1), the particles are homogeneous; i.e., the dynamics of every particle $Y^{N,n}$ are governed by the same parameter β. Section 9 discusses possible extensions to settings in which different parameter vectors govern the dynamics of different agents. Second, we assume that the systematic risk factor X is a one-dimensional diffusive process with unit volatility. The generalization to a positive volatility process is straightforward given that any one-dimensional diffusion process with strictly positive volatility can be transformed to a unit-volatility diffusion via the Lamperti transform; see Iacus (2008). The extension to cases with multiple independent systematic risk factors that follow SDE as in (1) is also straightforward. Third, if $\mathscr{Y}$ is a finite or semi-infinite domain, we assume that the boundary $\partial\mathscr{Y}$ is either non-attainable or, if it is attainable, it is absorbing. Finally, we assume that the system of SDEs (1) has a unique strong solution. Sufficient conditions guaranteeing strong existence and uniqueness in the case $\gamma = 0$ are coefficient functions satisfying Lipschitz and linear growth conditions; for instance, see Protter (2004). Sufficient conditions guaranteeing existence and uniqueness in a case with jumps can be found in Fernandez & Meleard (2000).

2.2. **Data.** We are interested in estimating the parameter $\theta = (\alpha, \beta)$ for the system (1). The data available for inference include observations of an ensemble feature Z^N of the system. Observations are not necessarily available for every particle in the system; instead, observations are available for the aggregate state of the system. Granular data for each agent is typically difficult and costly to collect for very large systems, motivating the data structure we adopt.

More precisely, the assumption is made that Z^N is a linear function of the empirical measure μ^N; i.e.,

$$Z_t^N = \langle f, \mu_t^N \rangle = \int_{\mathscr{Y}} f(y)\mu_t^N(dy) = \frac{1}{N}\sum_{n=1}^{N} f(Y_t^{N,n}) \tag{2}$$

for a twice continuously differentiable function $f : \mathbb{R}^{d_Y} \to \mathbb{R}$. For instance, if $d_Y = 1$ and $f(y) = y$, the observation process Z is the empirical mean of the particles' positions. For simplicity, we assume that Z^N is one-dimensional but generalizations to the multivariate case are straightforward. Let $(t_M)_{M\geq 0}$ be a sequence of deterministic times such that $0 \leq t_{i-1} < t_i < \infty$ for $i \geq 1$ and $t_M \to \infty$ as $M \to \infty$. The data available for inference is $\mathsf{D}_{M,N} = (\mathsf{Z}_{M,N}, \mathsf{X}_M)$, where $\mathsf{Z}_{M,N}$ includes observations of Z^N and X_M includes observations of X during the time interval $[0, t_M]$. The data $\mathsf{D}_{M,N}$ is assumed to be valued in a set $\mathcal{D}$ and measurable with respect to the Borel σ-algebra on $\mathcal{D}$.

We consider three alternative observational regimes of the processes Z^N and X that are common in empirical applications. In the first regime, the statistician observes Z^N discretely at times $t_0 < t_1 < \ldots < t_M$ while observing X continuously over $[0, t_M]$. In this regime, $\mathsf{Z}_{M,N} = (Z_{t_0}^N, Z_{t_1}^N, \ldots, Z_{t_M}^N)$ and $\mathsf{X}_M = (X_t)_{t\in[0,t_M]}$ with $\mathcal{D} = \mathbb{R}^{M+1} \times C([0, t_M])$. In the second regime, both Z^N and X are observed discretely at times $t_0 < t_1 < \cdots < t_M$. In this regime, $\mathsf{Z}_{M,N} = (Z_{t_0}^N, Z_{t_1}^N, \ldots, Z_{t_M}^N)$, $\mathsf{X}_M = (X_{t_0}, X_{t_1}, \ldots, X_{t_M})$, and $\mathcal{D} = \mathbb{R}^{M+1} \times \mathbb{R}^{M+1}$. In the third regime, both Z^N and X are observed continuously over $[0, t_M]$ so that $\mathsf{Z}_{M,N} = (Z_t^N)_{t\in[0,t_M]}$, $\mathsf{X}_M = (X_t)_{t\in[0,t_M]}$, and $\mathcal{D} = C([0, t_m]) \times C([0, t_m])$. The three observation regimes that we consider are summarized in Table 1.

Observational Regime	Z	X	Section of Paper
Regime 1	Discrete	Discrete	6
Regime 2	Discrete	Continuous	5
Regime 3	Continuous	Continuous	7

Table 1. The three different observational regimes which this paper develops parameter estimation methods for.

2.3. **Weak convergence assumptions.** We assume that there exist processes $\bar{\mu}$ and $\bar{\Xi}$ such that for fixed $t > 0$ and for almost every path $(X_s)_{s\in[0,t]}$,

$$\mu_t^N \overset{E}{\Rightarrow} \bar{\mu}_t, \tag{3}$$

$$\Xi_t^N \equiv \sqrt{N}(\mu_t^N - \bar{\mu}_t) \overset{\mathcal{W}}{\Rightarrow} \bar{\Xi}_t, \tag{4}$$

as $N \to \infty$. Here, "$\overset{E}{\Rightarrow}$" indicates weak convergence in the Skorohod space $D_E([0,T])$ where E is the space of measures on $\mathscr{Y}$. "$\overset{\mathcal{W}}{\Rightarrow}$" denotes weak convergence in a space $D_{\mathcal{W}}([0,T])$.[2] The limiting measure $\bar{\mu}$ is a law of large numbers while $\bar{\Xi}$ is a central limit theorem. The latter is also often referred to in the mathematical literature as a "fluctuation limit". The empirical fluctuation process Ξ_t^N describes the fluctuations of the finite system around the law of large numbers $\bar{\mu}$. Ξ_t^N is called the empirical fluctuation measure and is a signed measure.

For fixed $t > 0$ and conditional on $(X_s)_{s\in[0,t]}$, we assume that $\bar{\mu}$ and $\bar{\Xi}$ satisfy stochastic evolution equations of the form

$$\begin{aligned} d\bar{\mu}_s &= \mathcal{A}^1_{X_s,\bar{\mu}_s}\bar{\mu}_s ds + \mathcal{A}^2_{X_s,\bar{\mu}_s}\bar{\mu}_s dX_s, \\ d\bar{\Xi}_s &= \mathcal{G}^1_{X_s,\bar{\mu}_s}\bar{\Xi}_s ds + \mathcal{G}^2_{X_s,\bar{\mu}_s}\bar{\Xi}_s dX_s + d\bar{\mathcal{M}}_s, \end{aligned} \tag{5}$$

on $[0,t]$ with initial conditions $\bar{\mu}_0 \in E$ and $\bar{\Xi}_0 \in \mathcal{W}$, respectively. We have, of course, that $\mu_0^N \overset{E}{\Rightarrow} \bar{\mu}_0$ and $\Xi_0^N \overset{W}{\Rightarrow} \bar{\Xi}_0$. Conditional on $(X_s)_{s\in[0,t]}$, $(\bar{\mathcal{M}}_s)_{s\in[0,t]}$ is a centered Gaussian with covariance given by:

$$\operatorname{Cov}\left(\left\langle f, \bar{\mathcal{M}}_{t_1}\right\rangle, \left\langle g, \bar{\mathcal{M}}_{t_2}\right\rangle \,\middle|\, (X_s)_{s\in[0,t_2]}\right) = \mathbb{E}\left[\int_0^{t_1} H(f,g,\bar{\mu}_s)\mathrm{d}s \,\middle|\, (X_s)_{s\in[0,t_1]}\right], \tag{6}$$

for $0 \le t_1 < t_2 \le t$ and any $f, g \in \mathcal{W}'$, the dual space of $\mathcal{W}$. H is a function that depends upon the form of (1). Conditional on $(X_s)_{s\in[0,t]}$ and for any test function f in $\mathcal{W}'$, the process $(\langle f, \bar{\mathcal{M}}_s\rangle)_{s\in[0,t]}$ is a Brownian motion. The operators $\mathcal{A}^1_{X_s,\bar{\mu}_s}$ and $\mathcal{A}^2_{X_s,\bar{\mu}_s}$ may be nonlinear, and therefore the stochastic evolution equation for $\bar{\mu}$ can be nonlinear. The operators $\mathcal{G}^1_{X_s,\bar{\mu}_s}$ and $\mathcal{G}^2_{X_s,\bar{\mu}_s}$ are linear; they do not depend upon $\bar{\Xi}_s$. As a result, $(\bar{\Xi}_s)_{s\in[0,t]}$ is also a conditionally Gaussian processes given $(X_s)_{s\in[0,t]}$. Even though the stochastic evolution equation for $\bar{\mu}$ is non-linear, it is driven only by time and the systematic factor X. Consequently, the mapping $s \mapsto \bar{\mu}_s$ for $s \in [0,t]$ is deterministic conditional on $(X_s)_{s\in[0,t]}$. The stochastic evolution equation for $\bar{\mu}_t$ may admit a density; however, the stochastic evolution equation for $\bar{\Xi}_t$ will typically only have a distribution-valued solution.

Weak convergence results as in (3)-(4) that satisfy stochastic evolution equations as in (5) are often available for systems of the type (1). They have been proven for a number of settings, as illustrated in the following examples.

2.4. **Examples.** A number of interacting particle models for financial systems have recently been developed in the literature. We highlight a few here to illustrate how our likelihood estimation methods might be applied.

Example 2.1 (Systemic Risk)**.** *Garnier et al. (2013) develop a mean field model for systemic risk in a large system of interacting agents. Agents can move from a healthy state to a failed state. The authors study the probability of a transition of agents from the healthy state to the failed state. They prove law of large numbers, fluctuation limit, and large deviation results.*

Let $Y_t^{N,n}$ be the risk for agent n where

$$dY_t^{N,n} = -hU(Y_t^{N,n})dt + \theta(\left\langle y, \mu_t^N\right\rangle - Y_t^{N,n})dt + \sigma dW_t^{N,n}.$$

U is a potential function with two stable states. One stable state of U represents the normal state for agents while the other stable state represents the failed state for agents (i.e., default). The empirical mean $\left\langle y, \mu_t^N\right\rangle$ is taken to represent the systemic risk in the system. This particular paper does not include a systematic risk factor. The observed quantity might be the empirical mean. The empirical mean is a natural measure for the aggregate systemic risk in the system since it is the average across the risks of all the agents. The observation process would then be

$$Z_t^N = \left\langle f, \mu_t^N\right\rangle = \left\langle y, \mu_t^N\right\rangle,$$

where f is simply the identity function.

[2] Ξ_t^N is distribution-valued, and $\mathcal{W}$ will typically be chosen to be the dual of a suitable Hilbert space.

Example 2.2 (Interbank Lending). *Fouque & Ichiba (2013) Fouque & Ichiba (2013) propose a model of interbank lending. Banks with higher capital reserves lend to banks with lower capital reserves. Each bank's capital reserves also fluctuates according to a Brownian motion* $W_t^{N,n}$*. There is no systematic risk factor in the model. The monetary reserve of each bank satisfies the SDE*

$$dY_t^{N,n} = \left(\kappa + \langle y, \mu_t^N \rangle - Y_t^{N,n}\right)dt + 2\sqrt{Y_t^{N,n}}dW_t^{N,n},$$

where κ *is a drift. One might observe the average monetary reserve of the banking system, in which case the observation process is*

$$Z_t^N = \langle f, \mu_t^N \rangle,$$

where f *is again the identity function.*

Example 2.3 (Correlated Default Timing). *Giesecke et al. (2013) and Spiliopoulos et al. (2014) study defaults in a large credit pool where each name is subject to idiosyncratic and systematic risk. Furthermore, names interact in the sense that there is contagion risk. When one name defaults, the surviving names become more likely to default themselves. Specifically, the n-th name defaults at the time* $\tau^{N,n}$*, which designates the first arrival of a counting process driven by a stochastic intensity* $\lambda_t^{N,n}$*. Giesecke et al. (2013) and Spiliopoulos et al. (2014) are able to prove a law of large numbers and a central limit theorem for this model setting.*

The dynamics for each name is represented by a two-dimensional process $Y_t^{N,n} = (\lambda_t^{N,n}, M_t^{N,n})$*, where* $\lambda_t^{N,n}$ *is the stochastic intensity and* $M_t^{N,n} = \mathbf{1}_{\tau^{N,n} > t}$*. The total loss in the pool is*

$$L_t^N = \frac{1}{N}\sum_{n=1}^{N}(1 - M_t^{N,n}).$$

The parameter set is $\theta = (\alpha, \beta)$ *where* $\beta = (\kappa, m, \sigma, \beta^C, \beta^S)$*. The process* $Y_t^{N,n}$*, composed of the stochastic intensity* $\lambda_t^{N,n}$ *and the counting process* $M_t^{N,n}$*, has the dynamics*

$$\begin{aligned} d\lambda_t^{N,n} &= \kappa(m - \lambda_t^{N,n})dt + \sigma\sqrt{\lambda_t^{N,n}}dW_t^n + \beta^C dJ_t^{N,n,1} + \beta^S \lambda_t^{N,n} dX_t, \\ dM_t^{N,n} &= -dJ_t^{N,n,2}, \end{aligned} \tag{7}$$

where $J_t^{N,n,2}$ *is a point process with intensity* $M_t^{N,n}\lambda_t^{N,n}$*. The point process* $J_t^{N,n,1} = L_t^N$ *and has intensity* $\frac{1}{N}\sum_{n=1}^N \lambda_t^N M_t^{N,n} = \langle f_\Lambda(y), \mu_t^N \rangle$ *where* $f_\Lambda(y) = \lambda m$ *for* $y = (\lambda, m)$*.* X *is a systematic factor affecting all of the names in the pool. The systematic factor's impact on the default of names in the pool is governed by the parameter* β^S*. There is also a contagion effect where the default of names in the pool cause a spike in the intensity of the surviving names, making the remaining names morely likely to default. This is the mean field or interaction term and its strength is determined by the parameter* β^C*. The names' intensities also tend to revert to a mean intensity level* m*. The systematic risk dynamics satisfies a diffusion process. For instance,* X *could follow a (Lamperti-transformed) OU process. The data in this setting is the loss* L_t^N *and the systematic factor* X *observed at a set of discrete times* $t_1, \ldots, t_M$*. The loss can be written in the form of (2) by setting* $f(y) = 1 - m$*.*

$$Z_t^N = L_t^N = \langle f, \mu_t^N \rangle = \int_{\mathbb{R}^2} f(y)\mu_t^N(dy) = \int_{\mathbb{R}^2}(1 - m)\mu_t^N(d\lambda, dm) = J_t^N.$$

Example 2.4 (Structural Model for Portfolio Credit Risk). *Bush et al. (2011) develop a structural model for the distance to default processes* $V_t^{N,n}$ *of a pool of firms. The n-th firm defaults when the process* $V_t^{N,n}$ *first hits zero. Their model includes a systematic risk factor* X_t *which creates correlation between the default times of different names. However, there is no mean field interaction term.*

$$\begin{aligned} dV_t^{N,n} &= \mu dt + \sqrt{1 - \rho}dW_t^{N,n} + \sqrt{\rho}dX_t, \\ \tau^{N,n} &= \inf\{t : V_t^{N,n} = 0\}, \end{aligned}$$

where X *is a systematic process. Let* $Y_t^{N,n} = \left(V_t^{N,n}, M_t^{N,n}\right)$ *where* $M_t^{N,n} = \mathbf{1}_{\tau^{N,n} > t}$*. The n-th name is removed from the pool at time* $\tau^{N,n}$*. That is,* $y = 0$ *is an absorbing boundary. When the n-th name first hits* *the boundary at $y=0$, $M_t^{N,n}$ is instantly set to zero. The total loss in the pool is*

$$L_t^N = \frac{1}{N}\sum_{n=1}^{N}(1-M_t^{N,n}).$$

The observation process would again be the loss L_t^N at times $t_1,\ldots,t_M$. As before, the loss can be written in the form of (2) by setting $f(\lambda,m)=1-m$.

$$Z_t^N = \langle f,\mu_t^N\rangle = \int_{\mathbb{R}^2}(1-m)\mu_t^N(d\lambda,dm) = L_t^N.$$

Note that for this model, defaults are predictable events occuring when the processes $V_t^{N,n}$ first hit the zero boundary. This example fits into our general framework since the domain in equation 1 can be semi-infinite with an absorbing boundary.

Example 2.5 (Correlated Default Timing). *Cvitanic et al. (2012) develop a dynamic reduced form model for the default of names in a credit pool. The intensities for each name are a function of idiosyncratic and systematic risk factors as well as the loss in the pool. The impact of the loss on the intensities of the names in the pool is permanent. Define M_t^N and L_t^N as in the first example. Let $Y_t^{N,n}=(B_t^{N,n},Q_t^{N,n},\lambda_t^{N,n},M_t^N)$. The dynamics for $Y_t^{N,n}$ are*

$$\begin{aligned}
dB_t^{N,n} &= dW_t^{N,n},\\
dQ_t^{N,n} &= \nu(t,W_t,B_t^{N,n},X_t,L_t^N;\beta)dt+\sigma(t,W_t,B_t^{N,n},L_t^N;\beta)dW_t^{N,n},\\
d\lambda_t^{N,n} &= f^{N,n}(t,W_s,B_t^{N,n},X_t,Q_t^{N,n},L_t^N;\beta)dt,\\
dM_t^{N,n} &= dJ_t^{N,n,2},
\end{aligned}$$

where X is the systematic risk, $W^{N,n}$ is the idiosyncratic source of risk, and L_t^N is the loss. $J_t^{N,n,2}$ is a counting process with intensity $\lambda_t^{N,n}M_t^{N,n}$. The systematic risk could follow a (Lamperti-transformed) SDE. As before, one would observe L_t^N and X at a discrete set of times $t_1,\ldots,t_M$.[3]

Example 2.6 (Correlated Default Timing). *Dai Pra et al. (2009) develop a dynamic reduced form model for a credit portfolio. Each name lies in the state space $Y_t^{N,n}=(p_t^{N,n},q_t^{N,n})$ where $p_t^{N,n}$ and $q_t^{N,n}$ are binary $\{-1,1\}$ and indicate whether the name is in a "good" financial state or a distressed financial state. For instance, p could indicate whether a name has a good credit rating or a bad credit rating. The variable q is not directly observeable; it could represent, for example, the liquidity of the name.*

The global financial indicator is

$$m_t^N = \frac{1}{N}\sum_{n=1}^{N}p_t^{N,n}.$$

This indicator measures the "global financial health" or level of "systemic risk". Each name transitions from $p^{N,n}\to -p^{N,n}$ with intensity $\nu_p(p_t^{N,n},q_t^{N,n},m_t^N)$. Similarly, each name transitions from $q^{N,n}\to -q^{N,n}$ with intensity $\nu_q(p_t^{N,n},q_t^{N,n},m_t^N)$. Let $J_t^{N,n}=\left(J_t^{N,n,p},J_t^{N,n,q}\right)$ be a two-dimensional vector of point processes. The dynamics for $Y_t^{N,n}$ follow

$$\begin{aligned}
dp_t^{N,n} &= -2p_{t_-}^{N,n}dJ_t^{N,n,p},\\
dq_t^{N,n} &= -2q_{t_-}^{N,n}dJ_t^{N,n,q}.
\end{aligned}$$

$J_t^{N,n,p}$ is a counting process with intensity $\nu_p(p_t^{N,n},q_t^{N,n},m_t^N)$. Similarly, $J_t^{N,n,q}$ is a counting process with intensity $\nu_q(p_t^{N,n},q_t^{N,n},m_t^N)$. The observation process would be

$$Z_t^N = \int_{\mathbb{R}^2}p\mu_t^N(dp,dq) = m_t^N.$$

[3] The model (8) has been slightly adapted to the framework (1). In their original paper, Cvitanic et al. (2012) let their model be even more general by allowing for dependence of a and b on the loss L_t^N.

Example 2.7 (Mean Field Game for Energy or Consumer Goods Market). *Chan & Sircar (2014) consider an energy or consumer goods market where producers compete against each other. Each producer must choose what price they will sell the good at. Each producer's profit is affected by the prices and quantities of the good supplied by the other producers in the market. This effect is modeled as a mean field term by the authors.*

Each producer has capacity $Y_t^{N,n}$ at time t, which represents the amount of the good that they are able to sell. The fraction of producers still alive at time t (i.e., have positive capacity) is:

$$\eta_t^N = \int_{\mathbb{R}_+} d\mu_t^N.$$

Each producer chooses a price $p_t^{N,n}$ to sell the good at and sells amount:

$$q_t^{N,n} = a(\eta_t^N;\kappa) - p_t^{N,n} + c(\eta_t^N;\alpha)\bar{p}_t^N,$$

where $\bar{p}_t^N = \frac{1}{\eta_t^N}\frac{1}{N}\sum_{n=1}^N p_t^{N,n}$ and a, c are functions that must be specified. The capacities $Y_t^{N,n}$ have the dynamics:

$$dY_t^{N,n} = -q_t^{N,n}dt + \sigma \mathbf{1}_{Y_t^{N,n}>0} dW_t^{N,n}.$$

The optimal price, or control, satisfies the optimization problem:

$$p_t^{N,n} = \sup_{p\in\mathcal{P}} \mathbb{E}[\int_t^T e^{-r(s-t)} p_s^{N,n} q_s^{N,n} ds | \mathcal{F}_t, p^{N,n} = p],$$

where $\mathcal{P}$ is the class of admissible controls. The parameters to be estimated are $\theta = (\kappa, \alpha, \sigma)$. The observation process Z_t^N might be the total amount of remaining capacity $\langle 1, \mu_t^N \rangle$.

3. Misspecified model for approximate inference

Suppose the true parameter governing the system (1) is $\theta_0 = (\alpha_0, \beta_0)$. Define $\mathbb{D}_\theta^{M,N}$ as the $\mathbb{P}_\theta$-law of the data $\mathsf{D}_{M,N}$. Fix a parameter space $\Theta = \Theta_\alpha \times \Theta_\beta$ and assume that $\theta_0 \in \Theta^\circ$. A standard approach for parameter inference is maximum likelihood estimation. However, maximum likelihood estimation is infeasible due to the scale and complexity of the system (1). Except in the most trivial cases, the transition density for the system (1) cannot be calculated except via brute-force Monte Carlo simulation, rendering filtering and likelihood estimation computationally intractable for any more than a few agents.

We exploit the weak convergence results (3)-(4) to perform parameter inference. The law of large numbers is combined with the central limit theorem to yield a large system approximation for the system (1). Not surprisingly, evaluation of the approximate system distribution is computationally much more tractable than finding the true system distribution via direct Monte Carlo simulation. Numerical tests in Giesecke et al. (2013) and Spiliopoulos et al. (2014) indicate that the large system approximation provides a highly accurate estimate for the finite-dimensional distributions of Z_t^N, even in moderately-large systems of several hundred particles.

The large system approximation is used to develop an approximate likelihood function for the observations. Approximate MLEs are then found using the approximate likelihood function. In other words, parameters are estimated via maximum likelihood estimation from a misspecified model for the data. Due to the weak convergence results, the misspecified model is close to the true model for the data and converges to the true model as the size of the system grows. Thus, we expect the approximate likelihood estimators to be accurate estimators of the true parameters as the size of the system and number of observations become large.

Set

$$\bar{\mu}_t^N = \bar{\mu}_t + \frac{1}{\sqrt{N}}\bar{\Xi}_t. \tag{8}$$

Then $\bar{\mu}^N$ is a second-order approximation of the finite-system empirical measure μ^N for large values of N. Define the approximate observation process

$$\bar{Z}_t^N = \langle f, \bar{\mu}_t^N \rangle = \int_{\mathscr{Y}} f(u)\bar{\mu}_t(du) + \frac{1}{\sqrt{N}}\int_{\mathscr{Y}} f(u)\bar{\Xi}_t(du), \tag{9}$$

and the limiting observation process

$$Z_t^\infty = \langle f, \bar{\mu}_t \rangle = \int_{\mathscr{Y}} f(u) \bar{\mu}_t(du). \tag{10}$$

Define the approximate data $\bar{\mathsf{D}}_{M,N} = (\bar{\mathsf{Z}}_{M,N}, \mathsf{X}_M)$ and the limiting data $\mathsf{D}_{M,\infty} = (\mathsf{Z}_{M,\infty}, \mathsf{X}_M)$. By construction we have that $\mathsf{D}_{M,N} \to \mathsf{D}_{M,\infty}$ in distribution and $\bar{\mathsf{D}}_{M,N} \to \mathsf{D}_{M,\infty}$ almost surely as $N \to \infty$.

Write $\bar{\mathbb{D}}_\theta^{M,N}$ for the $\mathbb{P}_\theta$-law of $\bar{\mathsf{D}}_{M,N}$. Instead of estimating from the family $\mathbf{D}^{M,N} = \{\mathbb{D}_\theta^{M,N} : \theta \in \Theta\}$ that contains the true law $\mathbb{D}_{\theta_0}^{M,N}$ of the data, we will estimate the parameter θ from the approximate family $\bar{\mathbf{D}}^{M,N} = \{\bar{\mathbb{D}}_\theta^{M,N} : \theta \in \Theta\}$ via maximum likelihood. In other words, we perform maximum likelihood estimation using a misspecified model for the data $\mathsf{D}_{M,N}$, which for large N closely approximates the true model. The procedure is as follows. First, we construct the likelihood function for the misspecified family $\bar{\mathbf{D}}^{M,N}$. Then, we evaluate the likelihood of the misspecified family at the true data $\mathsf{Z}_{M,N}$. Finally, we maximize this approximate likelihood.

We assume that the laws of the true family $\mathbf{D}^{M,N}$ are equivalent; i.e., for $\theta, \theta' \in \Theta$, $\mathbb{D}_\theta^{M,N}$ and $\mathbb{D}_{\theta'}^{M,N}$ are equivalent laws. We make the same equivalence assumption for the laws in the misspecified family $\bar{\mathbf{D}}^{M,N}$; i.e., $\bar{\mathbb{D}}_\theta^{M,N}$ and $\bar{\mathbb{D}}_{\theta'}^{M,N}$ are equivalent for $\theta, \theta' \in \Theta$. For $\theta \in \Theta$ and $d \in \mathcal{D}$, the likelihood function for the true family $\mathbf{D}^{M,N}$ is defined as

$$\mathcal{L}_{M,N}(\theta \,|\, d) = \left. \frac{\mathrm{d}\mathbb{D}_\theta^{M,N}}{\mathrm{d}\mathbb{D}_{\theta_0}^{M,N}} \right|_{\mathsf{D}_{M,N} = d} > 0.$$

The likelihood function for the misspecified family $\bar{\mathbf{D}}^{M,N}$ is

$$\bar{\mathcal{L}}_{M,N}(\theta \,|\, d) = \left. \frac{\mathrm{d}\bar{\mathbb{D}}_\theta^{M,N}}{\mathrm{d}\bar{\mathbb{D}}_{\theta_0}^{M,N}} \right|_{\bar{\mathsf{D}}_{M,N} = d} > 0.$$

An *approximate likelihood estimator* $\theta_{M,N}^A$ is defined as a solution to the optimization problem

$$\theta_{M,N}^A \in \arg\max_{\theta \in \theta} \log \mathcal{L}_{M,N}^A(\theta), \tag{11}$$

where

$$\mathcal{L}_{M,N}^A(\theta) = \bar{\mathcal{L}}_{M,N}(\theta \,|\, \mathsf{D}_{M,N}) \tag{12}$$

is an *approximate likelihood*.

Figure 1 illustrates several features of our estimation procedure. We have $\mathsf{D}_{M,N} \to \mathsf{D}_{M,\infty}$ in distribution and $\bar{\mathsf{D}}_{M,N} \to \mathsf{D}_{M,\infty}$ almost surely as $N \to \infty$. The latter fact implies that the misspecified likelihood function converges pointwise to the limiting likelihood function; i.e., $\bar{\mathcal{L}}_{M,N}(\theta|d) \to \mathcal{L}_{M,\infty}(\theta|d)$ for almost all parameters $\theta \in \Theta$ and data realizations $d \in \mathcal{D}$ as $N \to \infty$, where

$$\mathcal{L}_{M,\infty}(\theta \,|\, d) = \left. \frac{\mathrm{d}\mathbb{D}_\theta^{M,\infty}}{\mathrm{d}\mathbb{D}_{\theta_0}^{M,\infty}} \right|_{\mathsf{D}_{M,\infty} = d} \geq 0$$

and $\mathbb{D}_\theta^{M,\infty}$ is the $\mathbb{P}_\theta$-law of the limiting data $\mathsf{D}_{M,\infty}$. Note that when $\mathcal{L}_{M,\infty}(\theta|d) = 0$, the limiting law $\mathbb{D}_\theta^{M,\infty}$ is absolutely continuous with respect to the true limiting law $\mathbb{D}_{\theta_0}^{M,\infty}$, but not vice versa. As a result, the limiting laws are not necessarily equivalent.

The pointwise convergence of the misspecified likelihood tells us that the misspecified family will become correctly specified as $N \to \infty$. Consequently, the approximate likelihood will lie close to the limiting likelihood for large but finite N. Thus, we can expect that the approximate likelihood estimator will be highly accurate when both M and N are large.

4. Asymptotic properties of parameter estimators for large systems

We study the asymptotic properties of the approximate likelihood estimator $\theta_{M,N}^A$ as $M \to \infty$ and $N \to \infty$. It is well known that $\theta_{M,N}^A$ is not consistent if we only let $M \to \infty$ because the family $\bar{\mathbf{D}}^{M,N}$

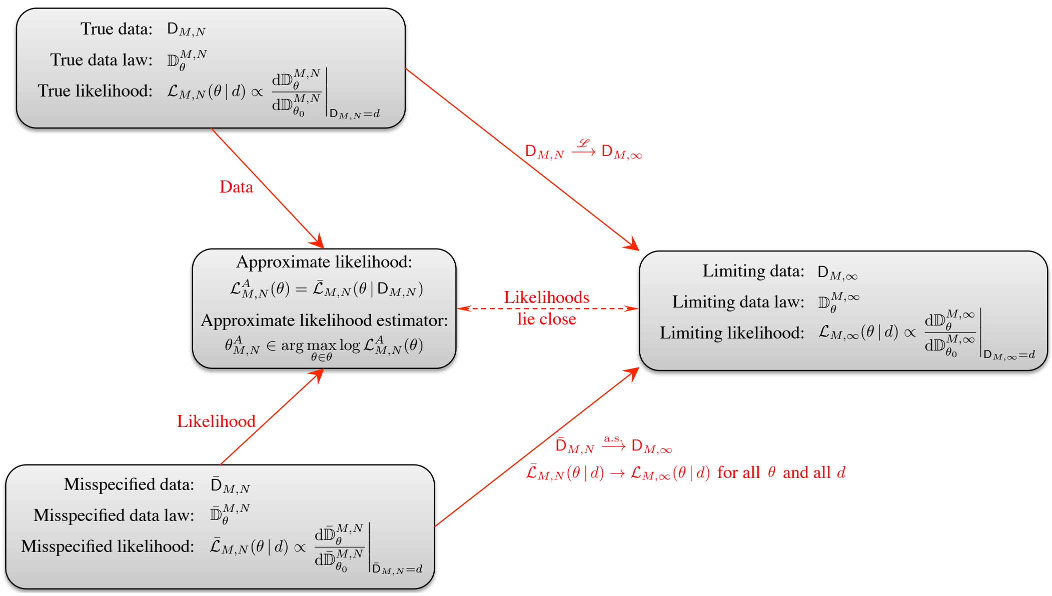

FIGURE 1. Data, data law, and likelihood functions for the true, the limiting, and the misspecified models.

is misspecified for finite N.[4] By letting $N \to \infty$ and $M \to \infty$ simultaneously, though, we can exploit the fact that the misspecified family becomes correctly specified in the limit when the system grows infinitely large. Consequently, we can expect that the approximate estimator will behave as a true maximum likelihood estimator when the system becomes large and more data becomes available.

The following proposition derives conditions for consistency and asymptotic efficiency of an approximate likelihood estimator. The conditions we formulate ensure that the approximate likelihood has a maximizer in a small neighborhood of the true parameter with large probability for large values of M and N. They also imply that an approximate likelihood estimator is asymptotically efficient in the sense that it attains the Cramér-Rao bound in the limit $M \to \infty$ and $N \to \infty$. The proposition follows the arguments in Bar-Shalom (1971) for deriving consistency and asymptotic efficiency for maximum likelihood estimators based on serially dependent observations of the data.

Proposition 4.1. *Suppose the following conditions hold:*

(A1) The mapping $\theta \mapsto \mathcal{L}^A_{M,N}(\theta)$ is almost surely three times continuously differentiable for all $M, N \geq 1$.

(A2) For any $\theta \in \Theta^\circ$, there exists a sequence $(C_{M,N})_{M,N\geq 1}$ of positive and finite constant such that the following limits hold in $\mathbb{P}_\theta$-probability as $M \to \infty$ and $N \to \infty$:

$$C_{M,N} \nabla \log \mathcal{L}^A_{M,N}(\theta) \to 0 \quad \text{and} \quad C_{M,N} \frac{\nabla^2 \mathcal{L}^A_{M,N}(\theta)}{\mathcal{L}^A_{M,N}(\theta)} \to 0.$$

(A3) For any $\theta \in \Theta^\circ$, the following matrix limit exists in $\mathbb{P}_\theta$-probability and is deterministic and positive definite:

$$\Sigma_\theta = \lim_{M,N\to\infty} C_{M,N} \nabla \log \mathcal{L}^A_{M,N}(\theta)^\top \nabla \log \mathcal{L}^A_{M,N}(\theta).$$

(A4) There exists a bounded-in-probability random variable H of the same dimension as $\nabla^3 \log \mathcal{L}^A_{M,N}$ such that almost surely for any $M, N \geq 1$ and $\theta \in \Theta^\circ$,

$$\left| C_{M,N} \nabla^3 \log \mathcal{L}^A_{M,N}(\theta) \right| \leq H.$$

The above inequality is understood componentwise.

---

[4]See Ogata (1980), White (1982), Gourieroux, Monfort & Trognon (1984), and McKeague (1984). Under suitable conditions, the approximate MLE $\theta^A_{M,N}$ converges in probability as $M \to \infty$ to the minimizer of the Kullback-Leibler divergence of the family $\bar{\mathbf{D}}^{M,N}$ relative to the true law $\mathbb{D}^{M,N}_{\theta_0}$. Thus, $\theta^A_{M,N}$ can be viewed as a minimum distance estimator as in Millar (1984).

Then any approximate likelihood estimator $\theta^A_{M,N}$ is consistent; i.e.,

$$\theta^A_{M,N} \to \theta_0 \quad \text{in} \quad \mathbb{P}_{\theta_0} - \text{probability}$$

as $M \to \infty$ and $N \to \infty$. In addition, an approximate likelihood estimator is asymptotically efficient in the sense that

$$\operatorname{Var}\left(C^{-1/2}_{M,N}\left(\theta^A_{M,N} - \theta_0\right)\right) \to \Sigma^{-1}_{\theta_0}$$

as $M \to \infty$ and $N \to \infty$.

Assumption (A1) is a standard differentiability assumption. Assumption (A4) is a standard boundedness assumption that ensures that a linear-quadratic Taylor expansion of the approximate likelihood around θ_0 is sufficiently accurate. Assumption (A3) is also standard and ensures that the asymptotic variance-covariance matrix of an approximate likelihood estimator is well-defined. Finally, Assumption (A2) is an identifiability condition. This condition ensures that the true parameter can be recovered given an infinite amount of data of an infinitely large system.

We remark that Proposition 4.1 only derives the asymptotic variance-covariance matrix of the approximate likelihood estimator $\theta^A_{M,N}$. We intentionally do not derive the asymptotic distribution of

$$C^{-1/2}_{M,N}\left(\theta^A_{M,N} - \theta_0\right)$$

with the goal of keeping the result as general as possible. In order to obtain asymptotic normality, for example, a necessary condition is that

$$C^{1/2}_{M,N} \nabla \log \mathcal{L}^A_{M,N}(\theta_0)$$

converges in $\mathbb{P}_{\theta_0}$-distribution to a normal random variable. Sufficient conditions for convergence in distribution to a normal random variable are formulated in standard central limit theorems; see Hall & Heyde (1980) for example. We refer to Jeganathan (1995) for alternative asymptotic distribution results.

A surprising implication of Proposition 4.1 is that one can obtain consistent and asymptotically efficient approximate likelihood estimators for large systems from the misspecified family even when no such estimators exist for the true family. Proposition 4.1 does not impose any conditions on the finite system (1) nor on the finite-system law $\mathbb{D}^{M,N}_\theta$. Informally speaking, when the system is large, one can neglect the finite-system law $\mathbb{D}^{M,N}_\theta$, replace it by the approximate law $\bar{\mathbb{D}}^{M,N}_\theta$, and obtain accuracate parameter estimators.

5. Discrete Observations of Z^N and continuous observations of X

In this section, we analyze the misspecified maximum likelihood estimation problem for the system (1) when the data $\mathsf{D}_{M,N}$ contains discrete observations of Z^N and continuous observations of X. In other words, $\mathsf{D}_{M,N} = (\mathsf{Z}_{M,N}, \mathsf{X}_M)$ for $\mathsf{Z}_{M,N} = (Z^N_{t_0}, Z^N_{t_1}, \ldots, Z^N_{t_M})$ and $\mathsf{X}_M = (X_t)_{t \in [0,t_M]}$.

Assume that the $\mathbb{P}_\theta$-law of the path X_M is absolutely continuous with respect to the Wiener measure $\mathbb{W}$ on $(C([0,t_M]), \sigma(C([0,t_M])))$. The approximate log-likelihood $\log \mathcal{L}^A_{M,N}(\theta)$ is proportional to

$$\begin{aligned} \ell^N_\theta(\mathsf{Z}_{M,N}, \mathsf{X}_M) &= \int_0^{t_M} a(X_u;\alpha)\mathrm{d}X_u - \frac{1}{2}\int_0^{t_M} a^2(X_u;\alpha)du + \log \bar{p}^{N,c}_\theta(\mathsf{Z}_{M,N}|\mathsf{X}_M) \\ &= \int_0^{t_M} a(X_u;\alpha)\mathrm{d}X_u - \frac{1}{2}\int_0^{t_M} a^2(X_u;\alpha)du + \sum_{m=1}^{M} \log \bar{p}^{N,c}_{m,\theta}(Z^N_{t_m}|\mathsf{X}_M, \mathsf{Z}_{m-1,N}), \end{aligned} \tag{13}$$

where $\bar{p}^{N,c}_\theta$ is the conditional density of $\bar{\mathsf{Z}}_{M,N}$ given the path X_M. In addition, we define $\bar{p}^{N,c}_{m,\theta}(\cdot|x,z)$ as the conditional density of $\bar{Z}^N_{t_m}$ given $\mathsf{X}_M = x$ and $\bar{\mathsf{Z}}_{m-1,N} = z$. These conditional densities are Gaussian. It follows from equation (8) that the mean for the multivariate Gaussian density $\bar{p}^{N,c}_\theta$ is simply the law of large numbers $\left(\langle f, \bar{\mu}_{t_0}\rangle, \ldots, \langle f, \bar{\mu}_{t_M}\rangle\right)$ for the path X_M. Its covariance is the covariance of $\left(\frac{1}{N}\left\langle f, \bar{\Xi}_{t_0}\right\rangle, \ldots, \frac{1}{N}\left\langle f, \bar{\Xi}_{t_M}\right\rangle\right)$, again conditional on X_M. As a result, in the limit $N \to \infty$, the density $\bar{p}^{N,c}_\theta$ becomes a delta function centered at the law of large numbers.

The first two terms in (13) can be calculated using standard methods; for instance, see Feigin (1976). Although the density $\bar{p}^{N,c}_{m,\theta}$ is Gaussian, calculating its mean and variance is a not a trivial task because it requires the solution of the stochastic evolution equations (5). Section 5.1 develops a semi-analytical method to calculate the mean and variance of $\bar{p}^{N,c}_{m,\theta}$. We remark that the methodology in Section 5.1 does not require any Monte-Carlo simulation; the computed means and variances are deterministic. As a result, the likelihood can be maximized with a number of standard optimization methods (for example, gradient descent).

5.1. **Numerical methods for evaluating the conditional densities.** We develop a numerical approach to calculate the mean and variance of the conditional densities $\bar{p}^{N,c}_{m,\theta}$. Evaluating the mean and the variance for $\bar{p}^{N,c}_{m,\theta}$ requires the numerical solution of the stochastic evolution equations for the law of large numbers and the central limit theorem. The LLN stochastic evolution equation can be solved by a number of standard PDE tools, including: finite difference, Galerkin methods, and method of moments. The numerical solution of the stochastic evolution equation for the central limit theorem is a much more challenging numerical problem since it has a distribution-valued solution. This means that a pointwise, differentiable solution may not exist. In order to numerically solve the fluctuation equation, one must reduce it to a system of SDEs. The distribution for the system of SDEs can then be semi-analytically calculated, conditional upon the path X_M. The semi-analytical solution exploits the fact that $\bar{\Xi}$ is conditionally Gaussian given X_M.

In particular, we highlight two numerical schemes which can be used to solve the fluctuation limit. One numerical approach to the fluctuation equation is to use a finite volume method which divides up the space into "boxes" or computational cells. One can then solve for the probability mass in each cell. Such finite volume schemes have been investigated in Donev, Vanden-Eijnden, Garcia & Bell (2010). Alternatively, in some cases, a method of moments can be used to solve for a collection of moments of the solution to the stochastic evolution equation. The moments are integrals of functions against the solution of the stochastic evolution equation for the central limit theorem (as well as the law of large numbers). This can be particularly useful when the quantity one is interested in is one of the moments. This happens to be the case in Example 2.3 where the limiting loss is one of the moments. For that model, very accurate solutions can be achieved even with only a few moments, making the approach highly tractable.

Both the method of moments and the finite volume method reduce the stochastic evolution equation to a system of SDEs of the form

$$d\bar{\mathbf{v}}_t = A_1(X_t)\bar{\mathbf{v}}_t dt + A_2(X_t)\bar{\mathbf{v}}_t dX_t + Bd\mathbf{B}_t, \tag{14}$$

where $\bar{\mathbf{v}}_t : [0,T]\times\Omega \to \mathbb{R}^K$, $A_1, A_2 : [0,T]\times\Omega \to \mathbb{R}^{K\times K}$, $B : [0,T]\times\Omega \to \mathbb{R}^{J\times K}$, and $\mathbf{B}_t \in C_{\mathbb{R}^J}([0,T])$ is, conditional upon X, a vector of Brownian motions with covariation matrix $d[\mathbf{B}_t, \mathbf{B}_t] = \Sigma_B(t)dt$. The random vector $\bar{\mathbf{v}}_{t=0}$ is mean-zero Gaussian with covariance Σ_0.

Recall that the approximate observation process is defined as $\bar{Z}^N_t = \langle f, \bar{\mu}_t\rangle + \frac{1}{\sqrt{N}}\left\langle f, \bar{\Xi}_t\right\rangle$. In the case of the method of moments, $\left(\bar{\mathbf{v}}_t\right)_k = \left\langle f_k, \bar{\Xi}_t\right\rangle$ for a sequence of functions $f_1, \ldots, f_K$. The method of moments is especially advantageous when one of the f_k equals f. Then, $\bar{Z}^N_t = \langle f, \bar{\mu}_t\rangle + \frac{1}{\sqrt{N}}\left(\bar{\mathbf{v}}_t\right)_k$ for the k such that $f_k = f$. In the finite volume method, one tracks the average of $\bar{\Xi}_t$ in a series of computational cells. For notational convenience, we present the method in one dimension.

$$\left(\bar{\mathbf{v}}_t\right)_k = \frac{1}{\Delta y}\int_{(k-1)\Delta y}^{k\Delta y} \bar{\Xi}_t(dy).$$

The stochastic evolution equation for $\bar{\Xi}_t$ can then be reduced to a system of SDEs for the finite volume averages $\bar{\mathbf{v}}_t$. Moreover, the integral $\left\langle f, \bar{\Xi}_t\right\rangle$ can be evaluated as a weighted average of the finite volume averages.

In either case, the approximate observation process $\bar{Z}^N_t$ is a linear transformation of the solution $\bar{\mathbf{v}}_t$ (up to the numerical error of the method of moments or finite volume scheme).

$$\bar{Z}^N_t = \langle f, \bar{\mu}_t\rangle + \frac{1}{\sqrt{N}} H\bar{\mathbf{v}}_t,$$

where $H \in \mathbb{R}^{1\times K}$.

Note that conditional on the path X_M, $\langle f, \bar{\mu}_t\rangle$ is deterministic and one therefore directly obtains the quantity $H\bar{\mathbf{v}}_t$ from observing $\bar{Z}^N_t$ at times $t_1, \ldots, t_M$. Furthermore, conditional on the path X_M, the solution of the SDE system (14) is Gaussian. Therefore, one can completely specify its conditional finite-dimensional distributions by finding the mean and covariance, which can be done in a straightforward manner using the fundamental solution for the SDE system (14). The fundamental solution $\Psi(s,t) : [s,T]\times\Omega \to \mathbb{R}^{K\times K}$ for the SDE system (14) satisfies

$$\begin{aligned} d\Psi(s,t) &= A_1\Psi(s,t)dt + A_2\Psi(s,t)dX_t, \quad t > s, \\ \Psi(s,s) &= I, \end{aligned}$$

where I is the identity matrix. The solution $\bar{\mathbf{v}}_t$ can be written as

$$\bar{\mathbf{v}}_t = \Psi(s,t)\bar{\mathbf{v}}_s + \Psi(s,t)\int_s^t \Psi^{-1}(s,u)Bd\mathbf{B}(u).$$

Conditional on X_M and $\bar{\mathbf{v}}_s \sim \mathcal{N}(\mu_s, \Sigma_s)$, $\bar{\mathbf{v}}_t \sim \mathcal{N}(\Psi(s,t)\mu_s, \Sigma(s,t))$ where

$$\Sigma(s,t) = \Psi(s,t)\Sigma_s\Psi(s,t)^\top + \Psi(s,t)\Big[\int_s^t \Psi^{-1}(s,u)B\Sigma_B(u)\big(\Psi^{-1}(s,u)B\big)^\top du\Big]\Psi(s,t)^\top.$$

Conditional on X_M, the observed process $\bar{Z}_t^N$ is also Gaussian with variance $\frac{1}{N}H\Sigma(0,t)H^\top$ and mean $\langle f, \bar{\mu}_t\rangle$. Given X_M, one can easily solve for Ψ and calculate the covariance $\Sigma(s,t)$.

The terms $\bar{p}_{m,\theta}^{N,c}(Z_{t_m}^N|\mathsf{X}_M, \mathsf{Z}_{m-1,N})$ can now be calculated semi-analytically due to $\bar{Z}_t^N$ being a conditionally Gaussian process. Let $\pi^{m|m}$ be the conditional distribution of $(\bar{Z}_{t_m}, \bar{\mathbf{v}}_{t_m})$ given $(\mathsf{X}_M, \bar{\mathsf{Z}}_{m,N})$ and let $\pi^{m+1|m}$ be the conditional density of $(\bar{Z}_{t_{m+1}}^N, \bar{\mathbf{v}}_{t_{m+1}})$ given $(\mathsf{X}_M, \bar{\mathsf{Z}}_{m,N})$. These correspond to Gaussian distributions and are therefore completely defined by their means and covariances. If one can calculate $\pi^{m|m-1}$, then one can immediately calculate $\bar{p}_{m,\theta}^{N,c}$ since the latter is simply the marginal density of the former.

We next develop a procedure to calculate in closed-form the means and covariances corresponding to $\{\pi^{m|m-1}\}_{m=1,\ldots,M}$. The method relies upon the solution for Ψ and is hence semi-analytic.

Theorem 5.1. *Let $\pi^{m|m}$ be the conditional distribution of $(\bar{Z}_{t_m}, \bar{\mathbf{v}}_{t_m})$ given $(\mathsf{X}_M, \bar{\mathsf{Z}}_{m,N})$ and let $\pi^{m+1|m}$ be the conditional density of $(\bar{Z}_{t_{m+1}}^N, \bar{\mathbf{v}}_{t_{m+1}})$ given $(\mathsf{X}_M, \bar{\mathsf{Z}}_{m,N})$. The filters can be updated according to the steps described below.*

Algorithm:

(i) $\pi^{1|0} \sim \mathcal{N}(\mu^{1|0}, \Sigma^{1|0})$ *where* $\mu^{1|0} = \big(\langle f, \bar{\mu}_{t_1}\rangle, 0, \ldots, 0\big)$ *and* $\Sigma_{1,1}^{1|0} = \frac{1}{N}H\Sigma(0,t_1)H^\top$, $\Sigma_{1:1+K,2:1+K}^{1|0} = \Sigma(0,t_1)$, *and* $\Sigma_{1,2:1+K}^{1|0} = H$.

(ii) For $m = 1, \ldots, M-1$

(a) $\pi_{m|m} = \mathcal{N}(\mu^{m|m}, \Sigma^{m|m})$ *where* $\pi_{m|m}$ *is a degenerate Gaussian with* $\mu_1^{m|m} = \bar{Z}_{t_m}^N$, $\Sigma_{1,1}^{m|m} = 0$, *and*

$$\begin{aligned} \mu_{2:1+K}^{m|m} &= \mu^{m|m-1} + \Sigma_{2:1+K,1}^{m|m-1}\big(\Sigma_{1,1}^{m|m-1}\big)^{-1}(\bar{Z}_{t_m}^N - \mu_1^{m|m-1}), \\ \Sigma_{2:1+K,2:1+K}^{m|m} &= \Sigma_{2:K,2:1+K}^{m|m-1} - \Sigma_{2:1+K,1}^{m|m-1}\big(\Sigma_{1,1}^{m|m-1}\big)^{-1}\Sigma_{1,2:1+K}^{m|m-1}. \end{aligned}$$

(b) $\pi_{m+1|m} = \mathcal{N}(\mu^{m+1|m}, \Sigma^{m+1|m})$ *where*

$$\begin{aligned} \mu_{2:1+K}^{m+1|m} &= \Psi(t_m, t_{m+1})\mu_{2:1+K}^{m|m}, \\ \mu_1^{m+1|m} &= \big\langle f, \bar{\mu}_{t_{m+1}}\big\rangle + \frac{1}{\sqrt{N}}H\mu_{2:1+K}^{m+1|m}, \\ \Sigma_{2:1+K,2:1+K}^{m+1|m} &= \Psi(t_m, t_{m+1})\Sigma_{2:1+K,2:1+K}^{m+1|m}\Psi(t_m, t_{m+1})^\top \\ &+ \Psi(t_m, t_{m+1})\Big[\int_{t_m}^{t_{m+1}} \Psi^{-1}(t_m,u)B\Sigma_B(u)\big(\Psi^{-1}(t_m,u)B\big)^\top du\Big]\Psi(t_m, t_{m+1})^\top, \\ \Sigma_{1,1}^{m+1|m} &= \frac{1}{N}H\Sigma_{2:1+K,2:1+K}^{m+1|m}H^\top. \end{aligned}$$

After the filters $\pi^{m+1|m}$ and $\pi^{m|m}$ have been calculated for $m = 0, \ldots, M-1$, one can easily find the desired terms $\{p_{m,\theta}^{N,c}\}_{m=1,\ldots,M}$ and evaluate the approximate likelihood (13). Note that the filter is very similar to the Kalman filter; the main difference is that the partial observations of the hidden state are exact (i.e., no observational noise).

5.2. **Finite sample properties of parameter estimators for large systems.** We prove finite sample asymptotics for the approximate estimator in the case where Z^N is observed discretely and X is observed completely. Here, "finite sample" means observations are made over a finite time period; the system can still grow large in N. Given a sample of data on the interval $[0, t_M]$, we are able to show, under mild technical conditions, that the approximate estimator for the parameter β is consistent and asymptotically normal as only $N \to \infty$ and $M < \infty$. These are very useful results which complement the results in Section 4 and allow the statistician to evaluate the accuracy of the approximate likelihood estimator given a finite data sample of a large system. In addition, it is shown that the approximate estimator converges to a weighted least squares estimator in the limit $N \longrightarrow \infty$.

5.2.1. *Relation to Least Squares.* The approximate estimator $\theta^A_{M,N}$ has an elegant relationship with least squares estimation in the limit $N \longrightarrow \infty$. For very large systems, this makes estimation particularly tractable.

The approximate estimator can alternatively be written as

$$\theta^A_{M,N} = (\alpha^A_M, \beta^A_{M,N}) \in \arg\max_{\theta \in \Theta} \frac{1}{N} \ell^N_\theta(\mathsf{Z}_{M,N}, \mathsf{X}_M).$$

Note that the estimators $\alpha^A_{M,N}$ and $\beta^A_{M,N}$ can be calculated separately.

$$\begin{aligned} \alpha^A_{M,N} &\in \arg\max\nolimits_{\alpha \in \Theta_\alpha} \tfrac{1}{N} \ell^N_\alpha(\mathsf{X}_M), \\ \beta^A_{M,N} &\in \arg\max\nolimits_{\beta \in \Theta_\beta} \tfrac{1}{N} \ell^N_\beta(\mathsf{Z}_{M,N}, \mathsf{X}_M), \end{aligned}$$

where $\ell^N_\alpha \equiv \int_0^{t_M} a(X_u;\alpha)\mathrm{d}X_u - \frac{1}{2}\int_0^{t_M} a^2(X_u;\alpha)du$ and $\ell^N_\beta \equiv \log \bar{p}^{N,c}_\theta(\mathsf{Z}_{M,N}|\mathsf{X}^c_M)$. In this section, we will also sometimes explicitly write the dependence of the estimators on the data:

$$\theta^A_{M,N} \equiv \theta^A_{M,N}(\mathsf{Z}^A_{M,N}, \mathsf{X}_M) = \big(\alpha^A_M(\mathsf{X}_M), \beta^A_M(\mathsf{Z}_{M,N}, \mathsf{X}_M)\big).$$

Let m^M_β and $\frac{1}{N}\Sigma^M_\beta$ be the mean and covariance of $\bar{Z}_{M,N}$ under the measure $\mathbb{D}^{M,N}_\theta$. The mean m^M_β is simply the law of large numbers $\langle f, \bar{\mu}\rangle$ while the covariance Σ^M_β is the covariance of the fluctuation process $\langle f, \bar{\Xi}\rangle$; neither depend upon N. Without the scaling factor $\frac{1}{N}$, the limiting approximate log-likelihood would be singular, making it difficult to prove the desired asymptotic properties. With the scaling factor, the limiting approximate log-likelihood converges to a very simple function (in fact, just a Gaussian density).

Lemma 5.2. *Assume the following conditions:*

- *(B1) The parameter sets Θ_α and Θ_β are compact.*
- *(B2) The functions $\beta \mapsto m^M_\beta$ and $\beta \mapsto \Sigma^M_\beta$ are almost surely continuous on Θ_β.*
- *(B3) The matrix Σ^M_β is almost surely positive definite for all $\beta \in \Theta_\beta$.*

Then, for each $z \in \mathbb{R}^M$ and $x \in C([0, t_M])$, $\theta^A_{M,N}(z,x) \in \arg\max_{\theta\in\Theta} \frac{1}{N}\ell^N_\theta(z,x)$ converges to $\theta^A_{M,\infty}(z,x) \in (\alpha^A_M(x), \beta^A_{M,\infty}(z,x))$ where

$$\begin{aligned} \alpha^A_M(x) \quad &\in \quad \arg\max_{\alpha \in \Theta_\alpha} \int_0^{t_M} a(x_u;\alpha)\mathrm{d}x_u - \frac{1}{2}\int_0^{t_M} a^2(x_u;\alpha)du, \\ \beta^A_{M,\infty}(z,x) \quad &\in \quad \arg\max_{\beta\in\Theta_\beta} -\frac{1}{2}(z - m^M_\beta)^\top (\Sigma^M_\beta)^{-1}(z - m^M_\beta). \end{aligned}$$

The limiting estimator is a weighted least squares regression. The weights are determined by the fluctuation limit. Typically, there will be a tradeoff between increasing the magnitude of the covariance matrix Σ^M_β and making the mean m^M_β closer to the observation z.

5.2.2. *Convergence of Estimators.* We next show that the estimator $\beta^A_{M,N}(\mathsf{Z}_{M,N}, \mathsf{X}_M)$ is consistent as $N \longrightarrow \infty$ and $M < \infty$; i.e., $\beta^A_{M,N} \to \beta_0$ in $\mathbb{P}_{\theta_0}$-probability as $N \to \infty$.

Theorem 5.3. *Assume the following conditions:*

- *(B1) The parameter sets Θ_α and Θ_β are compact.*
- *(B2) The functions $\beta \mapsto m^M_\beta$ and $\beta \mapsto \Sigma^M_\beta$ are almost surely continuous on Θ_β.*
- *(B3) The matrix Σ^M_β is almost surely positive definite for all $\beta \in \Theta_\beta$.*

(B4) There exists an $M_0 < \infty$ such that for almost every realization of X_M, the map $\mathsf{Z}_{M,\infty} : \Theta_\beta \mapsto \mathbb{R}^{M+1}$ is one-to-one for $M > M_0$.

If conditions (B1), (B2), and (B3) hold, the approximate estimator $\beta^A_{M,N}(\mathsf{Z}_{M,N}, \mathsf{X}_M)$ converges in probability to a limit point in the set $\beta^A_{M,\infty}(\mathsf{Z}_{M,\infty}, \mathsf{X}_M)$. Moreover, if condition (B4) holds and $M > M_0$, the approximate estimator $\beta^A_{M,N}(\mathsf{Z}_{M,N}, \mathsf{X}_M)$ converges in probability to the true parameters β_0 as $N \longrightarrow \infty$.

The assumptions required for Theorem 5.3 are mild. (B1), (B2), and (B3) are standard conditions. Condition (B4) implies that the true parameter β_0 is almost surely identifiable from the limiting likelihood $\mathcal{L}_{M,\infty}(\theta|\mathsf{D}_{M,\infty})$, which is a necessary condition (and a standard assumption) for the consistency of likelihood estimators. Conditional on a path of X, the limiting process $\mathsf{Z}_{M,N}$ is deterministic. If the map $\mathsf{Z}_{M,\infty} : \Theta_\beta \mapsto \mathbb{R}^{M+1}$ is one-to-one for a realization of the data X_M, then the corresponding realization of the limiting likelihood $\mathcal{L}_{M,\infty}(\theta|\mathsf{D}_{M,\infty})$ is proportional to a delta function centered at β_0 and the true parameter β_0 is clearly identifiable. If the sample size M_0 is larger than the degrees of freedom for the function $\mathsf{Z}_{M_0,\infty} : \Theta_\beta \mapsto \mathbb{R}^{M+1}$, then the true parameter β_0 is identifiable from the limiting likelihood. For instance, if there is only one parameter and $\mathsf{Z}_{M,\infty}$ is monotonic in that parameter, $M_0 = 0$. In practice, one can numerically test if Condition (B4) holds for the limiting process $\mathsf{Z}_{M_0,\infty}$ by numerically searching to see if there exist multiple sets of parameters producing the same solution. An example of a system where assumptions (B1), (B2), (B3), and (B4) can be (analytically) verified to hold is (21) in Section 8.

Remark 5.4. *If there are other parameters β besides β_0 which can produce the limiting data $\mathsf{Z}_{M,\infty}$, one cannot necessarily expect the approximate estimator to converge in probability to the "true parameter" β_0. In this case, one can only be assured that the approximate estimator converges in probability to $\beta^A_{M,\infty}(\mathsf{Z}_{M,\infty}, \mathsf{X}_M)$; i.e., the set of β which can produce the limiting data $\mathsf{Z}_{M,\infty}$. The uniqueness of the parameter which can produce the limting data $\mathsf{Z}_{M,\infty}$ therefore determines the identifiability of the true parameters β_0.*

Theorem 5.5. *Suppose that the conditions of Theorem 5.3 hold. In addition, suppose that the functions $\beta \mapsto m^M_\beta$ and $\beta \mapsto \Sigma^M_\beta$ are almost surely twice continuously differentiable on Θ_β. Then, the approximate estimator $\beta^A_{M,N}$ has the central limit theorem:*

$$\sqrt{N}(\beta^A_{M,N} - \beta_0) \xrightarrow{d} (\nabla^2 \ell^\infty_{\beta_0})^{-1} Q_{\beta_0},$$

where $Q_{\beta_0} \equiv \sum_{i,j} \nabla(m^M_{\beta_0})_i ((\Sigma^M_{\beta_0})^{-1})_{i,j} \left\langle f, \bar{\Xi}_{t_j} \right\rangle$.

The central limit theorem is given in closed-form in terms of $\bar{\mu}$ and $\bar{\Xi}$; standard errors are therefore easy to compute for the approximate estimator.

6. Discrete Observations of Z^N and X

In this section, we propose a methodology to carry out approximate maximum likelihood estimation in the observational regime where both Z^N and X are observed discretely. The data set available for inference is given by $\mathsf{D}_{M,N} = (\mathsf{Z}_{M,N}, \mathsf{X}_M)$, where $\mathsf{Z}_{M,N} = (Z^N_{t_1}, \ldots, Z^N_{t_M})$ and $\mathsf{X}_M = (X_{t_1}, \ldots, X_{t_M})$. The data $\mathsf{Z}_{M,N}$ is measurable on $(\mathbb{R}^M, \mathcal{B}^M)$ and X_M is measurable on $(\mathbb{R}^M, \mathcal{B}^M)$, where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}$.

As before, assume that the $\mathbb{P}_\theta$-law of a path of X is absolutely continuous with respect to the Wiener measure $\mathbb{W}$ on $(C([0, t_M), \mathbb{R}), \sigma(C([0, t_M), \mathbb{R})))$. In addition, assume that transition densities exist for the processes $\bar{Z}^N$ and Z^∞ relative to the Lebesgue measure $\mathbb{L}^M$ on $(\mathbb{R}^M, \mathcal{B}^M)$. Sufficient conditions are given in Feigin (1976), Komatsu & Takeuchi (2001), and Takeuchi (2002). For $\theta \in \Theta$ and $d = (z, x) \in \mathbb{R}^M \times \mathbb{R}^M$, the true likelihood function satisfies

$$\mathcal{L}_{M,N}(\theta \,|\, d) \propto p^N_\theta(z|x) p_\alpha(x)$$

and the likelihood function for the misspecified model satisfies

$$\bar{\mathcal{L}}_{M,N}(\theta \,|\, d) \propto \bar{p}^N_\theta(z|x) p_\alpha(x).$$

Here, p_α is the density of X_M, and p^N_θ and $\bar{p}^N_\theta$ are the conditional densities of $\mathsf{Z}_{M,N}$ and $\bar{\mathsf{Z}}_{M,N}$ given X_M, respectively.

The weak convergence results imply that the limiting observation process $\bar{Z}_t^N$ is conditionally Gaussian given $(X_s)_{s\in[0,t]}$. Thus, the density $\bar{p}_\theta^N$ can be written more explicitly. Let $\mathsf{X}_M^c = (X_t : 0 \le t \le t_M)$ denote the complete path of the systematic process X until time t_M. Then,

$$\bar{p}_\theta^N(z|x) = \mathbb{E}_\theta\left[\bar{p}_\theta^{N,c}(z|\mathsf{X}_M^c)\middle|\,\mathsf{X}_M = x\right] \tag{15}$$

where $z \in \mathbb{R}^M$, $x \in \mathbb{R}^M$, and $\bar{p}_\theta^{N,c}(z|\mathsf{X}_M^c)$ is the conditional density of $\bar{\mathsf{Z}}_{M,N}$ under $\bar{\mathbb{D}}_\theta^{M,N}$ given a complete path of X_M^c of the systematic process. Importantly, $\bar{p}_\theta^{N,c}(z|\mathsf{X}_M^c)$ is a Gaussian density so that $\bar{p}_\theta^N(z|\mathsf{X}_M)$ is an infinite mixture of Gaussian densities. We have already developed a numerical scheme in Section 5.1 to approximate the densities $\bar{p}_\theta^{N,c}(z|\mathsf{X}_M^c)$ in semi-analytical fashion. The quantity (15) can then, for example, be calculated by generating skeletons of X_M^c conditional on the observation X_M, semi-analytically evaluating $\bar{p}_\theta^{N,c}(z|\mathsf{X}_M^c)$ for each path, and taking the average across all paths. Skeletons of complete paths of X conditional on the observation X_M can be generated exactly, e.g., via the method in Beskos, Papaspiliopoulos & Roberts (2006). Therefore, the case with discrete observations of X builds upon the case with continuous observation of X.

6.1. **Evaluation of Estimators.** The approximate likelihood function is

$$\mathcal{L}_{M,N}^A \propto \bar{p}_\theta^N(\mathsf{Z}_{M,N}|\mathsf{X}_M)p_\alpha(\mathsf{X}_M). \tag{16}$$

The approximate likelihood can be easily evaluated for a fixed θ. The density $p_\alpha(\mathsf{X}_M)$ can be calculated using available methods from the literature (for instance, see Ait-Sahalia (2002)). The term $\bar{p}_\theta^N(\mathsf{Z}_{M,N}|\mathsf{X}_M)$ can be evaluated by simulating paths X_M^c from its conditional distribution given X_M, semi-analytically computing the density $\bar{p}_\theta^N(\mathsf{Z}_{M,N}|\mathsf{X}_M^c)$ for each path X_M^c, and then averaging over all the paths. The approximate likelihood (16) can then be evaluated as

$$\mathcal{L}_{M,N}^A \approx \frac{1}{L}\sum_{l=1}^{L}\bar{p}_\theta^N(\mathsf{Z}_{M,N}|\mathsf{X}_M^{c,l})p_\alpha(\mathsf{X}_M).$$

Of course, as $L \to \infty$, the average converges to the approximate likelihood $\mathcal{L}_{M,N}^A$. Since this method requires that (16) must be stochastically approximated, Monte Carlo error can cause numerical evaluations to be inconsistent from one step to the next. This can be potentially problematic when maximizing to find the approximate estimators. One can address this issue by using the EM algorithm, which we adapt to our setting in the next subsection.

6.2. **EM Algorithm.** The approximate likelihood (16) can also be maximized via the EM algorithm. The complete-path approximate log-likelihood $\ell_\theta^N(\mathsf{Z}_{M,N}, \mathsf{X}_M^c)$ (i.e., the approximate log-likelihood provided the complete path X_M^c) is given in equation (13). As described in Subsection 5.1, the complete log-likelihood $\ell_\theta^N(\mathsf{Z}_{M,N}, \mathsf{X}_M^c)$ can be computed semi-analytically. Let $Q(\theta', z|\theta) = \mathbb{E}_\theta[\ell_{\theta'}^N(\bar{\mathsf{Z}}_{M,N}, \mathsf{X}_M^c)|\bar{\mathsf{Z}}_{M,N} = z, \mathsf{X}_M]$. The EM algorithm then is

- E-step: compute $Q(\theta', \mathsf{Z}_{M,N}|\theta)$
- M-step: maximize $Q(\cdot, \mathsf{Z}_{M,N}|\theta)$.

These steps are repeated until the estimated parameters converge within some toleration. Convergence of the stochastic version of the EM algorithm above is demonstrated in Dembo & Zeitouni (1986). Nielsen (2000) derives conditions that ensure that the parameter estimator derived from the EM algorithm inherits the consistency and asymptotic normality properties of the approximate likelihood estimator $\theta_{M,N}^A$.

We propose to estimate the expectation in the E-step of the EM algorithm via Monte-Carlo simulation. To this extent, we generate skeletons of X_M^c conditional upon $\big(\mathsf{X}_M, \bar{\mathsf{Z}}_{M,N} = z\big)$. By taking advantage of the conditional Gaussian distribution of $\bar{Z}_\theta^{M,N}$, we calculate an estimator of $Q(\theta', z|\theta)$ via the following steps:

- Generate sample paths of $\mathsf{X}_M^{c,1}, \ldots, \mathsf{X}_M^{c,L}$ from the conditional law of X_M^c given X_M
- Compute the weights $w_\theta^1, \ldots, w_\theta^L$ where

$$w_\theta^l = \frac{\bar{p}_\theta^{N,c}(z|\mathsf{X}_M^{c,l})}{\sum_{k=1}^L \bar{p}_\theta^{N,c}(z|\mathsf{X}_M^{c,k})}.$$

- Calculate $Q(\theta', z|\theta)$:

$$Q(\theta', z|\theta) \approx \sum_{l=1}^{L} \ell_{\theta'}^{N}(z, \mathsf{X}_M^{c,l}) w_\theta^l,$$

where $\ell_{\theta'}^{N}(z, \mathsf{X}_M^{c,l})$ and the weights w_θ^l can be evaluated semi-analytically using the methodology in Section 5.1.

Proposition 6.1. *$\sum_{l=1}^{L} \ell_{\theta'}^{N}(z, \mathsf{X}_M^{c,l}) w_\theta^l$ converges almost surely to $Q(\theta, z|\theta')$.*

7. Continuous Observations of Z^N and X

In this section, we develop tools to evaluate the likelihood function in the case where both Z^N and X are observed continuously on $[0, t_M]$. The observed data is $\mathsf{D}_{M,N} = (\mathsf{Z}_{M,N}, \mathsf{X}_M)$ where $\mathsf{Z}_{M,N} = \big(Z_t^N : 0 \le t \le t_M\big)$ and $\mathsf{X}_M = \big(X_t : 0 \le t \le t_M\big)$.

For the remainder of this section, we assume that $\zeta = 0$ in the model (1). This restriction is necessary for our approach for continuously observed (Z, X) to work.[5] The strategy is the following: given the path X_M, $(\bar{Z}_t^N)_{t\in[0,t_M]}$ is a Gaussian process. If the measures for a Gaussian process and the Wiener process are equivalent, one can evaluate the Radon-Nikodym derivative of the measure for the Gaussian process with respect to the Wiener measure using the results from Shepp (1966). Therefore, we first transform $\bar{Z}_t^N$ into a process $\bar{Q}_t^N$ with quadratic variation t. Then, using Shepp's results, we find the Radon-Nikodym derivative and use it to form our approximate likelihood function.

Lemma 7.1. *Define m_t^β to be the mean of $\bar{Z}_t$ under the measure $\bar{\mathbb{D}}_\theta^{M,N}$, where as usual $\theta = (\alpha, \beta)$. There exists a deterministic process q_t^β such that $\bar{Q}_t^N = \sqrt{N} g_t^\beta(\bar{Z}_t^N - m_t^\beta)$ has quadratic variation t under $\bar{\mathbb{D}}_\theta^{M,N}$.*

Also, define $Q_t^N = \sqrt{N} g_t^\beta(Z_t^N - m_t^\beta)$. As usual, let $\mathbb{W}$ be the Wiener measure on the space of continuous paths $C([0, t_M])$. $\mathsf{Q}_{M,N}$ is the complete observation of Q_t^N, which is available due to the complete observation of Z_t^N. Similarly, $\bar{\mathsf{Q}}_{M,N}$ is the complete observation of $\bar{Q}_t^N$. Let $\bar{\mathbb{Q}}_\theta^{M,N}$ be the conditional distribution of $\bar{\mathsf{Q}}_{M,N}$ given X_M. Also, define $\mathbb{X}_\theta$ as the law of X and $\bar{\mathbb{G}}_\theta^{M,N}$ as the law of the data $(\bar{\mathsf{Q}}_{M,N}, \mathsf{X}_M)$.

Lemma 7.2. *The likelihood function for the misspecified model is proportional to the Radon-Nikodym derivative of the law of the process $(\bar{\mathsf{Q}}_{M,N}, X)$ with respect to the Wiener product measure $\mathbb{W} \times \mathbb{W}$:*

$$\frac{d\bar{\mathbb{D}}_\theta^{M,N}}{d\bar{\mathbb{D}}_{\theta_0}^{M,N}}\Big|_{(\bar{\mathsf{Z}}_{M,N},\mathsf{X}_M)} = \frac{d\bar{\mathbb{G}}_\theta^{M,N}}{d\bar{\mathbb{G}}_{\theta_0}^{M,N}}\Big|_{(\bar{\mathsf{Q}}_{M,N},\mathsf{X}_M)} \propto \frac{d\bar{\mathbb{G}}_\theta^{M,N}}{d(\mathbb{W}\times\mathbb{W})}\Big|_{(\bar{\mathsf{Q}}_{M,N},\mathsf{X}_M)} = \frac{d\bar{\mathbb{Q}}_\theta^{M,N}}{d\mathbb{W}}\Big|_{\bar{\mathsf{Q}}_{M,N}} \frac{d\mathbb{X}_\theta}{d\mathbb{W}}\Big|_{\mathsf{X}_M}.$$

Now, define $\Sigma^Q(s,t;\beta)$ to be the covariance for the process $\bar{Q}^N$ (conditional on X_M) and $K(s,t;\beta) = -\frac{\partial}{\partial t}\frac{\partial}{\partial s}\Sigma^Q(s,t;\beta)$. The covariance $\Sigma^Q(s,t;\beta)$ can be calculated using the methods in Section 5.1.

Proposition 7.3. *Assume the following conditions:*

(C1) $K(s,t;\beta)$ exists for almost every $(s,t) \in [0,t_M] \times [0,t_M]$, almost surely.

(C2) $K(s,t;\beta)$ is continuous and of trace class, almost surely.[6]

Then, it follows that

$$\frac{d\bar{\mathbb{Q}}_\theta^{M,N}}{d\mathbb{W}}\Big|_{\bar{\mathsf{Q}}_{M,N}=(q,x)} = d(1;\beta)^{-\frac{1}{2}} \exp[-\tfrac{1}{2}\textstyle\int_0^{t_M}\int_0^{t_M} H_1(s,u;\beta) dq_s dq_u], \tag{17}$$

where $d(\lambda;\beta)$ is the Fredholm determinant of K and $H_\lambda(s,u;\beta)$ is the Fredholm resolvent of K.

Define the misspecified log-likelihood function:

$$\ell_\theta^N(q,x) = \log(d(1;\beta)^{-\frac{1}{2}}) - \tfrac{1}{2}\textstyle\int_0^{t_M}\int_0^{t_M} H_1(s,u;\beta) dq_s dq_u + \log(\frac{d\mathbb{X}_\theta}{d\mathbb{W}}\Big|_x). \tag{18}$$

[5] Our approach involves scaling the observation process $\bar{Z}_t^N$ such that its quadratic variation is t. If $\zeta \neq 0$, it is not possible find such a scaling in the observation setting we have chosen.

[6] A bounded linear operator G in a Hilbert space H is of trace class if $\sum_k \left\langle (G^*G)^{\frac{1}{2}} e_k, e_k \right\rangle < \infty$ where $\{e_k\}_{k=1}^\infty$ is an orthonormal basis of H.

Equation (18) provides an explicit formula for calculating the likelihood given continuous observations of both $\bar{Z}^N$ and X. The approximate log-likelihood is:

$$\log \mathcal{L}^A_{M,N}(\theta) \propto \ell^N_\theta(\mathsf{Q}_{M,N}, \mathsf{X}_M). \tag{19}$$

The approximate estimator is:

$$\theta^A_{M,N} = (\alpha^A_{M,N}, \beta^A_{M,N}) \in \arg\max_{\theta\in\Theta} \ell^N_\theta(\mathsf{Q}_{M,N}, \mathsf{X}_M) = \arg\max_{\theta\in\Theta} \frac{1}{N}\ell^N_\theta(\mathsf{Q}_{M,N}, \mathsf{X}_M).$$

Lemma 7.4. *For each $\theta \in \Theta$, $\frac{1}{N}\ell^N_\theta(\mathsf{Q}_{M,N}, \mathsf{X}_M)$ converges in probability to:*

$$\ell^\infty_\beta \quad \equiv -\tfrac{1}{2}\int_0^{t_M}\int_0^{t_M} H_1(s,u;\beta)\tfrac{\partial}{\partial s}[g^\beta_s(Z^\infty_s - m^\beta_s)]\tfrac{\partial}{\partial u}[g^\beta_u(Z^\infty_u - m^\beta_u)]dsdu. \tag{20}$$

Theorem 7.5. *In addition to the assumptions of Lemma 7.3, suppose that:*

(C1) The Fredholm determinant $d(1;\beta)$ and the Fredholm resolvent $H_1(s,t;\beta)$ are almost surely continuous on Θ_β and $[0,t_M]\times[0,t_M]\times\Theta_\beta$, respectively.
(C2) The kernel $H_1(s,t;\beta)$ is almost surely positive definite.
(C3) There exists a $T_0 < \infty$ such that for almost every realization of X_M, the map $(Z^\infty_t)_{0\le t\le T} : \Theta_\beta \mapsto C([0,T])$ is one-to-one for $T > T_0$.

Then, the approximate estimator $\beta^A_{M,N}$ converges in probability to β_0 as $N \longrightarrow \infty$.

Theorem 7.6. *In addition to the assumptions of Theorem 7.5, assume that H_1 and g^β are almost surely twice continuously differentiable in β. Then, as $N \to \infty$, the estimator $\beta^A_{M,N}$ satisfies the following central limit theorem:*

$$\sqrt{N}(\beta^A_{M,N} - \beta_0) \xrightarrow{d} (\nabla^2 \ell^\infty_{\beta_0})^{-1} A_{\beta_0},$$

where $A_{\beta_0} = \int_0^{t_M}\int_0^{t_M} H_1(s,u;\beta_0) d[g^{\beta_0}_s \nabla m^{\beta_0}_s] d[g^{\beta_0}_u \langle f, \bar{\Xi}_u\rangle]$.

8. Numerical Studies

In this section, we numerically demonstrate the performance of the approximate likelihood approach for two models. The first is a model similar to Example 2.2 while the second is from Example 2.3.

8.1. Parameter Estimation for a model with mean-field interaction through the empirical mean.

Consider the mean-field system:

$$\begin{aligned} dY^{N,n}_t &= \beta^C \langle y, \mu^N_t\rangle dt + \sqrt{Y^{N,n}_t} dW^{N,n}_t + \beta^S Y^{N,n}_t dX_t, \\ dX_t &= dW_t. \end{aligned} \tag{21}$$

This model is similar to the interbank lending model of Fouque & Ichiba (2013) described in Example 2.2. A source of common risk, X_t, has been added which generates dependence between components beyond the mean-field term. Suppose we observe the empirical mean, $Z^N_t = \langle y, \mu^N_t\rangle$, and the systematic process X at discrete times $t_0, \ldots, t_M$. The parameters that need to be estimated are $\theta = (\beta^C, \beta^S)$. The true maximum likelihood estimator for (21) is not computationally tractable. The misspecified model is:

$$d\bar{Z}^N_t = \beta^C \bar{Z}^N_t dt + \beta^S \bar{Z}^N_t dX_t + \frac{1}{\sqrt{N}}\bar{Z}^\infty_t dV_t,$$

where V_t is a standard Brownian motion. The law of large numbers $\bar{Z}^\infty_t$ is a geometric Brownian motion. The assumptions of Theorem 5.3 hold for the model (21). In particular, the limiting process $\bar{Z}^\infty_t : \Theta \to \mathbb{R}^M$ is almost surely one-to-one for $M > M_0 = 1$. Therefore, the approximate estimator $\theta^A_{M,N}$ is consistent.

We numerically study the convergence of the estimator for $t_M = 10$ and observation intervals of length .025. The initial conditions are $Y^{N,n}_0 = 1$. Data is generated from the true model (21) with the true parameter $\theta_0 = (.25, .01)$. Then, the parameter vector θ is estimated using the approximate likelihood approach. Numerical results are reported in Table 2. The approximate estimators are highly accurate even for small N and converge to the true parameter as N grows large. The approximate likelihood approach has allowed for the accurate parameter estimation of a system for which maximum likelihood estimators cannot be tractably computed.

N	Approx. Likelihood Estimator for β^C	Approx. Likelihood Estimator for β^S
100	.2397	.008325
500	.2635	.008660
1,000	.2429	.01122
10,000	.2492	.01035
100,000	.2495	.009931

TABLE 2. Approximate estimators for the system (21) with true parameter $\theta_0 = (.25, .01)$. N is the size of the system.

8.2. **Parameter estimation for correlated default timing model.** We restate the model for correlated default timing from Example 2.3 for easy reference. We consider a pool of N names where each name defaults at a stochastic intensity $\lambda_t^{N,n}$. The indicator random variable $M_t^{N,n}$ is 1 if the n-th name is still alive at time t and 0 if it has already defaulted by time t. The loss L_t^N is the fraction of the names in the pool which have defaulted.

$$\begin{aligned} d\lambda_t^{N,n} &= \kappa(m - \lambda_t^{N,n})dt + \sigma\sqrt{\lambda_t^{N,n}}dW_t^n + \beta^C dJ_t^{N,n,1} + \beta^S\lambda_t^{N,n}dX_t, \\ dM_t^{N,n} &= -dJ_t^{N,n,2}, \\ dX_t &= dW_t, \end{aligned} \tag{22}$$

where $J_t^{N,n,2}$ is a point process with intensity $M_t^{N,n}\lambda_t^{N,n}$ and $J_t^{N,n,1} = L_t^N$. The observation process Z_t^N is the loss in the pool L_t^N. The limiting measure $\bar{\mu}$ for the system (22) satisfies the stochastic evolution equation:

$$d\langle f, \bar{\mu}_t\rangle = \left\{\langle \mathcal{L}_1 f, \bar{\mu}_t\rangle + \langle \mathcal{Q}, \bar{\mu}_t\rangle\langle \mathcal{L}_2 f, \bar{\mu}_t\rangle + \left\langle \mathcal{L}_3^{X_t} f, \bar{\mu}_t\right\rangle\right\}dt + \left\langle \mathcal{L}_4^{X_t} f, \bar{\mu}_t\right\rangle dW_t, \quad \text{a.s.,} \tag{23}$$

where we define the differential operators:

$$\begin{aligned} \mathcal{L}_1 f &= \frac{1}{2}\sigma^2\lambda\frac{\partial^2 f}{\partial\lambda^2} - \alpha(\lambda - \bar{\lambda})\frac{\partial f}{\partial\lambda} - \lambda f \\ \mathcal{L}_2 f &= \beta^C\frac{\partial f}{\partial\lambda} \\ \mathcal{L}_3^x f &= \frac{1}{2}(\beta^S)^2\lambda^2\frac{\partial^2 f}{\partial\lambda^2} \\ \mathcal{L}_4^x f &= \beta^S\lambda\frac{\partial f}{\partial\lambda}. \\ \mathcal{G}_{x,\mu} f &= \mathcal{L}_1 f + \mathcal{L}_3^x f + \langle \mathcal{Q}, \mu\rangle\mathcal{L}_2 f + \langle \mathcal{L}_2 f, \mu\rangle\mathcal{Q} \\ \mathcal{L}_5(f,g) &= \sigma^2\frac{\partial f}{\partial\lambda}\frac{\partial g}{\partial\lambda}\lambda \\ \mathcal{L}_6(f,g) &= fg\lambda \\ \mathcal{L}_7 f &= f\lambda. \end{aligned}$$

Also define

$$\mathcal{Q}(\lambda) \stackrel{\text{def}}{=} \lambda,$$

and the inner product

$$\langle f, \nu\rangle = \int_0^\infty f(\lambda)d\nu(\lambda).$$

The limiting measure $\bar{\Xi}$ for the system (22) satisfies the stochastic evolution equation:

$$d\left\langle f, \bar{\Xi}_t\right\rangle = \left\langle \mathcal{G}_{X_s,\bar{\mu}_s} f, \bar{\Xi}_s\right\rangle ds + \left\langle \mathcal{L}_4^{X_s} f, \bar{\Xi}_s\right\rangle dW_s + \left\langle f, \bar{\mathcal{M}}_t\right\rangle. \text{ a.s.} \tag{24}$$

N	Approx. Estimator for β^C	Approx. Estimator for β^S	Approx. Estimator for κ
1,000	.1920	.03277	.1236
10,000	.2975	.02121	.2058
100,000	.3102	.05165	.2068
500,000	.3061	.04855	.2046
2,000,000	.2986	.05021	.1989

TABLE 3. Approximate estimator for the system (22) with true parameters $\theta_0 = (.3, .05, .2)$. N is the size of the system.

Let $\mathcal{V}_t$ be the *sigma*-algebra generated by W_t. Then, conditional on $\mathcal{V}$, $\bar{\mathcal{M}}$ is, centered Gaussian with covariance function given by:

$$
\begin{aligned}
\operatorname{Cov}\left[\left\langle f, \bar{\mathcal{M}}_{t_1}\right\rangle, \left\langle g, \bar{\mathcal{M}}_{t_2}\right\rangle \,\middle|\, \mathcal{V}_{t_1\vee t_2}\right] = \mathbb{E}\Bigg[\int_0^{t_1\wedge t_2} & [\langle \mathcal{L}_5(f,g), \bar{\mu}_s\rangle + \langle \mathcal{L}_6(f,g), \bar{\mu}_s\rangle + \langle \mathcal{L}_2 f, \bar{\mu}_s\rangle \langle \mathcal{L}_2 g, \bar{\mu}_s\rangle \langle \mathcal{Q}, \bar{\mu}_s\rangle \\
& - \langle \mathcal{L}_7 g, \bar{\mu}_s\rangle \langle \mathcal{L}_2 f, \bar{\mu}_s\rangle - \langle \mathcal{L}_7 f, \bar{\mu}_s\rangle \langle \mathcal{L}_2 g, \bar{\mu}_s\rangle]\, ds \,\Bigg|\, \mathcal{V}_{t_1\vee t_2}\Bigg].
\end{aligned} \tag{25}
$$

Maximum likelihood estimation for the model (22) is computationally intractable. We implement the numerical methods described in Section 5.1 for the above model. Both the LLN and fluctuation stochastic evolution equations are solved using the method of moments. The stochastic evolution equations are reduced to systems of SDEs using the method of moments. The LLN system is solved using the Euler method. The distribution of the fluctuation SDE moment system is found using the fundamental solution of that system, as described in Subsection 5.1. The fundamental solution is calculated using the Euler method and X is simulated exactly. Data is simulated from the actual system in Example 2.3 using the timescaling approach described in Giesecke et al. (2013). We use 19 moments for the LLN and 6 moments for the fluctuation equation. We take X_t to be a standard Brownian motion W_t.

Unfortunately, the full model (22) is not identifiable. Multiple combinations of parameters can produce the same path for the limiting process $\mathsf{Z}_{M,\infty}$. Therefore, in order for the assumptions of Theorem 5.3 to be satisfied, we fix the parameters $(m, \sigma) = (0, .05)$ and estimate the parameters $\theta = (\beta^C, \beta^S, \kappa)$. The initial condtions are $Y_0^{N,n} = .05$. Data is generated from the true model (22) with the true parameters $\theta_0 = (.3, .05, .2)$. Observations of Z^N are discrete at time intervals of .025 and X is continuously observed. Observations are made over the time interval $[0, t_M]$ where $t_M = 25$. The parameter vector θ is then estimated using the approximate likelihood approach. Numerical results are reported in Table 3.

9. EXTENSIONS

Here we consider several extensions that further generalize our statistical approach for large interacting systems. We extend our approach to the heterogeneous parameter case. In addition, we show that the approximate estimation approach can also be applied to discrete-time systems.

9.1. **Heterogeneous parameters.** So far we have only considered the case where particles have homogeneous parameters in the system (1). This means that every particle has the same parameter θ. A more general model might consider several types of particles $1, \ldots, K$. Each type would have an associated parameter vector $\theta_k \in \Theta$. Consequently, the dynamics for particles of different types will be different. Let $k(n)$ be the type of the n-th particle. Then, the finite system is:

$$
\begin{aligned}
dX_t &= a(X_t; \alpha)dt + dW_t, \\
dY_t^{N,n} &= \nu(X_t, Y_t^{N,n}, \mu_t^N; \theta_{k(n)})dt + \sigma(X_t, Y_t^{N,n}, \mu_t^N; \theta_{k(n)})^\top dW_t^{N,n} + \gamma(X_t, Y_t^{N,n}, \mu_t^N; \theta_{k(n)})^\top dJ_t^{N,n} \\
&\quad + \zeta(X_t, Y_t^{N,n}, \mu_t^N; \theta_{k(n)})^\top dX_t,
\end{aligned}
$$

It is typically easy to extend the law of large numbers and fluctuation limit from the homogeneous parameter case to the heterogeneous parameter case for a finite set of different parameter types. With the heterogeneous LLN and fluctuation limit in hand, the same numerical approach developed for the homogeneous parameter case can be applied to the heterogeneous case in order to estimate $\{\theta_k\}_{k=1}^K$.

9.2. **Discrete Time.** The general framework considered in the main section of this paper is in continuous time. The approach can easily be applied to an interacting stochastic system with discrete-time dynamics. The discrete-time analog to the continuous-time framework (1) is the joint Markov process $\big(X_t, Y_t^{N,1}, \ldots, Y_t^{N,N}, \mu_t^N\big)$ for $t = 1, 2, 3, \ldots, T$.

If μ_t^N weakly converges to $\bar{\mu}_t$ and $\Xi_t^N = \sqrt{N}(\mu_t^N - \bar{\mu}_t)$ weakly converges to $\bar{\Xi}_t$, one can again use the large system approximation $\mu_t^N \approx \bar{\mu}_t^N = \bar{\mu}_t + \frac{1}{\sqrt{N}}\bar{\Xi}_t$ in order to perform parameter estimation. As in continuous time, $\bar{\mu}_t^N$ is conditionally Gaussian given X. Computations will in fact typically be considerably easier than in continuous time since the transition kernel for $\bar{\mu}_t^N$ will be known in closed-form. The transition kernel for $\bar{\mu}_t^N$ derives its form directly from the transition kernel for $Y_t^{N,n}$ in the finite system. An example of limiting laws for a discrete-time system can be found in Sirignano & Giesecke (2014).

9.3. **Least Squares Regression.** In some cases, the covariance of the central limit theorem $\bar{\Xi}$ may be computationally expensive to calculate. One can instead perform an unweighted least squares regression using only the law of large numbers. For instance, in the case of discretely observed Z^N and continuously observed X, the unweighted least squares minimization would be:

$$\beta_{M,N}^A = \arg\min_{\beta \in \Theta_\beta} \sum_{m=1}^{M} (Z_{t_m}^N - m_{t_m}^\beta)^2, \tag{26}$$

where here m_t^β is the process $\bar{Z}_t^\infty$ under the measure $\bar{\mathbb{D}}_\theta^{M,\infty}$. Recall that if the central limit theorem were included, the limiting approximate estimator would be a weighted least squares regression while here the limiting approximate estimator is an unweighted least squares regression. Although the approximate likelihood estimator will be more accurate, the least squares estimator (26) will have a lower computational expense. In addition, the least squares estimator (26) is consistent and asymptotically normal in the limit $N \to \infty$ provided that the conditions of Theorems 5.3 and 5.5 hold, respectively. An analogous formulation of (26) can obviously be constructed for continuous observations of Z^N.

Appendix A. Proofs

Proposition 4.1. Write $\mathcal{L}_{M,N}^A(\theta) = \prod_{m=1}^M \mathcal{I}_{m,N}^A(\theta)$ for the incremental approximate likelihood

$$\mathcal{I}_{m,N}^A(\theta) = \frac{\mathcal{L}_{m,N}^A(\theta)}{\mathcal{L}_{m-1,N}^A(\theta)}.$$

The incremental likelihoods $(\mathcal{I}_{m,N}^A(\theta))_{1 \le m \le M}$ are serially correlated. As a result, we find ourselves in the setting of Bar-Shalom (1971). We will show that equations (4.10)-(4.13) of Bar-Shalom (1971) hold in our setting. These equations imply that for any $\epsilon, \delta > 0$, the approximate likelihood $\mathcal{L}_{M,N}^A$ has a maximizer in a δ-neighborhood of θ_0 with probability larger than $1-\epsilon$ for sufficiently large M and N (see Bar-Shalom (1971, p. 76)). Therefore, equations (4.10)-(4.13) are sufficient conditions for consistency of a parameter estimator.

Because the third partial derivative of $C_{M,N}\mathcal{L}_{M,N}(\theta)$ is uniformly bounded for all $M, N \ge 1$ and $\theta \in \Theta$ by Assumption (A4), Assumption (A1) implies that

$$0 = C_{M,N}\nabla \log \mathcal{L}_{M,N}^A(\theta_{M,N}^A) = d_0 + d_1(\theta_{M,N}^A - \theta_0) + \frac{\zeta}{2}(\theta_{M,N}^A - \theta_0)^\top d_2(\theta_{M,N}^A - \theta_0)$$

for some $\zeta = \zeta(\theta_{M,N}^A, \theta_0)$ with $|\zeta| \le 1$, and

$$\begin{aligned}
d_0 &= C_{M,N}\nabla \log \mathcal{L}_{M,N}^A(\theta_0),\\
d_1 &= C_{M,N}\frac{\nabla^2 \mathcal{L}_{M,N}^A(\theta_0)}{\mathcal{L}_{M,N}^A(\theta_0)} - C_{M,N}\nabla \log \mathcal{L}_{M,N}^A(\theta_0)^\top \nabla \log \mathcal{L}_{M,N}^A(\theta_0),\\
d_2 &= C_{M,N}H.
\end{aligned}$$

Assumption (A2) immediately yields equation (4.10) of Bar-Shalom (1971), while Assumption (A3) directly implies (4.12) of Bar-Shalom (1971). Further, Equation (4.13) of Bar-Shalom (1971) holds because $C_{M,N}H$ is bounded in probability by Assumption (A4). For equation (4.11) of Bar-Shalom (1971), note that Assumption (A2) implies that the first term on the right-hand side of d_1 converges to zero in $\mathbb{P}_{\theta_0}$-probability, and

Assumption (A3) implies that the second term of the right-hand side of d_1 converges to $-\Sigma_{\theta_0}$ in $\mathbb{P}_{\theta_0}$-probability. Thus, consistency follows.

For asymptotic efficiency, note Assumption (A3) implies that

$$C_{M,N}^{-1/2}\left(\theta_{M,N}^A - \theta_0\right) = -\left(\Sigma_{\theta_0}^{-1} d_1 + \Sigma_{\theta_0}^{-1}\frac{\zeta}{2}(\theta_{M,N}^A - \theta_0)^\top d_2\right)^{-1} \Sigma_{\theta_0}^{-1} C_{M,N}^{-1/2} d_0.$$

The terms in parenthesis above converge to one in $\mathbb{P}_{\theta_0}$-probability as $M, N \to \infty$ given the consistency of $\theta_{M,N}^A$ and the fact that $d_1 \to -\Sigma_{\theta_0}$. As a result, the variance of $C_{M,N}^{-1/2}\left(\theta_{M,N}^A - \theta_0\right)$ is dominated by the variance of $\Sigma_{\theta_0}^{-1} C_{M,N}^{-1/2} d_0$. We have that

$$\mathbb{E}_{\theta_0}\left[C_{M,N}^{-1} d_0^\top d_0\right] = \mathbb{E}_{\theta_0}\left[C_{M,N} \nabla \log \mathcal{L}_{M,N}^A(\theta_0)^\top \nabla \log \mathcal{L}_{M,N}^A(\theta_0)\right].$$

Assumption (A3) tells us that the matrix $C_{M,N}\nabla \log \mathcal{L}_{M,N}^A(\theta_0)^\top \nabla \log \mathcal{L}_{M,N}^A(\theta_0)$ converges in $\mathbb{P}_{\theta_0}$-probability to the deterministic limit Σ_{θ_0}. Therefore, this matrix is bounded in probability for large values of M and N. The bounded convergence theorem then implies that

$$\lim_{M,N\to\infty} \mathbb{E}_{\theta_0}\left[C_{M,N}^{-1} d_0^\top d_0\right] = \Sigma_{\theta_0}.$$

Further, Assumption (A3) states that Σ_{θ_0} is positive definite. Thus, a square root matrix exists for Σ_{θ_0}. We conclude that

$$\mathrm{Var}\left(C_{M,N}^{-1/2}\left(\theta_{M,N}^A - \theta_0\right)\right) \to \Sigma_{\theta_0}^{-1}$$

as $M, N \to \infty$. Thus, the approximate likelihood estimator $\hat{\theta}_{M,N}^A$ is asymptotically efficient.

□

Proof of Theorem 5.1. Proof is omitted since it directly follows from well-known properties of Gaussian distributions. □

Proof of Proposition 5.2. First, fix $z \in \mathbb{R}^M$ and $x \in C([0, t_M])$ and consider $\frac{1}{N}\ell_\theta^N(z, x)$. We will show that $\frac{1}{N}\ell_\theta^N$ converges uniformly over Θ to $-\frac{1}{2}(z - m_\beta^M)^\top(\Sigma_\beta^M)^{-1}(z - m_\beta^M)$ for each $z \in \mathbb{R}^M$. For each N, we have that

$$\begin{aligned}\frac{1}{N}\ell_\theta^N(z,x) &= \tfrac{1}{N}\left[\int_0^{t_M} a(x_u;\alpha)\mathrm{d}x_u - \tfrac{1}{2}\int_0^{t_M} a^2(x_u;\alpha)du\right] + \tfrac{1}{N}[\log((2\pi)^{\frac{-M}{2}}) - M\log(N^{\frac{1}{2}}) - \tfrac{1}{2}\log(det(\Sigma_\beta^M))] \\ &\quad -\tfrac{1}{2}(z - m_\beta^M)^\top(\Sigma_\beta^M)^{-1}(z - m_\beta^M).\end{aligned}$$

Since m_β^M and Σ_β^M are continuous in θ and Θ is compact, they are both bounded. In addition, since Σ_β^M is positive definite and achieves its lower bound due to compactness, $det(\Sigma_\beta^M) > e_0$ where $e_0 > 0$ for $\beta \in \Theta_\beta$. It follows that $\frac{1}{N}\ell_\theta^N$ converges uniformly over Θ_β to $-\frac{1}{2}(z-m_\beta^M)^\top(\Sigma_\beta^M)^{-1}(z-m_\beta^M)$ for each $z \in \mathbb{R}^M$. The result then follows due to the continuity of $\frac{1}{N}\ell_\theta^N$ for each N, Θ_β being compact, and the uniform convergence. □

Next, consider the limits as $N \to \infty$ of the estimators $\alpha_{M,N}^A$ and $\beta_{M,N}^A$. Note that $\alpha_{M,N}^A$ is the same for every N because the partial log-likelihood ℓ_α does not depends on N. As a result, the limit of $\alpha_{M,N}^A$ is trivial. For the limit of $\beta_{M,N}^A$, consider the following facts. $\mathsf{Z}_{M,N}$ weakly converges to $\mathsf{Z}_{M,\infty}$. Since $\mathsf{Z}_{M,\infty}$ is deterministic given X_M, by Slutsky's theorem, we have that, conditional on X_M, $\mathsf{Z}_{M,N}$ converges in probability to $\mathsf{Z}_{M,\infty}$. Furthermore, the mean m_β^M and the matrix Σ_β^M depend only on X_M and not on $\mathsf{Z}_{M,N}$. The first part of this proof then implies that, for each realization of X_M in a set of probability 1,

$$\frac{1}{N}\ell_\theta^N(\mathsf{Z}_{M,N}, \mathsf{X}_M) \xrightarrow{p} -\frac{1}{2}(\mathsf{Z}_{M,\infty} - m_\beta^M)^\top(\Sigma_\beta^M)^{-1}(\mathsf{Z}_{M,\infty} - m_\beta^M).$$

The following lemmas will be used in the proof of Theorem 5.3.

Lemma A.1. *For each $\theta \in \Theta$ and almost every realization of X_M, $\frac{1}{N}\ell_\beta^N(\mathsf{Z}_{M,N}, \mathsf{X}_M)$ converges in probability to $-\frac{1}{2}(\mathsf{Z}_{M,\infty} - m_\beta^M)^\top(\Sigma_\beta^M)^{-1}(\mathsf{Z}_{M,\infty} - m_\beta^M)$.*

Proof. $\mathsf{Z}_{M,N}$ weakly converges to $\mathsf{Z}_{M,\infty}$. Since, conditional on X_M, $\mathsf{Z}_{M,\infty}$ is deterministic, by Slutsky's theorem, we have that, conditional on each realization of X_M in a set of probability 1, $\mathsf{Z}_{M,N}$ converges in probability to $\mathsf{Z}_{M,\infty}$. The result then follows in the same way as in Proposition 5.2. □

Lemma A.2. *For almost every X_M, the sequence $\{\frac{1}{N}\ell^N_\beta(\mathsf{Z}_{M,N},\mathsf{X}_M)\}_N$ is equicontinuous in probability.*

Proof. For each realization of X_M in a set of probability 1, we prove the stochastic equicontinuity of $\{\frac{1}{N}\ell^N_\beta(\mathsf{Z}_{M,N},\mathsf{X}_M)\}_N$. Stochastic equicontinuity means that for every $\epsilon>0$ and $\eta>0$, there exists a $\delta>0$ such that

$$\lim_{N\to\infty}\mathbb{D}^{M,N}_{\theta_0}\big[\sup_{|\beta_1-\beta_2|\le\delta,\ \beta_1,\beta_2\in\Theta_\beta}|\frac{1}{N}\ell^N_{\beta_1}(\mathsf{D}_{M,N})-\frac{1}{N}\ell^N_{\beta_2}(\mathsf{D}_{M,N})|>\epsilon\big|\mathsf{X}_M\big]<\eta.$$

By the mean value theorem and Markov's inequality, we have that

$$\begin{aligned}
&\lim_{N\to\infty}\mathbb{D}^{M,N}_{\theta_0}\big[\sup_{|\beta_1-\beta_2|\le\delta}|\frac{1}{N}\ell^N_{\beta_1}(\mathsf{D}_{M,N})-\frac{1}{N}\ell^N_{\beta_2}(\mathsf{D}_{M,N})|>\epsilon|\mathsf{X}_M\big]\\
=\ &\lim_{N\to\infty}\mathbb{D}^{M,N}_{\theta_0}\big[\sup_{|\beta_1-\beta_2|\le\delta}|(\mathsf{Z}_{M,N}-m^M_{\beta_1})^\top(\Sigma^M_{\beta_1})^{-1}(\mathsf{Z}_{M,N}-m^M_{\beta_1})+\frac{1}{2N}\log(det(\Sigma^M_{\beta_1}))\\
-\ &(\mathsf{Z}_{M,N}-m^M_{\beta_2})^\top(\Sigma^M_{\beta_2})^{-1}(\mathsf{Z}_{M,N}-m^M_{\beta_2})-\frac{1}{2N}\log(det(\Sigma^M_{\beta_2}))|>\epsilon|\mathsf{X}_M\big]\\
\le\ &\lim_{N\to\infty}\mathbb{D}^{M,N}_{\theta_0}\big[\sup_{|\beta_1-\beta_2|\le\delta}|(\mathsf{Z}_{M,N}-m^M_{\beta_1})^\top(\Sigma^M_{\beta_1})^{-1}(\mathsf{Z}_{M,N}-m^M_{\beta_1})-(\mathsf{Z}_{M,N}-m^M_{\beta_2})^\top(\Sigma^M_{\beta_2})^{-1}(\mathsf{Z}_{M,N}-m^M_{\beta_2})|>\frac{\epsilon}{2}|\mathsf{X}_M\big]\\
+\ &\lim_{N\to\infty}\frac{1}{2N}\sup_{|\beta_1-\beta_2|\le\delta}|\log(det(\Sigma^M_{\beta_1}))-\log(det(\Sigma^M_{\beta_2}))|\big]\\
=\ &\lim_{N\to\infty}\mathbb{D}^{M,N}_{\theta_0}\big[\sup_{|\beta_1-\beta_2|\le\delta}|(A_{\beta_1}-A_{\beta_2})+(B_{\beta_1}-B_{\beta_2})\mathsf{Z}_{M,N}+(C_{\beta_1}-C_{\beta_2})\mathsf{Z}^\top_{M,N}\mathsf{Z}_{M,N}|>\epsilon|\mathsf{X}_M\big]\\
\le\ &\lim_{N\to\infty}\mathbb{D}^{M,N}_{\theta_0}\big[\delta\big(\sup_{\beta\in\Theta_\beta}|\nabla A_\beta|+\sup_{\beta\in\Theta_\beta}|\nabla B_\beta\mathsf{Z}_{M,N}|+\sup_{\beta\in\Theta_\beta}|\nabla C_\beta\mathsf{Z}^\top_{M,N}\mathsf{Z}_{M,N}|\big)>\epsilon|\mathsf{X}_M\big]\\
\le\ &\lim_{N\to\infty}\big(\mathbb{D}^{M,N}_{\theta_0}\big[\sup_{\beta\in\Theta_\beta}|\nabla A_\beta|>\frac{\epsilon}{3\delta}|\mathsf{X}_M\big]+\mathbb{D}^{M,N}_{\theta_0}\big[\sup_{\beta\in\Theta_\beta}|\nabla B_\beta\mathsf{Z}_{M,N}|>\frac{\epsilon}{3\delta}|\mathsf{X}_M\big]+\mathbb{D}^{M,N}_{\theta_0}\big[\sup_{\beta\in\Theta_\beta}|\nabla C_\beta\mathsf{Z}^2_{M,N}|>\frac{\epsilon}{3\delta}|\mathsf{X}_M\big]\big).
\end{aligned}\tag{27}$$

First, $|\log(det(\Sigma^M_{\beta_1}))-\log(det(\Sigma^M_{\beta_2}))|$ is clearly bounded above and below due to the continuity of Σ^M_β, the compactness of Θ, and the positive definiteness of Σ^M_β. The terms A_β, B_β, and C_β are constants. Again due to the Assumptions of Proposition 5.2, $\sup_{\beta\in\Theta_\beta}|\nabla A_\beta|$, $\sup_{\beta\in\Theta_\beta}|(B_\beta)_k|$, and $\sup_{\beta\in\Theta_\beta}|(C_\beta)_{i,j}|$ are all bounded for all i,j,k. Since weak convergence implies tightness, for any $\kappa>0$, one can find a compact set K_κ such that $\mathsf{Z}_{M,N}$ lies in that set with probability $1-\kappa$. Therefore, for any ϵ and η, one can find a δ such that the last quantity in (27) is bounded by η. □

Proof of Theorem 5.3. It is sufficient to prove the convergence in probability of the estimator to θ_0 for each realization of X_M in a set of probability 1, since θ_0 is a constant. For each realization of X_M, the convergence of the approximate estimator to the limiting set $\beta^A_{M,\infty}(\mathsf{Z}_{M,\infty},\mathsf{X}_M)$ follows from the pointwise convergence and stochastic equicontinuity of the approximate likelihood function. Due to the positive definiteness of Σ_β, any β which maximizes the limiting likelihood in Lemma A.1 must satisfy $m^M_\beta=\mathsf{Z}_{M,\infty}$. By definition, β_0 is in this set of maximizers. If the only $\beta\in\Theta_\beta$ that could have produced the limiting path $\mathsf{Z}_{M,\infty}$ is β_0, then the limiting estimator is unique and equals β_0. Therefore, if the limiting process $\mathsf{Z}_{M,\infty}:\Theta_\beta\mapsto\mathbb{R}^{M+1}$ is almost surely one-to-one and $M>M_0$, the approximate estimator converges in probaiblity to β_0 in the limit $N\longrightarrow\infty$. □

Proof of Theorem 5.5. By the mean value theorem, we have that

$$0=\nabla\ell^N_{\beta^A_{M,N}}=\nabla\ell^N_{\beta_0}+\nabla^2\ell^N_{\beta^1_{M,N}}(\beta^A_{M,N}-\beta_0),$$

where $\beta^1_{T,N}\in[\beta^A_{M,N},\beta_0]$. One finds by rearranging that:

$$\sqrt{N}(\beta^A_{M,N}-\beta_0)=-(\nabla^2\ell^N_{\beta^1_{M,N}})^{-1}\sqrt{N}\nabla\ell^N_{\beta_0}.$$

Since the matrix inverse is a continuous function, $\nabla^2 \ell^N_{\beta^1_{M,N}}$ is continuous in both $\beta^1_{M,N}$ and $\mathsf{Z}_{M,N}$, and both $\beta^1_{M,N}$ and $\mathsf{Z}_{M,N}$ converge in probability, we have by Slutsky's theorem and the continuous mapping theorem that $(\nabla^2 \ell^N_{\beta^1_{M,N}})^{-1}$ converges in probability to $(\nabla^2 \ell^\infty_{\beta_0})^{-1}$.

Next, let's examine the term $\sqrt{N}\nabla \ell^N_{\beta_0}$.

$$
\begin{aligned}
\sqrt{N}\nabla \ell^N_{\beta_0} &= \frac{1}{\sqrt{N}}[\log((2\pi)^{\frac{-M}{2}}) - M\log(N^{\frac{1}{2}}) - \frac{1}{2}\log(det(\Sigma^M_\beta))] - \frac{\sqrt{N}}{2}\nabla(\mathsf{Z}_{M,N} - m^M_{\beta_0})^\top (\Sigma^M_{\beta_0})^{-1}(\mathsf{Z}_{M,N} - m^M_{\beta_0}) \\
&= \frac{1}{\sqrt{N}}[\log((2\pi)^{\frac{-M}{2}}) - M\log(N^{\frac{1}{2}}) - \frac{1}{2}\log(det(\Sigma^M_\beta))] \\
&- \frac{\sqrt{N}}{2}\nabla \sum_{i,j} (\mathsf{Z}_{M,N} - m^M_{\beta_0})_i ((\Sigma^M_{\beta_0})^{-1})_{i,j} (\mathsf{Z}_{M,N} - m^M_{\beta_0})_j.
\end{aligned}
$$

By the chain rule and the weak convergence results, one can see that:

$$\sqrt{N}\nabla \ell^N_{\beta_0} \xrightarrow{d} Q_{\beta_0} \equiv \sum_{i,j} \nabla (m^M_{\beta_0})_i ((\Sigma^M_{\beta_0})^{-1})_{i,j} \left\langle f, \bar{\Xi}_{t_j} \right\rangle.$$

Putting the pieces together, one has that

$$\sqrt{N}(\beta^A_{M,N} - \beta_0) \xrightarrow{d} (\nabla^2 \ell^\infty_{\beta_0})^{-1} Q_{\beta_0}.$$

□

Proof of Proposition 6.1. Rewrite it as:

$$\sum_{l=1}^L \ell^N_{\theta'}(z, \mathsf{X}^{c,l}_M) w^l_\theta = \frac{\frac{1}{L}\sum_{l=1}^L \ell^N_{\theta'}(z, \mathsf{X}^{c,l}_M) \bar{p}^{N,c}_\theta(z|\mathsf{X}^{c,l}_M)}{\frac{1}{L}\sum_{k=1}^L \bar{p}^{N,c}_\theta(z|\mathsf{X}^{c,k}_M)}.$$

Note that the denominator $\frac{1}{L}\sum_{k=1}^L \bar{p}^{N,c}_\theta(z|\mathsf{X}^{c,k}_M)$ is positive and converges almost surely to $\mathbb{E}_\theta[\bar{p}^{N,c}_\theta(z|\mathsf{X}^c_M)|\mathsf{X}_M]$ as $L \longrightarrow \infty$ by the strong law of large numbers. Similarly, the numerator $\frac{1}{L}\sum_{l=1}^L \ell^N_{\theta'}(z, \mathsf{X}^{c,l}_M)\bar{p}^{N,c}_\theta(z|\mathsf{X}^{c,l}_M)$ converges almost surely to $\mathbb{E}_\theta[\ell^N_{\theta'}(z, \mathsf{X}^c_M)\bar{p}^{N,c}_\theta(z|\mathsf{X}^c_M)|\mathsf{X}_M]$ as $L \longrightarrow \infty$. One then has that

$$
\begin{aligned}
\textstyle\sum_{l=1}^L \ell^N_{\theta'}(z, \mathsf{X}^{c,l}_M) w^l_\theta &\xrightarrow{a.s.} \mathbb{E}_\theta[\ell^N_{\theta'}(z, \mathsf{X}^c_M) \tfrac{\bar{p}^{N,c}_\theta(z|\mathsf{X}^c_M)}{\mathbb{E}_\theta[\bar{p}^{N,c}_\theta(z|\mathsf{X}^c_M)|\mathsf{X}_M]} |\mathsf{X}_M] \\
&= \mathbb{E}_\theta[\ell^N_{\theta'}(z, \mathsf{X}^c_M)|\mathsf{X}_M, \bar{\mathsf{Z}}_{M,N} = z] = Q(\theta', z|\theta)
\end{aligned}
$$

By Bayes' rule, one can see that $\frac{\bar{p}^{N,c}_\theta(z|\mathsf{X}^c_M)}{\mathbb{E}_\theta[\bar{p}^{N,c}_\theta(z|\mathsf{X}^c_M)|\mathsf{X}_M]}$ is the Radon-Nikodym derivative of the conditional $\mathbb{P}_\theta$-law of X^c_M given $(\bar{\mathsf{Z}}_{M,N}, \mathsf{X}_M)$ with respect to the conditional $\mathbb{P}_\theta$-law of X^c_M given X_M. □

Proof of Lemma 7.1. Let $g^\beta_t = H(f, f, \bar{\mu}_t)^{-1}$. The result is shown by a simple calculation of the quadratic variation $[\bar{Q}^N_t, \bar{Q}^N_t]$ from the fluctuation limit (5). □

Proof of Lemma 7.2. The Gaussian process $\bar{Q}^N_t$ has the same quadratic variation as the Wiener process. Therefore, $\bar{Q}^N$ and $\mathbb{W}$ are equivalent by the Feldman-Hajek theorem. Hence, by the results in Shepp (1966), the Radon-Nikodym derivative $\frac{d\bar{\mathbb{Q}}^{M,N}_\theta}{d\mathbb{W}}\Big|_{\bar{\mathsf{Q}}_{M,N}}$ exists. □

Proof of Proposition 7.3. From Lemma 7.2 and the results in Shepp (1966). □

Proof of Lemma 7.4. Directly follows from weak convergence of Z^N_t to Z^∞_t. Since Z^∞_t is deterministic given the complete path X_M, this implies convergence in probability. □

The following lemma is used in the proof of Theorem 7.5.

Lemma A.3. *Let the first two assumptions in Theorem 7.5 hold. Then, $\{\frac{1}{N}\ell^N_\theta(\mathsf{Q}_{M,N}, \mathsf{X}_M)\}_N$ is equicontinuous in probability.*

Proof. The proof is similar to the finite-dimensional case; the result follows from the continuity of H_1, the compactness of Θ, and the tightness of the sequence Z^N_t. □

Proof of Theorem 7.5. Due to assumptions (C2) and (C3) and $m_t^{\beta_0} = Z_t^{\infty}$, β_0 is the unique maximizer of the limiting likelihood ℓ^{∞}. $\beta_{M,N}^A \xrightarrow{p} \beta_0$ then follows from the pointwise convergence of the likelihood and the stochastic equicontinuity. $\square$

Proof of Theorem 7.6. The theorem can be proven using the same procedure as in Theorem 5.5. $\square$

References

Achdou, Y. & I. Capuzzo-Dolcetta (2010), 'Mean field games: Numerical methods', *SIAM Journal on Numerical Analysis* **48**(3), 1136–1162.

Ait-Sahalia, Yacine (2002), 'Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach', *Econometrica* **70**(1), 223–262.

Bar-Shalom, Yaakov (1971), 'On the asymptotic properties of the maximum-likelihood estimate obtained from dependent observations', *Journal of the Royal Statistical Society. Series B (Methodological)* **33**(1), 72–77.

Beskos, A., O. Papaspiliopoulos & G. Roberts (2006), 'Retrospective exact simulation of diffusion sample paths with applications', *Bernoulli* pp. 1077–1098.

Bishwal, J. (2011), 'Estimation in interacting diffusions: Continuous and discrete sampling', *Applied Mathematics* **2**, 1154–1158.

Bo, L. & A. Capponi (2014*a*), 'Bilateral credit valuation adjustment for large credit derivatives portfolios', *Finance and Stochastics* **18**(2), 431–482.

Bo, L. & A. Capponi (2014*b*), Systemic risk in interbanking networks.

Bossy, M. & D. Talay (1997), 'A stochastic particle method for the mckean-vlasov and the burgers equation', *Mathematics of Computation* **66**(217), 157–192.

Bossy, M., L. Fezoui & S. Piperno (2009), 'Comparison of a stochastic particle method and a finite volue deterministic method applied to burgers equation', *Monte Carlo Methods and Applications* **3**(2), 113–140.

Bush, N., B. M. Hambly, H. Haworth, L. Jin & C. Reisinger (2011), 'Stochastic evolution equations in portfolio credit modelling', *SIAM J. Fin. Math.* **2**(1), 627–664.

Carmona, R., J. Fouque & L. Sun (2013), Mean field games and systemic risk.

Carmona, Rene & Francois Delarue (2013), 'Probabilistic analysis of mean-field games', *SIAM Journal on Control and Optimization* **51**(4), 2705–2734.

Carmona, René, François Delarue & Aimé Lachapelle (2012), 'Control of mckean-vlasov dynamics versus mean field games', *Mathematics and Financial Economics* pp. 1–36.

Chan, P. & R. Sircar (2014), Bertrand and cournot mean field games.

Collet, F., P. Dai Pra & E Sartori (2010), 'A simple mean field model for social interactions: dynamics, fluctuations criticality', *Journal of Statistical Physics* **139**(5), 820–858.

Crimaldi, I., P. Dai Pra & I. Minelli (2014), Fluctuation theorems for synchronization of interacting polya's urns.

Crisan, D., T. Kurtz & Y. Lee (2012), Conditional distributions, exchangeable particle systems, and stochastic partial differential equations.

Cvitanic, Jaksa, Jin Ma & Jianfeng Zhang (2012), 'The law of large numbers for self-exciting correlated defaults', *Stochastic Processes and their Applications* **122**.

Dai Pra, P., W.J. Runggaldier, E. Sartori & M. Tolotti (2009), 'Large portfolio losses: A dynamic contagion model', *The Annals of Applied Probability* **19**(1), 347–394.

Dai Pra, Paolo & Frank den Hollander (1996), 'Mckean-vlasov lmit for interacting random processes in random media', *Journal of Statistical Physics* **84**(3), 735–772.

Dawson, D. (1983), 'Critical dynamics and fluctuations for a mean-field model of cooperative behavior', *Journal of Statistical Physics* **31**(1), 29–85.

Delarue, F., J. Inglis, S. Rubenthaler & E. Tanre (2014), Particle systems with a singular mean-field self-excitation. application to neuronal networks.

Dembo, A. & O. Zeitouni (1986), 'Parameter estimation of partially observed continuous time stochastic processes via the em algorithm', *Stochastic Processes and their Applications* **23**(1), 91–113.

Donev, A., E. Vanden-Eijnden, A. Garcia & J. Bell (2010), 'On the accuracy of finite-volume schmes for fluctuating hydrodynamics', *Communications in Applied Mathematics and Computational Science* **5**(2), 149–197.

Duffie, Darrell, Jun Pan & Kenneth Singleton (2000), 'Transform analysis and asset pricing for affine jump-diffusions', *Econometrica* **68**, 1343–1376.

Feigin, P.D. (1976), 'Maximum likelihood estimation for continuous-time stochastic processes', *Advances in Applied Probability* **8**(4), 712–736.

Fernandez, B. & S. Meleard (2000), 'Asymptotic behaviour for interacting diffusion processes with space-time random birth', *Bernoulli* **6**(1), 91–111.

Fontbona, J. (2006), 'A probabilistic interpretation and stochastic particle approximations of the three-dimensional navier-stokes equations', *Probability theory and related fields* **136**(1), 102–156.

Fouque, J-P. & T. Ichiba (2013), 'Stability in a model of inter-bank lending', *SIAM Journal on Financial Mathematics* **4**, 784–803.

Garnier, J., G. Papanicolaou & T. Yang (2013), 'Large deviations for a mean field model of systemic risk', *SIAM J. Fin. Math.* **4**(1), 151–184.

Giesecke, K., K. Spiliopoulos & R.B. Sowers (2012), 'Default clustering in large portfolios: Typical events', *Annals of Applied Probability* **23**(1), 348–385.

Giesecke, K., K. Spiliopoulos, R.B. Sowers & J.A. Sirignano (2013), 'Large portfolio asymptotics for loss from default', *Mathematical Finance (in press)* .

Gourieroux, C., A. Monfort & A. Trognon (1984), 'Pseudo maximum likelihood methods: Theory', *Econometrica* **52**(3), 681–700.

Guéant, Olivier (2009), 'A reference case for mean field games models', *Journal de Mathématiques Pures et Appliquées* **92**(3), 276 – 294.

Guéant, Olivier (2012), 'Mean field games equations with quadratic hamiltonian: a specific approach', *Mathematical Models and Methods in Applied Sciences* **22**(9).

Hall, P. & C.C. Heyde (1980), *Martingale limit theory and its application*, Probability and mathematical statistics, Academic Press.

Haynatzka, V., J. Gani & S. Rachev (2000), 'The spread of aids among interactive transmission groups', *Mathematical and Computer Modelling* **32**, 169–180.

Iacus, Stefano M. (2008), *Simulation and inference for stochastic differential equations: with R examples*, Springer series in statistics, Springer.

Jeganathan, P. (1995), 'Some aspects of asymptotic theory with applications to time series models', *Econometric Theory* **11**(5), 818–887.

Kasonga, R. (1990), 'Maximum likelihood theory for large interacting systems', *SIAM Journal on Applied Mathematics* **50**(3), 865–875.

Komatsu, T. & A. Takeuchi (2001), 'On the smoothness of pdf of solutions to sde of jump type', *International Journal of Differential Equations and Applications* **2**, 141–197.

Kurtz, T. & J. Xiong (2004), 'A stochastic evolution equation arising from the fluctuations of a class of interacting particle systems', *Communications in Mathematical Sciences* **2**(3), 325–358.

Lasry, J-M. & P-L Lions (2007), 'Mean field games', *Japanese Journal of Mathematics* **2**(1), 229–260.

McKeague, Ian W. (1984), 'Estimation for diffusion processes under misspecified models', *Journal of Applied Probability* **21**(3), 511–520.

Meleard, Sylvie (1996), 'Asymptotic behaviour of some interacting particle systems; mckean-vlasov and boltzmann models', *Probabilistic models for nonlinear partial differential equations* **42**(95), 42–95.

Millar, P. W. (1984), 'A general approach to the optimality of minimum distance estimators', *Transactions of the American Mathematical Society* **286**(1), 377–418.

Morale, D., V. Capasso & K. Oelschlager (2005), 'An interacting particle system modelling aggregation behavior: from individuals to populations', *Journal of Mathematical Biology* **50**(1), 49–66.

Nagasawa, M. & H. Tanaka (1987), 'Diffusion with interactions and collisions between coloured particles and the propagation of chaos', *Probability Theory and Related Fields* **74**(2), 161–198.

Nielsen, S. F. (2000), 'The stochastic EM Algorithm: Estimation and asymptotic results', *Bernoulli* **6**(3), 457–489.

Ogata, Yosihiko (1980), 'Maximum likelihood estimates of incorrect markov models for time series and the derivation of aic', *Journal of Applied Probability* **17**(1), 59–72.

Protter, P. (2004), *Stochastic Integration and Differential Equations*, Springer-Verlag, New York.

Shepp, L.A. (1966), 'Radon-nikodym derivatives of gaussian measures', *The Annals of Mathematical Statistics* **37**(2), 321–354.

Sirignano, J. & K. Giesecke (2014), Efficient risk analysis for mortgage pools.

Spiliopoulos, K., J.A. Sirignano & K. Giesecke (2014), 'Fluctuation analysis for the loss from default', *Stochastic Processes and their Applications* **124**(7), 2322–2362.

Sznitman, A. (1984), 'Nonlinear reflecting diffusion process, and the propagation of chaos and fluctuations associated', *Journal of functional analysis* **56**(3), 311–336.

Takeuchi, Atsushi (2002), 'The malliavin calculus for sde with jumps and the partially hypoelliptic problem', *Osaka J. Math.* **39**(3), 523–559.

White, Halbert (1982), 'Maximum likelihood estimation of misspecified models', *Econometrica* **50**(1), 1–25.

Department of Management Science and Engineering, Stanford University, Stanford, CA 94305
E-mail address: jasirign@stanford.edu

Boston University School of Management, Boston University, Boston, MA 02115
E-mail address: gas@bu.edu

Department of Management Science and Engineering, Stanford University, Stanford, CA 94305
E-mail address: giesecke@stanford.edu