Simulated Nonparametric Estimation of Dynamic Models with Applications to Finance*

Filippo Altissimo
European Central Bank, and CEPR

Antonio Mele
London School of Economics

First draft: April 11, 2001
This version: May 19, 2005

Abstract

This paper introduces a new class of parameter estimators for dynamic models, called Simulated Nonparametric Estimators (SNE). The SNE minimizes appropriate distances between nonparametric joint (or conditional) densities estimated from sample data and nonparametric joint (or conditional) densities estimated from data simulated out of the model of interest. Sample data and model-simulated data are smoothed with the same kernel. This makes the SNE: 1) consistent independently of the amount of smoothing (up to identifiability); and 2) asymptotically root-T normal when the smoothing parameter goes to zero at a reasonably mild rate. Furthermore, the estimator displays the same asymptotic efficiency properties as the maximum-likelihood estimator as soon as the model is Markov in the observable variables. The methods are flexible, simple to implement, and fairly fast; furthermore, they possess finite sample properties that are well approximated by the asymptotic theory. These features are illustrated within the typical estimation problems arising in financial economics.

JEL: C14, C15, C32, G12

Keywords: nonparametric estimation, asset pricing, continuum of moments, simulations

Corresponding author: Antonio Mele, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. Email: a.mele@lse.ac.uk.

*We wish to thank the editor, Bernard Salanié, and two anonymous referees for very helpful comments and suggestions; and Yacine Aït-Sahalia, Torben Andersen, Marine Carrasco, Mikhail Chernov, Frank Diebold, Cristian Huse, Dennis Kristensen, Oliver Linton, Nour Meddahi, Angelo Melino, Alex Michaelides, Eric Renault, Christopher Sims, Sam Thompson, seminar participants at CORE, the LSE, the Norwegian School of Economics and Business, Princeton University, the University of Pennsylvania and, especially, Valentina Corradi for her valuable comments and Fabio Fornari who worked with us in the initial phase of this project. We remain, however, solely responsible for any omission and mistakes, as well as for the views expressed in the paper, which do not reflect those of the European Central Bank.

1 Introduction

This paper introduces a new class of parameter estimators for dynamic models with possibly unobserved components, called Simulated Nonparametric Estimators (hereafter SNE). The SNE works by making the finite dimensional distributions of the model's observables as close as possible to their empirical counterparts estimated through standard nonparametric techniques. Since the distribution of the model's observables is in general analytically intractable, we recover it through two steps. In the first step, we simulate the system of interest. In the second step, we obtain model's density estimates through the application of the same nonparametric devices used to smooth the sample data. The result is a consistent and root-T asymptotically normal estimator displaying a number of attractive properties. First, our estimator is based on simulations; consequently, it can be implemented in a straightforward manner to cope with a variety of estimation problems. Second, the SNE is purposely designed to minimize distances of densities smoothed with the same kernel; therefore, up to identifiability, it is consistent regardless of the smoothing parameter behavior. Third, if the SNE is taken to match conditional densities and the model is Markov in the observables, it achieves the same asymptotic efficiency as the maximum-likelihood estimator (MLE). Finally, Monte Carlo experiments reveal that our estimator does exhibit a proper finite sample behavior.

Systems with unobserved components arise naturally in many areas of economics. Examples in macroeconomics include models of stochastic growth with human capital and/or sunspots, job duration models, or models of investment-specific technological changes. Examples arising in finance include latent factor models, processes with jumps, continuous-time Markov chains, and even scalar diffusions. While the general theory we develop in this article is well suited to address estimation issues in all such areas, the specific applications we choose to illustrate our methods cover the typical models arising in financial economics (latent factor models and diffusion models).

As is well-known, the major difficulty arising from the estimation of dynamic models with unobserved components is related to the complexity of evaluating the criterion functions. A natural remedy to this difficulty is to make use of simulation-based methods. The simulated method of moments (McFadden (1989), Pakes and Pollard (1989), Lee and Ingram (1991) and Duffie and Singleton (1993)), the simulated pseudo-maximum likelihood method of Laroque and Salanié (1989, 1993, 1994), the indirect inference approach of Gouriéroux, Monfort and Renault (1993) and Smith (1993), and the efficient method of moments (EMM) of Gallant and Tauchen (1996) represent the first attempts at addressing this problem through extensions of the generalized method of moments. The main characteristic of these approaches is that they are general-purpose. Their drawback is that they lead to inefficient estimators even in the case of fully observed systems. The only exception is the EMM, which becomes indeed efficient as the (parameter) dimension of the auxiliary score gets larger and larger - a condition known as "smooth embedding". There exist alternative simulation-based econometric methods, which directly approximate the likelihood function through simulations (e.g., Lee (1995) or Hajivassiliou and McFadden (1998)). These methods do lead to asymptotic efficiency. Yet all the estimators arising within this class of methods are designed to address very specific estimation problems.

More recently, the focus of the literature has shifted towards a search for estimators combining the attractive features of both moments generating techniques and ML. In addition to the EMM, two particularly important contributions in this area are Fermanian and Salanié (2004) and Carrasco, Chernov, Florens and Ghysels (2004). Precisely, Fermanian and Salanié (2004) introduced a general-purpose method in which the (intractable) likelihood function is approximated by kernel estimates obtained through simulations of the model of interest. The resulting estimator, called nonparametric simulated ML (NPSML) estimator, is then both consistent and asymptotically efficient as the number of simulations goes to infinity and the smoothing parameter goes to zero at some (typical) convergence rate. Carrasco, Chernov, Florens and Ghysels (2004) developed a general estimation technology which also leads to asymptotic efficiency in the case of fully observed Markov processes. Their method leads to a "continuum of moment conditions" matching model-based (simulated) characteristic functions to data-based characteristic functions.

This article belongs to this new strand of the literature. Our strategy is indeed to construct criterion functions leading to a general estimation approach. And in many cases of interest, these criterion functions are asymptotically equivalent to Neyman's chi-square measures of distance. It is precisely such an asymptotic equivalence which makes our resulting estimators asymptotically efficient. However, we emphasize that our estimators are quite distinct from any possible approximation to the MLE - they thus work rather differently from the Fermanian and Salanié NPSML estimator. In the language of indirect inference theory, we rely on "auxiliary criterion functions", which generally give rise to asymptotically inefficient but consistent estimators. But as soon as model and data's transition densities are estimated with a smoothing parameter converging to zero, these criterion functions converge to Neyman's chi-squares, and our estimator becomes efficient. In this sense, the role played by the smoothing parameter in our context parallels the role played by the smooth embedding condition within the EMM.[1] One distinctive feature of our method is that we allow the smoothing parameter to go to zero at a reasonably mild rate. Furthermore, we smooth model-generated data and observations with the same kernel. Therefore, the behavior of the smoothing parameter does not affect the *consistency* of the estimator. An asymptotically shrinking smoothing parameter can only affect the *precision* of our estimator.

Our method is also related to the estimators introduced by Carrasco, Chernov, Florens and Ghysels (2004). Indeed, our SNE also relies on a "continuum of moments", but in a very different manner. First, we do not need an infinite number of simulations to ensure consistency and

[1]We are grateful to one anonymous referee and Christopher Sims for bringing this point to our attention.

asymptotic normality of our estimators. Second, we use more classical ideas from the statistical literature, and develop estimating equations leading to match model-based density estimates (not characteristic functions) to their empirical counterparts. As for the NPSML estimator, the SNE is thus both conceptually very simple and fairly easy to implement. Earlier estimators based on ideas similar to ours include the ones introduced by Gallant (2001) and Billio and Monfort (2003). Precisely, Gallant (2001) estimator matches cumulative distributions, but it does not lead to asymptotic efficiency. Billio and Monfort (2003) estimator minimizes distances between observation-based and simulated-based expectations of test functions smoothed with kernel methods. While their estimator is not asymptotically efficient, it is still (up to identifiability) consistent independently of the amount of smoothing. Yet the rate of convergence of their estimator is nonparametric - although the rate of convergence to zero of their smoothing parameter can be made very slow. As noted earlier, the convergence rate of our estimator is the usual parametric one, but this attractive feature of our methods is obtained with one additional computational cost: To match nonparametric density estimates, the evaluation of our objective functions requires the computation of a Riemann integral.

Finally, Aït-Sahalia (1996) is one additional fundamental contribution which this article is clearly related to. Aït-Sahalia developed a minimum distance estimator in which the *measure of distance* is a special case of the general class of measures of distance we consider here. But our estimator is different for three additional important reasons. First, the asymptotic behavior of Aït-Sahalia's estimator critically depends on the smoothing parameter; as we argued earlier, our estimator is designed in a way that the smoothing parameter plays a relatively more marginal role. Second, Aït-Sahalia's estimator only matches marginal densities. Third, Aït-Sahalia's method does not rely on simulations; therefore, it is feasible only when the density implied by the model has a fairly tractable form.

The paper is organized in the following manner. Section 2 introduces basic notation and assumptions for the model of interest. Section 3 provides large sample theory. Section 4 illustrates how our methods can be used to estimate the typical diffusion models arising in financial economics. Section 5 assesses the finite sample properties of our estimators. Section 6 concludes. The appendix gathers proofs and regularity conditions omitted in the main text; and an unpublished appendix available upon request provides additional technical details.

2 The model of interest

Let $\Theta \subset \mathbb{R}^{p_\theta}$ be a compact parameter set, and for a given parameter vector $\theta_0 \in \Theta$, consider the following reduced-form data generating process:

$$y_{t+1} = f\left(y_t, \epsilon_{t+1}; \theta_0\right), \qquad t = 0, 1, \cdots, \tag{1}$$

where $y_t \in \mathbb{R}^d$, f is known and $\{\epsilon_t\}_{t=1,\cdots}$ is a sequence of $\mathbb{R}^d$-valued identically and independently distributed random variables (with known distribution). The purpose of this paper is to provide estimators of the true parameter vector θ_0. We consider a general situation in which some components of y are not observed. Accordingly, we partition vector y as:

$$y = \begin{pmatrix} y^o \\ \cdots \\ y^u \end{pmatrix},$$

where $y^o \in Y^o \subseteq \mathbb{R}^{q^*}$ is the vector of observable variables and $y^u \in Y^u \subseteq \mathbb{R}^{d-q^*}$ is the vector of unobservable variables. Data are collected in a $q^* \times T$ matrix with elements $\{y^o_{j,t}\}_{j=1,\cdots,q^*;t=1,\cdots,T}$, where $y^o_{j,t}$ denotes the t-th observation of the j-th component of vector y^o, and T is the sample size. Since our general interest lies in the estimation of partially observed processes, we may wish to recover as much information as possible about the dependence structure of the observables in (1). We thus set $q = q^*(1 + l)$, for some $l \geq 1$, let $y^o_t = (y^o_{1,t}, \cdots, y^o_{q^*,t})$ and

$$x_t \equiv \left(y^o_t, \cdots, y^o_{t-l}\right), \quad t = t_l \equiv 1 + l, \cdots, T, \tag{2}$$

and define $X \subseteq \mathbb{R}^q$ as the domain of x_t. In practice, there is a clear trade-off between increasing the highest lag l and both speed of computations *and* the curse of dimensionality. In Section 3.2, we succinctly present a few practical devices on how to cope with the curse of dimensionality.

Let $\pi(x; \theta)$ denote the joint density induced by (1) on x when the parameter vector is $\theta \in \Theta$. Let $\pi_0(x) \equiv \pi(x; \theta_0)$ and let $|\nabla_\theta \pi\,(x; \theta)|_2$ denote the outer product of vector $\nabla_\theta \pi\,(x; \theta)$. We now make assumptions further characterizing the family of processes we are investigating.

Assumption 1 (a) *$\pi(x; \theta)$ is continuous and bounded on $X \times \Theta$.* **(b)** *For all $x \in X$, function $\theta \mapsto \pi(x; \theta)$ is twice differentiable and its derivatives are bounded on Θ. Furthermore, f is continuous and twice differentiable on Θ.*

To ensure the feasibility of the asymptotic theory related to our estimation methods, we also need to make the following assumption on the decay of dependence in the observables in (1):

Assumption 2. *Vector y is a Markov β-mixing sequence with mixing coefficients β_k satisfying* $\lim_{k\to\infty} k^{\mu}\beta_k \to 0$, *for some* $\mu > 1$.

The mixing condition of assumption 2 is critical for the application of a functional central limit theorem due to Arcones and Yu (1994). Precisely, assumption 2 ensures convergence of suitably rescaled integrals of kernel functions to stochastic integrals involving generalized Brownian Bridges. This kind of convergence is exactly what we need to prove asymptotic normality of our estimators.

3 Theory

3.1 "Twin-smoothing"

Our estimation methodology is related to the classical literature on goodness-of-fit tests initiated by Bickel and Rosenblatt (1973). Let π_T be a nonparametric estimator of π_0, obtained as $\pi_T(x) \equiv (T\lambda^q)^{-1}\sum_{t=t_l}^{T} K\left((x_t - x)/\lambda\right)$, where $x \in \mathbb{R}^q$, the bandwidth $\lambda > 0$, and K is a symmetric bounded kernel of the r-th order.[2] Consider the following empirical measure of distance:

$$I_T(\theta) = \int_{\mathbb{R}^q} \left[\pi(x;\theta) - \pi_T(x)\right]^2 w_T(x)dx, \tag{3}$$

where $w_T > 0$ is a weighting function possibly depending on data, and θ is a given parameter value. Let $\hat{\theta}$ be some consistent estimator of θ_0. Typical measures of fit of the parametric model $\{\pi(\cdot;\theta),\ \theta \in \Theta\}$ to data are based on the empirical distance $I_T(\hat{\theta})$.[3] Alternatively, the empirical distance in (3) can be utilized to *estimate* the unknown parameter vector θ_0. For example, Aït-Sahalia (1996) defined an estimator minimizing (3) (with weighting function $w_T \equiv \pi_T$) in the context of scalar diffusions:

$$\theta_T^I = \arg\min_{\theta\in\Theta} I_T(\theta). \tag{4}$$

An important feature of the empirical measure of distance $I_T(\hat{\theta})$ is that a *parametric* density estimate, $\pi(\cdot;\hat{\theta})$, is matched to a *nonparametric* one, $\pi_T(\cdot)$. Under correct model specification, $\pi_T(x) \xrightarrow{p} K * \pi(x;\theta_0) \equiv \int_{\mathbb{R}^q} \lambda^{-q} K\left((u-x)/\lambda\right)\pi(u;\theta_0)\,du$ (x-pointwise). As is well-known, the result that $\pi_T(\cdot) \xrightarrow{p} \pi(\cdot;\theta_0)$ only holds if the bandwidth satisfies $\lambda \equiv \lambda_T$, $\lim_{T\to\infty}\lambda_T \to 0$ and $\lim_{T\to\infty} T\lambda_T^q \to \infty$. Therefore, bandwidth choice is critical for (3) and (4) to be really informative in finite samples.

[2] A symmetric kernel K is a symmetric function around zero that integrates to one. It is said to be of the r-th order if: 1) $\forall\mu \in \mathbb{N}^q : |\mu| \in \{1,\cdots,r-1\}$ ($|\mu| \equiv \sum_{j=1}^{q}\mu_j$), $\int u_1^{\mu_1}\cdots u_q^{\mu_q}K(u)du = 0$; 2) $\exists\mu \in \mathbb{N}^q : |\mu| = r$ and $\int u_1^{\mu_1}\cdots u_q^{\mu_q}K(u)du \neq 0$; and 3) $\int \|u\|^r K(u)du < \infty$.

[3] Precisely, rescaled versions of (3) are classically used to implement tests of model misspecification (see, e.g., Pagan and Ullah (1999) for a comprehensive survey on those tests). Corradi and Swanson (2005) have recently developed new specification tests for diffusion processes based on *cumulative* probability functions.

To circumvent this problem, we consider a measure of distance alternative to (3). A simple possibility is an empirical distance in which the nonparametric estimate π_T is matched by the model's density smoothed with the same kernel and conditional on a given bandwidth value:

$$L_T(\theta) = \int_{\mathbb{R}^q} [K * \pi(x;\theta) - \pi_T(x)]^2 w_T(x)\, dx. \tag{5}$$

Fan (1994) developed a class of bias-corrected goodness of fit tests based on the previous empirical distance and weighting function $w_T \equiv \pi_T$. And Härdle and Mammen (1993) devised a similar bias-correction procedure for testing the closeness of a parametric regression function to a nonparametric one.

A key idea in this paper is to combine the appealing idea underlying the estimator θ_T^I in (4) with the bias-corrected empirical measure in (5). To achieve this objective, we consider an estimator minimizing the distance in (5) rather than in (3), and consider a general empirical weighting function w_T. Specifically, define the following estimator:

$$\theta_T^L = \arg\min_{\theta\in\Theta} L_T(\theta), \tag{6}$$

where $w_T(x) \xrightarrow{p} w(x)$ uniformly, and w is another positive function. In (5), kernel smoothing operates in the same manner on model-implied density and on data-based density estimates. Therefore, bandwidth conditions affect the two estimators θ_T^I and θ_T^L in a quite different manner. Table 1 summarizes these bandwidth conditions. Under our regularity conditions, consistency of θ_T^L holds *independently of bandwidth behavior*. That is, up to identifiability (see assumption 3-(a) below), λ can be any strictly positive number. On the contrary, consistency of θ_T^I requires the additional conditions that $\lambda_T \to 0$ and $T\lambda_T^q \to \infty$.[4]

The "twin-smoothing" procedure underlying the estimator θ_T^L in (6) is intimately related to the general indirect inference strategy put forward in the seminal papers of Gouriéroux, Monfort and Renault (1993) and Smith (1993). In the language of indirect inference, we are matching a model-implied (infinite-dimensional) auxiliary parameter $(K * \pi(x;\theta))$ to the corresponding (infinite-dimensional) parameter computed on real data $(\pi_T(x))$. These auxiliary parameters can be estimated with an arbitrary bandwidth choice; yet, and up to identifiability, our estimator is still consistent in exactly the same spirit of the indirect inference principle.

Our basic idea is also related to the kernel-based indirect inference approach developed by Billio and Monfort (2003). Billio-Monfort estimator matches conditional expectations of arbitrary test-functions estimated through nonparametric methods - one conditional expectation computed

[4]Other estimators related to (4) suffer from exactly the same drawback. Two examples are 1) estimators based on nonparametric density estimates of the log-likelihood function obtained through simulations; and 2) estimators based on the so-called Kullback-Leibler distance (or relative entropy) $\int_{\mathbb{R}^q} \log[\pi(x,\theta)/\pi_0(x)]\pi(x,\theta)dx$. We are grateful to Oliver Linton for having suggested the latter example to us.

on true data and one conditional expectation computed on simulated data. This makes asymptotic bias issues irrelevant for their estimator. One important difference between our estimator θ_T^L in (6) and Billio-Monfort estimator is that our estimator is consistent at the usual parametric rate. The rate of convergence of Billio-Monfort estimator is contaminated by the rate of convergence of their bandwidth sequence to zero - although in practice the convergence of their bandwidth can be made very slow. Intuitively, Billio-Monfort estimator matches a finite number of test-functions. Instead, we match a continuum of moment conditions. But at the same time, this attractive feature of our estimator (matching a continuum of moments) brings an additional computational cost related to the evaluation of the Riemann integral in (5). Finally, our idea to directly focus on matching objects related to density functions resembles the "effective calibration" strategy of Gallant (2001). The main difference is that Gallant (2001) considers matching cumulative distribution functions. As we demonstrate in later sections, the advantage to focus on density functions is that it allows us to address efficiency issues.

Similarly as for consistency, θ_T^L and θ_T^I are asymptotically normally distributed under different bandwidth restrictions. Our estimator θ_T^L is asymptotically normal under the standard assumptions that $\lambda_T \to 0$ and $\sqrt{T}\lambda_T^q \to \infty$.[5] Instead, θ_T^I is asymptotically normal under one additional condition on the order of the kernel (i.e. $\sqrt{T}\lambda_T^r \to 0$). Intuitively, this order condition guarantees that a density bias estimate vanishes at an appropriate rate without affecting the asymptotic behavior of θ_T^I. In contrast, density bias issues are totally absent if one implements estimator θ_T^L. As summarized in Table 1, bandwidth restrictions are only required to make our estimator θ_T^L asymptotically normal - not consistent. And as we demonstrate in the Monte Carlo experiments of Section 5, bandwidth restrictions in Table 1 are considerably less critical for asymptotic normality than for consistency.

Table 1 - Bandwidth assumptions and asymptotic behavior of θ_T^I and θ_T^L

	Consistency	Asymptotic normality
θ_T^I	$T\lambda_T^q \to \infty$, $\lambda_T \to 0$	$\sqrt{T}\lambda_T^q \to \infty$, $\lambda_T \to 0$, and $\sqrt{T}\lambda_T^r \to 0$
θ_T^L	no asymptotic bandwidth restrictions	$\sqrt{T}\lambda_T^q \to \infty$, $\lambda_T \to 0$

3.2 Simulated Nonparametric Estimators

Our fundamental objective is to extend the previous ideas to general situations. Specifically, suppose that the analytical solution for density $\pi(x;\theta)$ in (5) is unknown, but that it is still possible to simulate from that density. Accordingly, the first step of our estimation strategy requires simulated paths of the observable variables in (1). To generate S simulated paths for a

[5] More sophisticated versions of our estimator are asymptotically normal under an additional assumption guaranteeing that certain derivatives of density estimates are well-behaved (i.e. $\sqrt{T}\lambda_T^{q+1} \to \infty$) (see theorem 1).

given parameter value θ, we draw $y_0(\theta)$ from its stationary distribution, and compute recursively

$$y_{t+1}(\theta) = f\left(y_t(\theta), \tilde{\epsilon}_{t+1}; \theta\right), \qquad t = 0, 1, \cdots, T,$$

where $\{\tilde{\epsilon}_t\}_{t=1}^{T+1}$ is a sequence of random numbers drawn from the distribution of ϵ. Let $x^i(\theta) = \left\{x_t^i(\theta)\right\}_{t=t_l}^{T}$ $(i = 1, \cdots, S)$, where $x_t^i(\theta)$ is the i-th simulation of the t-th observation when the parameter vector is θ, and define $y^i(\theta)$ in a similar way. Let $\pi_T^i(x;\theta) \equiv (T\lambda_T^q)^{-1} \sum_{t=t_l}^{T} K((x_t^i(\theta) - x)/\lambda_T)$, where K and λ are the same kernel and bandwidth functions used to compute the nonparametric density estimate $\pi_T(\cdot)$ on sample data.

We are now in a position to provide the definition of the first estimator considered in this paper:

Definition 1. (SNE) *For each fixed integer S, the* Simulated Nonparametric Estimator (SNE) *is the sequence $\{\theta_{T,S}\}_T$ given by*:

$$\theta_{T,S} = \arg\min_{\theta\in\Theta} \int_X \left[\tilde{\pi}_T(x;\theta) - \pi_T(x)\right]^2 w_T(x)dx, \tag{7}$$

where $\tilde{\pi}_T(\cdot;\cdot) \equiv S^{-1}\sum_{i=1}^{S}\pi_T^i(\cdot;\cdot)$ and $w_T(\cdot) > 0$ is a sequence of bounded and integrable functions satisfying $w_T(x) \xrightarrow{p} w(x)$, x-pointwise, for some function w.

The appealing feature of this estimator is that π_T^i and π_T are computed with the *same kernel* and *bandwidth*. Such a twin kernel smoothing procedure operates on sample and model generated data in exactly the same manner as in (5). Consequently, the asymptotic properties of θ_T in (7) and θ_T^L in (6) are quite comparable. Moreover, consistency of θ_T does not require an infinite number of simulations S. Even in correspondence of a finite number of simulations, the objective function in (7) is asymptotically equivalent to the objective function in (5). These two features of the SNE make our estimation strategy quite distinct from the estimation strategy introduced by Fermanian and Salanié (2004) - in which the likelihood function is directly approximated by kernel estimates of model-simulated data. But our approach also entails the additional computational cost related to the evaluation of the Riemann integral in (7).

We consider kernels satisfying the following regularity conditions:[6]

[6]Assumption K is needed to prove the lemmata in appendix A through Andrews (1995) strategy of proof. Andrews (1995) (assumption NP4(b), p. 566-567) required that $\int (1+\|z\|^r)\sup_{b\geq 1}|\Phi(bz)|\,dz < \infty$. Assumption K imposes the weaker condition that $\int |\Phi(z)|\,dz < \infty$. We deleted the $(1+\|z\|^r)$ multiplicand because y is strong mixing by assumption 2. The $\sup_{b\geq 1}$ requirement is not to be ignored in all applications with data-dependent bandwidths.

Assumption K. *Kernels K are bounded, continuously differentiable with bounded derivatives up to the fourth order, and absolutely integrable with an absolutely integrable Fourier transform* $\Phi(z) \equiv (2\pi)^q \int \exp\left(\mathtt{i}z^\top u\right) K(u)\, du$.

Let $L_T(\theta) \equiv \int \left[\tilde{\pi}_T(x;\theta) - \pi_T(x)\right]^2 w_T(x)dx$. Let the expectation of the kernel for a given bandwidth value λ be denoted as:

$$m(x;\theta) \equiv K * \pi(x;\theta) = \frac{1}{\lambda^q} \int K\left(\frac{x-u}{\lambda}\right) \pi(u;\theta)du.$$

Accordingly, set $L(\theta) \equiv \int \left[m(x;\theta) - m(x;\theta_0)\right]^2 w(x)dx$. Criteria are required to satisfy the following regularity *and* identifiability conditions:

Assumption 3 (a). *For all $\theta \in \Theta$, $L_T(\theta)$ is measurable and continuous on Θ a.s. Moreover, $L(\theta)$ is continuous on Θ, $\exists$ unique $\theta_0 : L(\theta_0) = 0$, and* $\liminf_{T\to\infty} \min_{\theta\in N_T^c} L_T(\theta) > 0$, *where N_T^c is the complement in Θ of a neighborhood of $\theta_{T,S}$.*

The first part of assumption 3-(a) is needed to ensure existence of our SNE, and holds under mild conditions on the primitive model. For example, it holds under the previous kernel assumption K, and the assumption that function f in (1) is continuous on Θ. The second part of this assumption merits further discussion. We are designing our estimator in such a way that bandwidth choice is virtually irrelevant for consistency. But to accomplish this task, we need to make sure that the (infinite-dimensional) "auxiliary" parameter $K * \pi$ has information content on the "structural" parameter θ. The last part of assumption 3-(a) then makes our SNE *identifiably unique*.[7] Consistency of the SNE requires the following additional assumption:

Assumption 3 (b). *There exists a $\alpha > 0$ and a sequence κ_T bounded in probability as T becomes large such that for all $(\varphi, \theta) \in \Theta \times \Theta$, $|L_T(\varphi) - L_T(\theta)| \leq \kappa_T \cdot \|\varphi - \theta\|_2^\alpha$.*

Assumption 3-(b) is a standard high level assumption. In Altissimo and Mele (2005, appendix F), we have developed specific examples of primitive conditions ensuring that assumption 3-(b) does hold. We now formulate one assumption we use to prove asymptotic normality of the SNE. Let $\mathcal{K}_T^j(x;\theta) \equiv \left| K'((x_t^i(\theta) - x)/\lambda_T)(\partial y_{\ell,t}^i(\theta)/\partial\theta_j) \right|$ ($j = 1, \cdots, p_\theta$ and $i = 1, \cdots, S$), where $y_{\ell,t}^i(\theta)$ is the i-th simulation at t of the ℓ-th component of x_t in (2) ($\ell = 1, \cdots, q^*$). We have:

[7] See, e.g., Gallant and White (1988, definition 3.2 p. 19). One referee suggested that identifiability may break down if the bandwidth λ is larger than the support of data. In Altissimo and Mele (2005, appendix F.1), we formalize this referee's suggestion and provide an example of kernels, bandwidth *levels* and data generating process (with *bounded* support) such that identifiability does break down. In Altissimo and Mele (2005), we also argue that if kernels satisfy assumption K and have for example unbounded support, the identifiably uniqueness condition in assumption 3-(a) holds with *sufficiently small* bandwidth values (not necessarily shrinking to zero).

Assumption 4 (a). *For all* $j = 1, \cdots, p_\theta$ *and* $(x, \theta) \in X \times \Theta$, $\mathcal{K}_T^j(x; \theta)$ *is continuous, bounded and satisfies assumption 2; and* $\partial \mathcal{K}_T^j(x; \theta) / \partial \theta_m$ *is bounded for all* $m = 1, \cdots, p_\theta$; $\partial^{\rho+1} \pi(x; \theta) / \partial \theta \partial x^\rho$ *is uniformly bounded for some* $\rho \geq r$. **(b)** $\sup_{x \in X} |w_T(x) - w(x)| = O_p(T^{-\frac{1}{2}} \lambda_T^{-q}) + O_p(\lambda_T^r)$.

All in all, assumption 4 on $\mathcal{K}_T^j$ is needed to make the first order conditions satisfied by the SNE analytically tractable. (Basically, it allows one to interchange the order of derivation and integration in $\nabla_\theta L_T(\theta)$.) The assumption on $\partial^{\rho+1} \pi(x; \theta) / \partial \theta \partial x^\rho$ ensures uniform convergence of score functions to their asymptotic counterparts (see lemmata 5 to 10 in appendix A). Finally, the assumption on the weighting function w_T is obviously under the investigator's control. As an example, one may take $w_T(x) \equiv \pi_T(x) \gamma(x)$, where γ is another function. By lemma 1 in appendix A, this choice satisfies assumption 4-(b).

The following result provides the asymptotic properties of the SNE:

Theorem 1. *Let assumptions 1-(a), 2 and 3 hold*; *then, the SNE is* (*weakly*) *consistent. Furthermore, let* $\Psi(x) \equiv [\int |\nabla_\theta \pi(u; \theta_0)|_2 w(u) du]^{-1} \nabla_\theta \pi(x; \theta_0) w(x)$. *Then, under the additional assumption 1-(b) and 4, and the conditions that* $\lambda \equiv \lambda_T \to 0$ *and* $T^{\frac{1}{2}} \lambda_T^{q+1} \to \infty$ *as* $T \to \infty$,

$$\sqrt{T}(\theta_{T,S} - \theta_0) \xrightarrow{d} N\left(0, \left(1 + \frac{1}{S}\right) V\right),$$

where $V \equiv var[\Psi(x_1)] + \sum_{k=1}^{\infty} \{cov[\Psi(x_1), \Psi(x_{1+k})] + cov[\Psi(x_{1+k}), \Psi(x_1)]\}$, *provided it exists finitely.*

Proof. In appendix B. ■

The asymptotic theory underlying the SNE displays four basic distinctive features. First, *and up to identifiability* (see assumption 3-(a)), consistency does *not* rely on any condition regarding the bandwidth parameter. The only bandwidth conditions we actually need only ensure that the SNE is asymptotically normal. In particular, the order of the kernel plays no role within our asymptotic theory.[8] We shall see that this conclusion is only slightly modified even in more sophisticated versions of our basic estimator (see theorems 2 and 3 below).

Second, the (unscaled) variance V of theorem 1 collapses to the variance of the estimator in Aït-Sahalia (1996) in the scalar case and when $w_T = \pi_T$. However, we emphasize that the two estimators are radically different. Aït-Sahalia (1996) requires an analytical form of the model's density and, consequently, consistency of his estimator may only follow if both $\lambda_T \to 0$ and $T\lambda_T^q \to \infty$. The twin-smoothing procedure makes our SNE considerably less sensitive to

[8]The main technical reason explaining this result is that conditions such as $\sqrt{T} \cdot \lambda_T^r \to 0$ would be important if the theory required a functional limit theorem for $\sqrt{T}(\int \pi_T - \int \pi_0)$. We do not need such a demanding result. We only need a functional limit theorem for $\sqrt{T}(\int \pi_T - \int E(\pi_T))$.

bandwidth issues - a fact also documented in our Monte Carlo experiments.[9] Furthermore, the SNE can address estimation of multivariate models driven by partially observed state variables with unknown distribution. Also, we explicitly consider matching *joint* densities of data, not marginal densities. Finally, the SNE minimizes a measure of closeness of two nonparametric density estimates - one on true data and the second on simulated data. Under correct model's specification, the resulting biases in the two kernel estimates cancel out each other, and asymptotic normality can then be obtained *without* relying on any bias-reducing devices. For all these reasons, the SNE is potentially apt to exhibit a finite sample behavior that is well approximated by the asymptotic theory. And such a finite sample behavior is indeed documented by our Monte Carlo experiments in Section 5.

Third, our SNE makes use of general weighting functions. If $w_T = \pi_T$, the corresponding SNE would overweight discrepancies occurring where observed data have more mass. More generally, Theorem 1 reveals that the asymptotic variance of the estimator depends indeed on the limiting weighting function w at hand. However, a weighting function minimizing such an asymptotic variance is unknown, even in the case of fully observable processes.[10] In the next Section, we show that this problem can considerably be simplified through an appropriate change of the objective function in (7).[11]

Fourth, the estimator's variance has to be rescaled by $(1 + S^{-1})$ - similarly as in the familiar asymptotics of Indirect Inference estimators (e.g., Gouriéroux, Monfort and Renault (1993)). This scaling term arises because the model's joint density is recovered by means of simulations.

As for other nonparametric density based-estimators, the SNE is subject to the curse of dimensionality. But as in related contexts, the SNE can be extended to mitigate this issue. As an example, we may let,

$$\theta_{T,S} = \arg\min_{\theta\in\Theta} \sum_{\ell=1}^{l} L_T^{(\ell)}(\theta), \quad L_T^{(\ell)}(\theta) \equiv \int_{X^\ell} \left[\tilde{\pi}_T(x^\ell;\theta) - \pi_T(x^\ell)\right]^2 w_T(x^\ell)dx^\ell, \quad X^\ell \subseteq \mathbb{R}^{2q^*},$$

where $x_t^\ell = (y_t^o,\, y_{t-\ell}^o)$ (see eq. (2)). In proposing the above estimator, we imitated Fermanian and Salanié (2004, Section 4), who also considered addressing dimensionality issues through the use of lagged observable variables. But even when the dimension of the model's observables q^* is small, in practice l should be a small number given the current state of computational power. In

[9]Accordingly, the Aït-Sahalia's estimator could also be modified through the bias correction procedure we suggested in (5).

[10]An exception arises exactly in the i.i.d. case. Under regularity conditions given in section 3.3, the optimal weighting function would be given by $w_T(x) = \mathbb{T}_T(x) \cdot \pi_T(x)^{-1}$, where $\mathbb{T}_T(\cdot)$ is a trimming function converging pointwise to 1 as $T \to \infty$.

[11]Pastorello, Patilea and Renault (2003) have recently proposed a "latent backfitting" method to estimate partially observed systems through information provided by standard economic theory. In Altissimo and Mele (2005, appendix G), we have extended the theory in this paper to the ideal situation in which partially observed systems are estimated in conjunction with asset pricing models holding without measurement error.

tests involving stochastic volatility models, the SNE computed with $l = 1$ (i.e. with a matching of the joint density of two adjacent observations) had a very encouraging behavior (see Section 5). Dimensionality issues related to the spatial dimension q^* can be mitigated in the same vein.

3.3 Conditional Density SNE, and Efficiency

This section introduces a modification of the SNE, and addresses efficiency issues within the case of fully observable diffusions. We show that by casting the estimation problem as a matching of *conditional* densities (instead of *joint* ones), our resulting estimator is asymptotically (first-order) efficient whenever the state y in (1) is fully observable.

To prepare the analysis, consider again vector $x \in X \subseteq \mathbb{R}^q$ in (2). For each t, partition x_t as $x_t = (z_t, v_t)$, where $z_t \equiv y_t^o \in Z \subseteq \mathbb{R}^{q^*}$ is the vector of observable variables, and $v_t \in V \subseteq \mathbb{R}^{q-q^*}$, is the vector of predetermined variables:

$$v_t \equiv \left(y_{t-1}^o, \cdots, y_{t-l}^o\right), \quad t = t_l \equiv 1 + l, \cdots, T.$$

Consider the following conditional density matching estimator:

Definition 2. (CD-SNE) *For each fixed integer S, the* Conditional Density SNE (CD-SNE) *is the sequence $\{\theta_{T,S}\}_T$ given by*:

$$\theta_{T,S} = \arg\min_{\theta \in \Theta} \int_Z \int_V \left[\tilde{\pi}_T\left(z|\, v;\theta\right) - \pi_T\left(z|\, v\right)\right]^2 w_T(z,v) \mathbb{T}_{T,\delta}^2\left(v;\theta\right) dz dv, \tag{8}$$

where $\pi_T\left(z|\, v\right) \equiv \pi_T\left(z,v\right)/\,\pi_T\left(v\right)$, $\tilde{\pi}_T\left(z|\, v;\theta\right) \equiv S^{-1}\sum_{i=1}^S \pi_T^i\left(z,v;\theta\right)/\,\pi_T^i\left(v;\theta\right)$, $\{\mathbb{T}_{T,\delta}\}_T$ is a sequence of trimming functions satisfying assumption **T** *below, and $w_T > 0$ is a sequence of weighting functions satisfying assumption 4-(b).*

The CD-SNE relies on nonparametric conditional density estimates obtained as ratios between joints over marginals. Small values of the denominators in $\pi_T\left(z|\, v\right)$ may hinder numerical stability of the estimator, and the asymptotic theory. Therefore, we need to control the tail behavior of marginal density estimates. The role of trimming function $\mathbb{T}_{T,\delta}$ is to accommodate this task. Trimming functions are widely used in related contexts (see, e.g., Stone (1975), Bickel (1982), or more recently, Linton and Xiao (2000) and Fermanian and Salanié (2004)). In this paper, we consider trimming devices related to the original work of Andrews (1995).

Assumption T. *Let g be a bounded, twice differentiable density function with support* $[0,1]$, $g\left(0\right) = g\left(1\right) = 0$, *and let $g_\delta\left(u\right) \equiv \frac{1}{\delta}g(\frac{u}{\delta} - 1)$. We set, $\mathbb{T}_{T,\delta}\left(v;\theta\right) \equiv \prod_{i=0}^S \mathbb{T}\left(\pi_T^i\left(v;\theta\right)\right)$ $(\pi_T^0\left(\cdot\right) \equiv \pi_T\left(\cdot\right))$, where $\mathbb{T}\left(\ell\right) \equiv \int_0^\ell g_{\delta_T}\left(u\right) du$, for some sequence $\delta_T \rightarrow 0$.*

By construction, $\mathbb{T}_{T,\delta}$ is increasing, smooth and satisfies, $\mathbb{T}_{T,\delta}(v;\theta) = 0$ on $\{v : \pi_T^i(v;\theta) < \delta_T, i = 0, 1, \cdots, S\}$; and $\mathbb{T}_{T,\delta}(v;\theta) = 1$ on $\{v : \pi_T^i(v;\theta) > 2\delta_T, i = 0, 1, \cdots, S\}$. As $T \to \infty$ and $\delta_T \to 0$, and under additional regularity conditions, $\pi_T(z|v) \xrightarrow{p} \pi(z|v)$ - uniformly over expanding sets on which the trimming function $\mathbb{T}_{T,\delta}$ is nonzero (see lemma 3 in appendix A). In appendix C (see assumptions *T1-(a,b)*) we gather all regularity conditions on the asymptotic behavior of δ_T we need to demonstrate consistency and asymptotic normality of the CD-SNE.[12]

Let $\bar{L}_T(\theta) \equiv \iint [\tilde{\pi}_T(z|v;\theta) - \pi_T(z|v)]^2 w_T(x)dx$. We define the asymptotic counterpart of $\bar{L}_T$ as $\bar{L}(\theta) \equiv \iint [n(z,v;\theta) - n(z,v;\theta_0)]^2 w(x)dx$, where $n(z,v;\theta) \equiv m(z,v;\theta)/m(v;\theta)$. To prove consistency of the CD-SNE, we need conditions paralleling the ones in assumption 3:

Assumption 5. *$\bar{L}_T$ and $\bar{L}$ are as L_T and L in assumption 3-(a), and for all $(\varphi,\theta) \in \Theta \times \Theta$, $\left|\bar{L}_T(\varphi) - \bar{L}_T(\theta)\right| \leq \kappa_T \cdot \|\varphi - \theta\|_2^\alpha$, where α and κ_T are as in assumption 3-(b).*

The following result provides the asymptotic properties of the CD-SNE.

Theorem 2. *Let assumptions 1-(a), 2, 5 and assumption T1-(a) in appendix C hold; then the CD-SNE is (weakly) consistent. Under the additional assumptions 1-(b) and 4, and assumption T1-(b) in appendix C,*

$$\sqrt{T}(\theta_{T,S} - \theta_0) \xrightarrow{d} N(0, V),$$

where $V \equiv D_3^{-1} \cdot var[\frac{1}{S}\sum_{i=1}^S (D_1^i - D_2^i) - (D_1^0 - D_2^0)] \cdot D_3^{\top -1}$, provided it exists finitely; and the terms $\{D_1^i\}_{i=0}^S$, $\{D_2^i\}_{i=0}^S$ and D_3 are given in appendix C.2.

Proof. In appendix C. ■

The variance structure of the CD-SNE differs from the one in the asymptotic distribution of the SNE (see Section 3.2). In the CD-SNE case, one has to cope with additional terms arising because conditional densities are estimated as ratios of two densities (joints over marginals). These additional terms are $\{D_2^i\}_{i=0}^S$. As we show in appendix D.2, there exist weighting functions w_T making these terms identically zero. In those cases, the variance terms in theorem 2 have the same representation as the variance terms in Section 3.2. Proposition 2 in appendix D.2 summarizes our results on these issues.

We now argue that as soon as y in (1) is fully observable, there exists a weighting function w_T making the CD-SNE asymptotically attain the Cramer-Rao lower bound. Precisely, let,

$$w_T(z,v) = \frac{\pi_T(v)^2}{\pi_T(z,v)} \mathbb{T}_{T,\alpha}(z,v), \quad \mathbb{T}_{T,\alpha}(z,v) \equiv \mathbb{T}_\alpha(\pi_T(z,v;\theta)), \tag{9}$$

[12] Linton and Xiao (2000) suggested the following example of trimming functions with a closed-form solution. Let the Beta-type density $g(u) \propto z^k(1-z)^k$ (for some integer k); then $\mathbb{T}(\ell)$ is a $(2k+1)$-polynomial in $(\ell - \delta_T)/\delta_T$.

where $\mathbb{T}_{\alpha}(\ell) \equiv \int_0^{\ell} g_{\alpha_T}(u)\,du$, and g_{α_T} is as in assumption **T**. Similarly as for the CD-SNE in definition 2, $\mathbb{T}_{T,\alpha}(z,v)$ is a trimming function needed to control the tail behavior of the joint density estimate on sample data. If w_T is as in (9), the criterion in (8) reduces to:

$$\int_Z \int_V \left[\frac{\tilde{\pi}_T(z|v;\theta)}{\pi_T(z|v)} - 1\right]^2 \pi_T(z,v)\,\mathbb{T}^2_{T,\delta}(v;\theta)\,\mathbb{T}_{T,\alpha}(z,v)\,dzdv,$$

which asymptotically becomes a Neyman's chi-squared measure of distance. A Taylor's expansion of the first order conditions satisfied by the CD-SNE around θ_0 yields that in large samples,

$$\begin{aligned} -J_T(\theta_0)\cdot\sqrt{T}\,(\theta_{T,S}-\theta_0) &\overset{d}{=} \iint_{\mathbb{T}} \left[\frac{\tilde{\pi}_T(z|v;\theta_0)}{\pi_T(z|v)} - 1\right]\left[\frac{\nabla_\theta\tilde{\pi}_T(z|v;\theta_0)}{\pi_T(z|v)}\right]\pi_T(z,v)\,dzdv \\ &\overset{d}{=} \iint_{\mathbb{T}} \left[\tilde{\pi}_T(z,v;\theta_0) - \pi_T(z,v)\right]\cdot\nabla_\theta \ln\pi(z|v;\theta_0)\cdot dzdv \\ &= \frac{1}{S}\sum_{i=1}^{S} H_T^i(\theta_0) - H_T^0(\theta_0) \end{aligned} \qquad (10)$$

where

$$\begin{aligned} J_T(\theta_0) &= \iint_{\mathbb{T}} \left|\frac{\nabla_\theta\tilde{\pi}_T(z|v;\theta_0)}{\pi_T(z|v)}\right|_2 \pi_T(z,v)\,dzdv \\ H_T^i(\theta_0) &= \iint_{\mathbb{T}} \left\{\pi_T^i(z,v;\theta_0) - E\left[\pi_T^i(z,v;\theta_0)\right]\right\}\cdot\nabla_\theta\ln\pi(z|v;\theta_0)\cdot dzdv, \quad i=0,1,\cdots,S \end{aligned}$$

(with $\pi_T^0 \equiv \pi_T$) and integrals with a subscript $\mathbb{T}$ are integrals trimmed under the action of functions $\mathbb{T}_{T,\alpha}$ and $\mathbb{T}_{T,\delta}$. But $J_T(\theta_0)$ and $H_T^i(\theta_0)$ satisfy $J_T(\theta_0) \overset{p}{\to} E\left[|\nabla_\theta \ln\pi(z_1|v_1;\theta_0)|_2\right]$ and $H_T^i(\theta_0) \overset{d}{\to} N(0, \text{var}(\nabla_\theta\ln\pi(z|v;\theta_0)))$ $(i=0,1,\cdots,S)$ (see appendixes C.2 and D.2 for technical details on such a law of large numbers and central limit theorem[13]). Since the system is fully observable and Markov, $z_t = y_t$, and $\nabla_\theta \ln\pi(y_t|y_{t-1};\theta_0)$ is a martingale difference with respect to the sigma-fields generated by y. Therefore, the variance of the CD-SNE (rescaled by $(1+S^{-1})$) does attain the Cramer-Rao lower bound $E\left[|\nabla_\theta\ln\pi(y_2|y_1;\theta_0)|_2\right]^{-1}$.

The previous arguments are obviously heuristic. For example, one critical issue is to ensure that as $\alpha_T \to 0$, the weighting function in (9) $w_T(z,v) \overset{p}{\to} w(z,v)$ - uniformly over expanding sets on which $\mathbb{T}_{T,\alpha}$ is nonzero (see lemma 2 in appendix A). In appendix D, we gather all joint asymptotic restrictions on α_T and δ_T leading to consistency and asymptotic normality of the CD-SNE with weighting function as in (9) (see assumption *T-2(a,b)*). We have:

[13] The central limit theorem can be understood heuristically as follows. Consider approximating $H_T^{\cdot}(\theta_0)$ with $\tilde{H}_T \equiv \iint \omega(x)\,dA_T(x)$, where $x \equiv (z,v)$, $A_T(x) = \sqrt{T}\left[F_T(x) - E(F_T(x))\right]$, $\omega \equiv \nabla_\theta\ln\pi$, and $F_T(x) = \frac{1}{T}\sum_{t=t_l}^{T}\mathbb{I}_{x_t\le x}$. We have, $\tilde{H}_T = T^{-\frac{1}{2}}\sum_{t=t_l}^{T}\iint\omega(x)\left(d\mathbb{I}_{x_t\le x} - E\left(d\mathbb{I}_{x_t\le x}\right)\right) = T^{-\frac{1}{2}}\sum_{t=t_l}^{T}\left[\omega(x_t) - E(\omega(x_t))\right]$, where the last equality holds because $d\mathbb{I}_{x_t\le x} = \delta(x-x_t)\,dx$, where $\delta(\cdot)$ is the Dirac's delta. Now apply the central limit theorem to conclude.

Theorem 3. (Cramer-Rao lower bound) *Suppose that the state is fully observable* (*i.e.*, $q^* = d$). *Let the CD-SNE match one-step ahead conditional densities (i.e.,* $(z, v) \equiv (y_t, y_{t-1})$ *in (8)) and let* w_T *be as in (9). Let assumptions 1-(a), 2, 5 and assumption T2-(a) in appendix D hold*; *then, the CD-SNE is* (*weakly*) *consistent. Under the additional assumptions 1-(b), 4-(a) and assumption T2-(b) in appendix D, the CD-SNE is as in theorem 2, and it attains the Cramer-Rao lower bound as* $S \to \infty$.

Proof. In appendix D. ■

The previous efficiency result follows because the weighting function in (9) makes the CD-SNE asymptotically equivalent to the score as soon as the system is *fully* observable (see eq. (10)). We emphasize that this property corresponds to the classical first-order efficiency criterion in Rao (1962). Furthermore, results by which estimators based on closeness-of-density retain efficiency properties are not a novelty in the statistical literature. In the context of independent observations with discrete distributions, Lindsay (1994) presented a class of estimators encompassing a number of minimum disparity estimators based on Hellinger's distance, Pearson's chi-square, Neyman's chi-square, Kullback-Leibler distance, and maximum likelihood. Lindsay showed that while all these estimators are first-order efficient, they may differ in terms of second-order efficiency, and robustness. Basu and Lindsay (1994) extended this theory to the case of continuous densities. Such an extension can be used to illustrate some fundamental properties of our estimator. In the i.i.d. case, our CD-SNE can be thought of as a member belonging to a general class of minimum disparity estimators θ_T defined by the following estimating equation:

$$0 = \int_{\mathbb{T}} \mathcal{A}\left(\phi(x)\right)\left[\nabla_\theta\left(K * \pi\left(x; \theta_T\right)\right)\right] dx, \qquad \phi(x) \equiv \frac{\pi_T(x) - K * \pi\left(x; \theta_T\right)}{K * \pi\left(x; \theta_T\right)},$$

where $\mathcal{A}$ is an increasing continuous function in $(-1, \infty)$.[14] Under regularity conditions, function $\mathcal{A}$ determines how sensitive an estimator is to the presence of outliers. Indeed, function ϕ is high exactly when a point in the sample space has been accounted much more than predicted by the model. Accordingly, a robust estimator is one able to mitigate the effect of large values of ϕ. As a benchmark example, the likelihood disparity sets $\mathcal{A}(\phi) = \phi$. Estimators with the property that $\mathcal{A}(\phi) \ll \phi$ for large ϕ are more robust to the presence of outliers than maximum likelihood. For instance, the Hellinger's distance sets $\mathcal{A}(\phi) = 2[\sqrt{\phi + 1} - 1]$, and the Kullback-Leibler distance has $\mathcal{A}(\phi) = \ln\left(1 + \phi\right)$. It is easily seen that if $w_T = \pi_T\left(x\right)^{-1} \mathbb{T}_{T,\alpha}\left(x\right)$, our L_T is asymptotically a Neyman's chi-squared measure of distance, with $\mathcal{A}\left(\phi\right) = \phi / \left(1 + \phi\right)$. These simple facts suggest that the class of estimators that we consider displays interesting robustness properties.

Naturally, the aim of theorem 3 was to extend the above class of estimators to the case of dynamic models. However, we do not further investigate robustness properties of our estimators.

[14] As $\lambda \downarrow 0$, $\mathcal{A}$ and ϕ collapse to Lindsay's (1994) adjustment function and Pearson's residual, respectively.

Using robustness, and/or second-order efficiency criteria as discrimination devices of alternative parameter estimators of dynamic models is an interesting area that we leave for future research.

4 Applications to continuous-time financial models

All available simulation-based techniques (and the methods developed in this article) rest on the obvious assumption that the model of interest can be simulated in a simple manner. Unfortunately, continuous-time models can not even be simulated - except in the trivial case in which the transition density is known.[15] The simple reason is that a continuous-time model can only be *imperfectly* simulated through some discretization device. In this section, we show that our theory still works if we allow the discretization to shrink to zero at an appropriate rate.

4.1 The model

Let $\Theta \subset \mathbb{R}^{p_\theta}$ be a compact parameter set, and for a given parameter vector $\theta_0 \in \Theta$, consider the following data generating process $y = \{y(\tau)\}_{\tau \geq 0}$:

$$dy(\tau) = b\left(y(\tau), \theta_0\right) d\tau + a\left(y(\tau), \theta_0\right) dW(\tau),\ \tau \geq 0, \tag{11}$$

where W is a standard d-dimensional Brownian motion; b and a are vector and matrix valued functions in $\mathbb{R}^d$ and $\mathbb{R}^{d\times d}$, respectively; a is full rank almost surely; and y takes values in $Y \subseteq \mathbb{R}^d$. As in Section 3, we partition y as $y = (y^o \vdots y^u)$, where $y^o \in Y^o \subseteq \mathbb{R}^{q^*}$ is the subvector of observable variables. Data are assumed to be sampled at regular intervals, and we still let $q \equiv q^*(1+l)$ and $x_t \equiv \left(y_t^o, \cdots, y_{t-l}^o\right)$ $(t = 1+l, \cdots, T)$, where $\{y_t^o\}_{t=1}^T$ is the observations sequence and T is the sample size. We consider the following regularity condition:

Maintained assumptions. *System (11) has a strong solution and it is strictly stationary. Furthermore, assumptions 1 and 2 (with mixing coefficients $\bar{\beta}_k$ and exponent $\bar{\mu} > 1$, say) hold in the context of model (11).*

Chen, Hansen and Carrasco (1999) provide primitive conditions guaranteeing that assumption 2 holds in the case of scalar diffusions. A scalar diffusion is β-mixing with exponential decay if their "pull measure", defined as $\frac{b}{a} - \frac{1}{2}\frac{\partial a}{\partial y}$, is negative (positive) at the right (left) boundary (the authors also provide conditions ensuring β-mixing with polynomial decay in the case of zero pull measure at one of the boundaries (see their remark 5)). As regards multidimensional diffusions,

[15] To date, estimation methods specifically designed to deal with diffusion processes include moments generating techniques (e.g., Hansen and Scheinkman (1995), Singleton (2001)), approximations to maximum likelihood (e.g., Pedersen (1995) and Santa-Clara (1995), and Aït-Sahalia (2002, 2003)) and, on a radically different perspective, Markov Chain Monte Carlo approaches (e.g., Elerian, Chib and Shephard (2001)).

β-mixing with exponential decay can be checked through results developed by Meyn and Tweedie (1993) for exponential ergodicity, as in Carrasco, Hansen and Chen (1999). Finally, Carrasco, Hansen and Chen (1999) provide more specific results pertaining to partially observed diffusions.

4.2 Estimation

To generate simulated paths of the observable variables in (11), various discretization schemes can be used (see, e.g., Kloeden and Platen (1999)). In this paper, we consider the simple Euler-Maruyama discrete time approximation to (11):

$$_h y_{h(k+1)} - {}_h y_{hk} = b\left({}_h y_{hk}, \theta\right) \cdot h + a\left({}_h y_{hk}, \theta\right) \cdot \sqrt{h} \cdot \epsilon_{k+1}, \qquad k = 0, 1, \cdots, \tag{12}$$

where h is the discretization step and $\{\epsilon_k\}_{k=1,\cdots}$ is a sequence of independent $\mathbb{R}^d$-valued i.i.d. random variables. Let $x_h^i(\theta) = \{x_{t,h}^i(\theta)\}_{t=t_l}^T$ denote the "pseudo"-skeleton of the i-th simulation path ($i = 1, \cdots, S$) at the parameter value θ.[16] That is, $x_{t,h}^i(\theta)$ is the i-th simulation of the t-th observation when the parameter vector is θ. Finally, define $y_h^i(\theta)$ in the same way.

The behavior of the high frequency simulator is regulated by the following conditions:

Assumption D.1. *For all $\theta \in \Theta$,* **(a)** *The high frequency simulator (12) converges weakly*[17] *to the solution of (11) i.e., for each i, $y_h^i(\theta) \Rightarrow y(\theta)$ as $h \downarrow 0$.* **(b)** *The diffusion and drift functions a and b are Lipschitz continuous in y; their components are four times continuously differentiable in y; and a, b and their partial derivatives up to the fourth order have polynomial growth in y.* **(c)** *Finally, as $h \downarrow 0$ and $T \to \infty$:* **(c.1)** $h \cdot \sqrt{T} \to 0$; *or* **(c.2)** $h \cdot T \to 0$.

The maintained assumption that (11) is stationary implies that the "observed skeleton" of the diffusion inherits the same features of the continuous-time process. Since the simulation step h can not be zero in practice, we extend assumption 2 to cover the "pseudo"-skeleton behavior:

Assumption D.2. *For all $\theta \in \Theta$, $\exists h^0 > 0$ depending on θ : for all $h \in \left(0, h^0\right)$, $y_h^i(\theta)$ is β-mixing with mixing coefficients $\beta_k(h)$: $\lim_{k\to\infty} \max_{h\in(0,h^0)} k^{\mu_h} \beta_k(h) \to 0$ for some sequence $\{\mu_h\}_h > 1$; and $\lim_{h\downarrow 0} \mu_h = \bar{\mu}$, $\lim_{h\downarrow 0} \beta_k(h) = \bar{\beta}_k$, where $\bar{\mu}$ and $\bar{\beta}_k$ are as in the maintained assumptions.*

Primitive conditions ensuring that assumption D.1-(a) holds are well-known and can be found, for instance, in Kloeden and Platen (1999). Primitive conditions guaranteeing that assumption D.2 holds are also well-known (see, e.g., Tjøstheim (1990) for conditions ensuring that (12)

[16] We used the wording "pseudo"-skeleton because h is nonzero.

[17] Let $(y_{hk})_{k=1}^{\infty}$ be a discrete time Markov process, and $(y(\tau))_{\tau\geq 0}$ be a diffusion process. When the probability laws generating the entire sample paths of $(y_{hk})_{k=1}^{\infty}$ converge to the probability laws generating $(y(\tau))_{\tau\geq 0}$ as $h \downarrow 0$, $(y_{hk})_{k=1}^{\infty}$ is said to converge weakly to $(y(\tau))_{\tau\geq 0}$.

is exponentially ergodic for fixed h). Assumptions D.1-(b,c) make our estimators asymptotically free of biases arising from the *imperfect* simulation of model (11) (model (11) is imperfectly simulated so long as $h > 0$). Precisely, such biases arise through terms taking the form $\sqrt{T}[E(K(x_{t,h}^{i}(\theta_0)) - E(K(x_t))]$, where K is a symmetric bounded kernel. But by results summarized in Kloeden and Platen (1999, chapter 14), $\sqrt{T}[E(K(x_{t,h}^{i}(\theta_0)) - E(K(x_t))] = O(h \cdot \sqrt{T})$ whenever assumptions D.1-(a,b) hold and K is as differentiable as a and b are in assumption D.1-(b). The role of assumption D.1-(c) is then to asymptotically eliminate such bias terms. Naturally, more precise high frequency simulators would allow h to shrink to zero at an even lower rate. Finally, assumption D.1-(b) can considerably be weakened. For example, one may simply require that a, b be Hölder continuous, as in Kloeden and Platen (1999, theorem 14.1.5 p. 460). These extensions are not considered here to keep the presentation as simple as possible.

Let $L_{T,h}$ and $\bar{L}_{T,h}$ be the criterions of the SNE (definition 1) and the CD-SNE (definition 2), and consider a sequence $\{h_T\}_T$ of discretization stepsizes converging to zero. Let $\mathcal{K}_{T,h}^{j}(x;\theta)$ be defined similarly as $\mathcal{K}_{T}^{j}(x;\theta)$ in Section 3.2. We need the following regularity conditions:

Assumption D.3 **(a)** *Either* **(a.1)** $L_{T,h}$ *satisfies assumption 3; or* **(a.2)** $\bar{L}_{T,h}$ *satisfies assumption 5.* **(b)** *With* $\mathcal{K}_{T,h}^{j}$ *replacing* $\mathcal{K}_{T}^{j}$, **(b.1)** *assumption 4 holds; or* **(b.2)** *assumption 4-(a) holds.*

Assumptions D.1-D.3 are the additional assumptions we need to prove that our estimators work as in the previous Section 3. Precisely, the following theorem is proven in Altissimo and Mele (2005, appendix E).

Theorem D.1. *Let assumptions D.1-(a,b) and D.2 hold. Then, under the additional assumptions D.1-(c.1) and D.3-(a.1,b.1), the SNE is as in theorem 1; under the additional assumptions D.1-(c.2) and D.3-(a.2,b.1), the CD-SNE is as in theorem 2; and under the additional assumptions D.1-(c.2) and D.3-(a.2,b.2), the CD-SNE is as in theorem 3.*

5 Monte Carlo experiments

In this section we conduct Monte Carlo experiments to investigate finite sample properties of our estimators. We wish to address four points: First, we wish to ascertain whether the finite sample properties of our estimators are accurately approximated by the asymptotic theory. Second, we study how our SNE and CD-SNE compare with alternative estimators such as the Fermanian and Salanié (2004) NPSML estimator, and even the MLE. Third, we examine how the SNE and the CD-SNE compare with each other. And fourth, we investigate how bandwidth choice and the possible curse of dimensionality impart on our estimators' finite sample performance.

To address these points, we consider four distinct models: Two continuous-time models commonly utilized in finance (namely, the standard Vasicek model and one simple extension of the

Vasicek model with stochastic volatility); and two discrete-time stochastic volatility models (one univariate and one bivariate). Our experiments on all these models share some common features. First, nonparametric density estimates are implemented through Gaussian kernels. Second, our bandwidth choice closely follows the suggestions made by Chen, Linton and Robinson (2001) in the context of conditional density estimation with dependent data; precisely, for each Monte Carlo replication, we select the bandwidth by searching over values minimizing the asymptotic mean integrated squared error of the conditional density estimated on *sample* data. Third, we trim 2% of the observations. Fourth, we set the number of path simulations equal to 5 in all experiments (i.e. $S = 5$). Fifth, in cases in which our estimators can not be efficient, asymptotic standard deviations are approximated through Newey-West windows of ∓ 12. Sixth, we run 1000 Monte Carlo replications in each experiment. Finally, the experiments related to the continuous-time models are implemented with data sampled at weekly frequency; and models simulated through the Euler-Maruyama scheme (12) with stepsize $h = 1/(5 \times 52)$.[18]

5.1 Continuous-time models

We start by considering the celebrated Vasicek model of the short-term interest rate,

$$dr(\tau) = (b_1 - b_2 r(\tau))\, d\tau + a_1 \times dW(\tau), \tag{13}$$

where b_1, b_2 and a_1 are parameters and W is a Brownian motion. This model is the continuous-time counterpart of a discrete-time AR(1) model. Given its simplicity, it is a natural starting point. Moreover, this model can also be easily estimated by maximum likelihood. Therefore, it is a useful benchmark. The parametrization we choose for this model is $b_1 = 3.00$, $b_2 = 0.50$ and $a_1 = 3.00$. These parameter values imply that the model-generated data have approximately the same mean, variance and autocorrelations as the US short-term interest rate.

We consider four estimators. The first estimator is the CD-SNE in (8) implemented with the weighting function in (9). As we explained in Section 3.3, this estimator matches the model conditional density to the conditional density $\pi_T(r_t|r_{t-1})$ estimated from sample data. As we also demonstrated in Section 3.3, the use of the weighting function $w_T(r_t, r_{t-1}) = \frac{\pi_T(r_{t-1})^2}{\pi_T(r_t, r_{t-1})}$ makes the resulting CD-SNE first order efficient in this case.

The second estimator is the SNE in (7) obtained by matching the joint density of any two adjacent observations $\pi_T(r_t, r_{t-1})$. We use $w_T(r_t, r_{t-1}) = \pi_T(r_t, r_{t-1})$ as a weighting function. According to our theory, the resulting estimator is not first-order efficient. This experiment will

[18] In the most demanding applications (diffusion processes and sample sizes of 1000 observations), computation time on a Pentium 4 with 1.7GHz is between 3 and 6 minutes. Computation time may vary according to the dimension of the parameter vector, the programming language, the optimization algorithm and sometimes, the spread of the uniform distribution we draw the initial guesses from (see footnotes 20-22 below).

In the Monte Carlo experiments of this section, our estimators are implemented with Fortran-90. The objective functions are optimized through a Quasi-Newton algorithm, with a convergence criterion of the order of 10^{-5}.

thus help us to understand the effects of suboptimal choice of the objective function on the finite sample properties of our estimators.

The third estimator, labelled Analytical-NE, is a modification of the SNE in which the simulated nonparametric estimate $S^{-1}\sum_{i=1}^{S}\pi_T^i(r_t, r_{t-1};\theta)$ is replaced with its analytical counterpart $\pi^{\text{vas}}(r_t, r_{t-1};\theta)$.[19] Precisely, the objective function of the Analytical-NE takes the form,

$$\int_{(r_t,r_{t-1})\in\mathbb{R}^2}\left[\pi^{\text{vas}}(r_t, r_{t-1};\theta)-\pi_T(r_t, r_{t-1})\right]^2\pi_T(r_t, r_{t-1})dr_t dr_{t-1}. \tag{14}$$

Naturally, the Analytical-NE is practically unfeasible in most models of interest. We consider this estimator because it provides us with useful information about the importance of the "twin-smoothing" procedure discussed in Section 3.1 - i.e. the importance to apply the same kernel smoothing procedure to sample data and model-related data.

The fourth and last estimator we consider is the MLE.

Table 2 provides results of our Monte Carlo experiments when model (13) is estimated through the previous methods. We report mean, median, and sample standard deviation of the estimates over the Monte Carlo replications.[20] As regards the CD-SNE and the SNE, Table 2 also reports: 1) asymptotic standard deviations (obtained through the relevant theory developed in Section 3); and 2) coverage rates for 90% confidence intervals computed through the usual asymptotic approximation to the distribution of the estimator - that is, the estimate plus or minus 1.645 times the asymptotic standard deviation.

When the size of the simulated samples is 1000, the performance of the CD-SNE and MLE are comparable in terms of variability of the estimates. Specifically, the CD-SNE has a lower standard deviation than the MLE as regards the estimation of the parameter b_2 tuning the persistence of r; and the MLE is more precise than the CD-SNE as regards the estimation of the diffusion parameter a_1. As it turns out, the sample standard deviation of the CD-SNE estimates of a_1 is larger than its asymptotic counterpart, and this is reflected in a coverage rate below the nominal one. As regards biases, the MLE tends to under-estimate the dependence of the data and largely over-estimate the constant b_1 in the drift term. Interestingly, this phenomenon does not emerge when the model is estimated with the CD-SNE.

As expected, the results in Table 2 clearly demonstrate that moving from CD-SNE to SNE causes an increase in the variability of the estimates; this result is pronounced for the diffusion parameter a_1. Furthermore, the Analytical-NE produces a much larger variability of the estimates; even more interestingly, it estimates the parameters with large biases: in particular, minimizing

[19] As is well known, the transition density $\pi^{\text{vas}}(r_s|r_t;\theta)$ from date t to date s ($s>t$) is Gaussian with expectation equal to $b_1/b_2+[r(t)-(b_1/b_2)]\exp(-b_2(s-t))$ and variance equal to $\left[a_1^2/(2b_2)\right][1-\exp(-2b_2(s-t))]$. The marginal density is obtained by letting $s\to\infty$.

[20] Initial values of the parameters are drawn from a uniform distribution on $[1.5, 4.5]$ (for b_1 and a_1); and on $[0.1, 0.9]$ (for b_2). The correlations (over the Monte Carlo replications) between initial values and final estimates are 0.07 (for the SNE) and 0.08 (for the CD-SNE) on average over the parameters.

(14) over-estimates the diffusion coefficient a_1 by 0.55 and the constant b_1 in the drift term by 0.47. These results are perfectly consistent with our theoretical explanation of a second order biases arising when the model density and the sample density are not smoothed with the same kernel. As is well-known, the practical performance of nonparametric methods hinges on the proper choice of the bandwidth parameter. Table 2 also shows the effects of bandwidth selection on the small samples performance on the CD-SNE. We have implemented two experiments: in the first one, estimation is performed with a bandwidth level which is double the size suggested by Chen, Linton and Robinson (2001) - which we utilized earlier; in the second experiment, the bandwidth size is half the one we utilized earlier. The results in Table 2 suggest that while these bandwidth choices produce some effects on the estimates, those effects are marginal. In particular, under-smoothing the data introduces some volatility in the density estimates - which is reflected in a higher standard deviation of the parameters estimates. And over-smoothing the data tends to increase the mean bias of the parameter estimates.

Finally, Table 2 also documents the performance of the CD-SNE, SNE and MLE in shorter samples of 500 observations. As expected, the variability of the estimates increases with all these methods. As regards the estimates of the b_1 and b_2 parameters, the mean bias of the MLE almost doubles with respect to the longer sample; and the mean biases of the CD-SNE remain small relatively to the corresponding MLE mean biases.

A simple extension of model (13) is one in which the instantaneous volatility of the short-term rate r is proportional to an unobservable process $\{\sigma(\tau)\}_{t\geq 0}$ with constant elasticity of variance,

$$\begin{cases} dr(\tau) & = & (b_1 - b_2 r(\tau))\, d\tau + a_1 \times \sigma(\tau) dW_1(\tau) \\ d\sigma(\tau) & = & b_3 \times (1 - \sigma(\tau))\, d\tau + a_2 \times \sigma(\tau) dW_2(\tau) \end{cases} \tag{15}$$

where W_1 and W_2 are two uncorrelated Brownian motions, and b_3 and a_2 are parameters related to the volatility dynamics. Naturally, the presence of the unobservable volatility component in model (15) now makes MLE an unfeasible estimation alternative.

The parametrization of the stochastic volatility model (15) is $b_1 = 3.00$, $b_2 = 0.5$, $a_1 = 3.00$, $b_3 = 1.0$ and $a_2 = 0.3$. This parametrization implies that the unobservable volatility process is strongly dependent, but not as strongly as the observable process r itself. The parameters' values we are using are consistent with estimates of similar models on US short-term interest rates data.

We consider two estimators. The first estimator is the CD-SNE matching the model's conditional density to the conditional density $\pi_T(r_t|r_{t-1})$ of any two adjacent observations; we implement the CD-SNE with the weighting function in (9) of Section 3.3. The second estimator is the SNE implemented by matching the joint density $\pi_T(r_t, r_{t-1})$ of two adjacent observations; we use $\pi_T(r_t, r_{t-1})$ as a weighting function. The performance of both estimators is gauged in samples of 1000 and 500 observations, and the results are reported in Table 3.[21]

[21] Initial values of the parameters are drawn from a uniform distribution on $[1.5, 4.5]$ (for b_1 and a_1); on $[0.1, 0.9]$

As regards the larger simple size case and the CD-SNE, the standard deviation and the bias associated with the parameters b_1 and b_2 of the observable process are of the same order of magnitude as in Table 2; the presence of the unobservable volatility component makes the estimate of a_1 become more imprecise than the corresponding estimates in Table 2. As regards the bias terms, there is a tendency to over-estimate the parameter b_3; this phenomenon becomes more pronounced in the smaller sample size.

In contrast with our previous results obtained with the Vasicek model (13), we do not observe a clear ranking between the properties of the CD-SNE and the SNE. This phenomenon is particularly clear when the two estimators' properties are compared in terms of the variance of the estimates. Intuitively, the unobservable volatility process $\{\sigma(\tau)\}$ destroys the Markovianity property of the short-term interest rate $\{r(\tau)\}$. Precisely, the joint process $\{r(\tau), \sigma(\tau)\}$ in (15) is clearly Markov, but the "marginal" process $\{r(\tau)\}$ is not. Therefore, the conditions in Theorem 3 for asymptotic efficiency of the CD-SNE are not met. As a result, there is no reason for the CD-SNE to outperform the SNE. This makes the SNE an interesting alternative to look at in practical applications such as the ones considered in this section. The Monte Carlo experiments for discrete-time models reported below do reinforce this conclusion.

5.2 Discrete-time models

Discrete-time stochastic volatility models are also very often utilized in financial applications. The first model we consider in this section is the following one,

$$\begin{cases} y_t &= \sigma_b \times \exp(y_t^*/2) \times \epsilon_{1t} \\ y_t^* &= \phi \times y_{t-1}^* + \sigma_e \times \epsilon_{2t} \end{cases} \tag{16}$$

where $\{y_t\}_{t=1,2,\cdots}$ is the observable variable; $\{y_t^*\}_{t=1,2,\cdots}$ is the (latent) volatility process; ϵ_{1t} and ϵ_{2t} are two standard normal i.d. innovations; and ϕ, σ_b and σ_e are the parameters of interest. Our economic interpretation of the observable variable y_t is one of the unpredictable part of some long-lived asset return. One important reason leading us to focus on model (16) is that this model has become a workhorse in previous Monte Carlo studies - for example, Fermanian and Salanié (2004) tested their NPSML estimator on this model.

The parametrization of the discrete-time model (16) is $\phi = 0.95$, $\sigma_b = 0.025$ and $\sigma_e = 0.260$. We consider sample sizes of 500 observations. Table 4 reports the results of our Monte Carlo experiments when model (16) is estimated through the CD-SNE and the SNE. As in our previous Monte Carlo experiments on continuous-time models, we implement the CD-SNE by matching the model's conditional density to the conditional density $\pi_T(y_t|y_{t-1})$ of two adjacent observations, and utilize the weighting function (9) $\frac{\pi_T(y_{t-1})^2}{\pi_T(y_t, y_{t-1})}$ of Section 3.3. Similarly, we implement the

(for b_2); on $[0.5, 1.5]$ (for b_3); and on $[0.1, 0.5]$ (for a_2). The correlation (over the Monte Carlo replications) between initial values and final estimates are 0.12 (for the SNE) and 0.11 (for the CD-SNE) on average over the parameters.

SNE by matching the model's joint density to the joint density $\pi_T(y_t, y_{t-1})$ of two adjacent observations, and use $\pi_T(y_t, y_{t-1})$ as a weighting function.[22] Table 4 also reports the finite sample properties of three alternative estimation methods available in the literature, and summarized by Fermanian and Salanié (2004) (see *their* Table 4).

The results in Table 4 reveal that the finite sample properties of the CD-SNE and the SNE are very satisfactory, also in comparison with alternative estimation methods. In particular, the sample variability of the estimates of ϕ and σ_b obtained with our methods is in line with the asymptotic counterpart. As it turns out, it is relatively more difficult to estimate the volatility parameter σ_e of the latent process $\{y_t^*\}$; this results in a sample standard deviation larger than its asymptotic counterpart for both the CD-SNE and the SNE.

In our last Monte Carlo experiment, we explore how our methods are affected by the dimensionality of nonparametric density estimates. We consider a simple model in which *two* (unpredictable parts of) asset returns exhibit stochastic volatility. We make the simplifying assumption that the two asset returns volatilities are driven by a common volatility factor,

$$\begin{cases} y_{1t} &= \sigma_{b1} \times \exp(y_t^*/2) \times \epsilon_{1t} \\ y_{2t} &= \sigma_{b2} \times \exp\left(y_t^*/2\right) \times \epsilon_{2t} \\ y_t^* &= \phi \times y_{t-1}^* + \sigma_e \times \epsilon_{3t} \end{cases} \tag{17}$$

where $\{y_{it}\}_{t=1,2,\cdots}$ $(i = 1, 2)$ are the observable variables; $\{y_t^*\}_{t=1,2,\cdots}$ is the (latent) volatility process; ϵ_{1t}, ϵ_{2t} and ϵ_{3t} are three standard normal i.d. innovations; and σ_{bi} $(i = 1, 2)$, ϕ and σ_e are the parameters of interest.

The presence of a common source of stochastic volatility in asset returns can be rationalized by many recent theoretical models of long-lived asset price fluctuations. For example, models with external habit formation predict that a common volatility factor arises because all assets in the economy are consistently priced by a single pricing kernel. Therefore, time-varying volatility in the pricing kernel induced by habit formation propagates to all the asset returns (see, e.g., Menzly, Santos and Veronesi (2004)). Naturally, a sensible model for applied work is one in which returns volatilities also feature idiosyncratic components. But here we simply aim at isolating the effects of the curse of dimensionality on our estimators finite sample performance and, for obvious computational reasons, the Monte Carlo design has to be as simple as possible.

Similarly as for the previous experiments, we consider sample sizes of 500 observations, and parametrize model (17) as follows: $\phi = 0.95$, $\sigma_{b1} = \sigma_{b2} = 0.025$ and $\sigma_e = 0.260$. We examine finite sample properties of both the CD-SNE and the SNE. The CD-SNE is implemented by matching the conditional density $\pi_T\left(y_{1t}, y_{2t} | y_{1t-1}, y_{2t-1}\right) = \frac{\pi_T(y_{1t}, y_{2t}, y_{1t-1}, y_{2t-1})}{\pi_T(y_{1t-1}, y_{2t-1})}$ of two adjacent pairs of observations - with the weighting function (9). The SNE is implemented by matching

[22] Initial values of the the parameters are drawn from a uniform distribution on $[0.15, 0.35]$ (for σ_e); on $[0.9, 0.99]$ (for ϕ); and on $[0.015, 0.035]$ (for σ_b). The correlation (over the Monte Carlo replications) between initial values and final estimates are 0.09 (for the SNE) and 0.11 (for the CD-SNE) on average over the parameters.

the joint density $\pi_T(y_{1t}, y_{2t}, y_{1t-1}, y_{2t-1})$ of two adjacent pairs of observations - with weighting function $\pi_T(y_{1t}, y_{2t}, y_{1t-1}, y_{2t-1})$.[23] The results are displayed in Table 5.

The increase in dimensionality may produce two effects on the estimates. On the one hand, the observation of two asset returns may facilitate our understanding of the dynamic properties of the common unobserved volatility process. On the other hand, the larger dimension of the nonparametric density estimates may impinge upon the precision of the estimates. The results in Table 5 suggest that these effects do arise in our experiments. Overall, an increase in dimensionality does not seem to have jeopardized the performance of our estimators in this experiment.

6 Conclusions

This paper has introduced new methods to estimate the parameters of partially observed dynamic models. The building block of these methods is indeed very simple. It consists in simulating the model of interest for the purpose of recovering the corresponding density function. Our estimators are the ones which make densities on simulated data as close as possible to their empirical counterparts. We made use of classical ideas in the statistical literature to build up convenient measures of closeness of densities. Our estimators are easy to implement, fast to compute and in the special case of fully observed Markov systems, they can attain the same asymptotic efficiency as the maximum likelihood estimator. Furthermore, Monte Carlo experiments revealed that their finite sample performance is very satisfactory, even in comparison to maximum likelihood.

Using simulations to recover model-implied density is not only convenient "just" because it allows one to recover estimates of densities unknown in closed-form. We demonstrated that our "twin-smoothing" procedure makes this feature of our methods stands as a great improvement upon alternative techniques matching "closed-form" model-implied densities to data-implied densities. Consistently with our asymptotic theory, finite sample results suggest that a careful choice of both the measures of closeness of density functions and the bandwidth functions does enhance the performance of our estimators, but mainly in terms of their precision. Furthermore, our trick to use simulations to recover model-implied densities makes our estimators attain a high degree of accuracy in terms of unbiasedness, even in cases of unsophisticated objective functions and/or bandwidth selection procedures.

In our numerical experiments, we emphasized applications related to some typical models arising in financial economics. But we also demonstrated that our approach is quite general, and can be used to address related estimation problems. As an example, the typical Markov models arising in applied macroeconomics may also be estimated with our methods. In these cases, too, the previous asymptotic efficiency and encouraging finite sample properties make our methods stand as a promising advance into the literature of simulation-based inference methods.

[23]Initial values of the parameters are drawn as in the previous footnote. Correlations between initial guesses and final estimates are also of the same order of magnitude as in the previous footnote.

Appendix

An extensive appendix including the proofs of lemmata 1-10 in appendix A and further computations in appendixes B, C and D can be found in Altissimo and Mele (2005) (hereafter Al-M05).

A. Lemmata

Lemma 1. *Let assumptions 1-(a), 2 and K hold and for each t, let* $x_t \equiv (z_t, v_t)$*, as in the main text. We have,*

(a) $\sup_{x\in\mathbb{R}^q} |\pi_T(x) - m_0(x)| = O_p\left(T^{-\frac{1}{2}}\lambda_T^{-q}\right)$.

(b) $\sup_{x\in\mathbb{R}^q} |\pi_T(x) - \pi_0(x)| = O_p\left(T^{-\frac{1}{2}}\lambda_T^{-q}\right) + O_p(\lambda_T^r)$.

Lemma 2. *Let assumptions 1-(a), 2 and K hold, and set* $A_T \equiv \{(z,v) \in Z \times V : \pi_T(z,v) > \alpha_T\}$*, where* $\lim_{T\to\infty} \alpha_T \to 0$, $\lim_{T\to\infty} T^{\frac{1}{2}}\lambda_T^q \alpha_T^3 \to \infty$ *and* $\lim_{T\to\infty} \lambda_T^{q^*}\alpha_T \to 0$. *We have:*

(a) *Let* $\lim_{T\to\infty} \lambda_T \geq 0$; *then,*

$$\sup_{(z,v)\in A_T}\left[\frac{1}{m_0(z,v)}\left|\frac{\pi_T(v)}{\pi_T(z|v)} - \frac{m_0(v)}{n_0(z,v)}\right|\right] \xrightarrow{p} 0,$$

where $m_0(\cdot) \equiv m(\cdot;\theta_0)$ *and* $n_0(\cdot) \equiv n(\cdot;\theta_0)$.

(b) *Let* $\lim_{T\to\infty} \lambda_T = 0$, *and* $\lim_{T\to\infty} \alpha_T^3 \lambda_T^{-r} = \infty$; *then,*

$$\sup_{(z,v)\in A_T}\left[\frac{1}{\pi_0(z,v)}\left|\frac{\pi_T(v)}{\pi_T(z|v)} - \frac{\pi_0(v)}{\pi_0(z|v)}\right|\right] \xrightarrow{p} 0.$$

Lemma 3. *Let assumptions 1-(a), 2 and K hold, and set* $B_T \equiv \{v \in V : \pi_T^i(v;\theta) > \delta_T$, $i = 0,1,\cdots,S$, *all* $\theta \in \Theta\}$ $(\pi_T^0 \equiv \pi_T)$, *where* $\lim_{T\to\infty} \delta_T \to 0$, $\lim_{T\to\infty} T^{\frac{1}{2}}\lambda_T^{q-q^*}\delta_T^2 \to \infty$ *and* $\lim_{T\to\infty} T^{\frac{1}{2}}\lambda_T^q \delta_T \to \infty$. *We have:*

(a) *Let* $\lim_{T\to\infty} \lambda_T \geq 0$; *then* $\sup_{(z,v)\in Z\times B_T}\left|\frac{\pi_T(z,v)}{\pi_T(v)} - n_0(z,v)\right| \xrightarrow{p} 0$.

(b) *Let* $\lim_{T\to\infty} \lambda_T = 0$, *and* $\lim_{T\to\infty} \delta_T^2 \lambda_T^{-r} = \infty$; *then* $\sup_{(z,v)\in B_T}\left|\frac{\pi_T(z,v)}{\pi_T(v)} - \pi(z|v)\right| \xrightarrow{p} 0$.

Lemma 4. *Let assumptions 1-(a), 2 and K hold, and let* $\lim_{T\to\infty} \alpha_T \to 0$, $\lim_{T\to\infty} \delta_T \to 0$, $\lim_{T\to\infty} T^{\frac{1}{2}}\lambda_T^q \alpha_T^2 \delta_T \to \infty$ *and* $\lim_{T\to\infty} T^{\frac{1}{2}}\lambda_T^{q-q^*}\alpha_T^2\delta_T^2 \to \infty$. *We have:*

(a) *Let* $\lim_{T\to\infty}\lambda_T \geq 0$; *then, for each* $i = 1, \cdots, S$, *and* $\theta \in \Theta$,

$$\sup_{(z,v)\in A_T\cap B_T}\left[\frac{1}{m_0(z,v)\,n_0(z,v)}\left|\frac{\pi_T^i(z,v;\theta)}{\pi_T^i(v;\theta)} - n(z,v;\theta)\right|\right] \xrightarrow{p} 0.$$

(b) *Let* $\lim_{T\to\infty}\lambda_T = 0$, $\lim_{T\to\infty}\alpha_T^2\delta_T\lambda_T^{-r} = \infty$ *and* $\lim_{T\to\infty}\alpha_T^2\delta_T^2\lambda_T^{-r} = \infty$; *then, for each* $i = 1, \cdots, S$, *and* $\theta \in \Theta$,

$$\sup_{(z,v)\in A_T\cap B_T}\left[\frac{1}{m_0(z,v)\,n_0(z,v)}\left|\frac{\pi_T^i(z,v;\theta)}{\pi_T^i(v;\theta)} - \pi(z|\,v;\theta)\right|\right] \xrightarrow{p} 0.$$

Lemma 5. *Let assumptions 1, 2, K hold. For each* t, *let* $x_t \equiv (z_t, v_t)$, *as in the main text, and let* $\mathcal{K}_T^j(z,v;\theta)$ *satisfy the mixing condition in assumption 2 (*$j = 1, \cdots, p_\theta$*). Finally let* $\partial^{\rho+1}\pi(x;\theta)/\partial\theta\partial x^\rho$ *be uniformly bounded for some* $\rho \geq r$. *Then, for all* $\theta \in \Theta$ *and* $i = 1, \cdots, S$,

$$\sup_{x\in\mathbb{R}^q}\left|\nabla_{\theta_j}\tilde{\pi}_T(x;\theta) - \nabla_{\theta_j}\pi(x;\theta)\right| = O_p\left(T^{-\frac{1}{2}}\lambda_T^{-q-1}\right) + O_p\left(\lambda_T^r\right), \qquad j = 1, \cdots, p_\theta.$$

In lemmas 6 through 10 below, α_T and δ_T denote the same sequences introduced in the previous lemmas 2 and 3.

Lemma 6. *Let the assumptions in lemma 5 hold. Then, for all* $\theta \in \Theta$ *and* $j = 1, \cdots, p_\theta$

$$\begin{aligned}&\sup_{(z,v)\in Z\times B_T}\left|\nabla_{\theta_j}\tilde{\pi}_T(z|\,v;\theta) - \nabla_{\theta_j}\pi(z|\,v;\theta)\right|\\ =\;& O_p\left(T^{-\frac{1}{2}}\lambda_T^{-q-1}\delta_T^{-2}\right) + O_p\left(T^{-\frac{1}{2}}\lambda_T^{-(q-q^*)-1}\delta_T^{-2}\right) + O_p\left(T^{-\frac{1}{2}}\lambda_T^{-(q-q^*)}\delta_T^{-3}\right) + O_p\left(\lambda_T^r\delta_T^{-3}\right).\end{aligned}$$

Lemma 7. *Let the assumptions in lemma 5, and assumption 4-(b) hold. Then, for all* $\theta \in \Theta$ *and* $j = 1, \cdots, p_\theta$

$$\begin{aligned}&\sup_{(z,v)\in Z\times B_T}\left|\frac{\nabla_{\theta_j}\tilde{\pi}_T(z|\,v;\theta_0)w_T(z,v)}{\pi_T(v)} - \frac{\nabla_{\theta_j}\pi(z|\,v;\theta_0)w(z,v)}{\pi(v;\theta_0)}\right|\\ =\;& O_p\left(T^{-\frac{1}{2}}\lambda_T^{-q-1}\delta_T^{-3}\right) + O_p\left(T^{-\frac{1}{2}}\lambda_T^{-(q-q^*)-1}\delta_T^{-3}\right) + O_p\left(T^{-\frac{1}{2}}\lambda_T^{-(q-q^*)}\delta_T^{-4}\right) + O_p\left(\lambda_T^r\delta_T^{-4}\right).\end{aligned}$$

Lemma 8. *Let the assumptions in lemma 7 hold. Then, for all* $\theta \in \Theta$ *and* $j = 1, \cdots, p_\theta$

$$\begin{aligned}&\sup_{(z,v)\in Z\times B_T}\left|\frac{\nabla_{\theta_j}\tilde{\pi}_T(z|\,v;\theta_0)E\left[\pi_T(z,v)\right]w_T(z,v)}{\pi_T^1(v;\theta_0)\cdot\pi_T(v)} - \frac{\nabla_{\theta_j}\pi(z|\,v;\theta_0)\pi_0(z,v)\,w(z,v)}{\pi(v;\theta_0)^2}\right|\\ =\;& O_p\left(T^{-\frac{1}{2}}\lambda_T^{-q-1}\delta_T^{-4}\right) + O_p\left(T^{-\frac{1}{2}}\lambda_T^{-(q-q^*)-1}\delta_T^{-4}\right) + O_p\left(T^{-\frac{1}{2}}\lambda_T^{-(q-q^*)}\delta_T^{-5}\right) + O_p\left(\lambda_T^r\delta_T^{-5}\right).\end{aligned}$$

Lemma 9. *Let the assumptions in lemma 5 hold. Let* $v \mapsto \xi_{1T}(v)$ $(v \in V \subseteq \mathbb{R}^{q-q^*})$ *be a sequence of real, bounded functions satisfying* $\sup_{v\in V} |\xi_{1T}(v) - \xi_1(v)| = O_p(T^{-\frac{1}{2}}\lambda_T^{-(q-q^*)}) + O_p(\lambda_T^r)$*, for some function* ξ_1*. Then, for all* $\theta \in \Theta$ *and* $j = 1, \cdots, p_\theta$*,*

$$\sup_{(z,v)\in A_T\times B_T} \left| \frac{\nabla_{\theta_j}\tilde{\pi}_T(z|\, v;\theta_0)\pi_T(v)\,\xi_{1T}(v)}{\pi_T(z,v)} - \frac{\nabla_{\theta_j}\pi(z|\, v;\theta_0)\pi_0(v)\,\xi_1(v)}{\pi_0(z,v)} \right|$$
$$= O_p\left(T^{-\frac{1}{2}}\lambda_T^{-q-1}\alpha_T^{-1}\delta_T^{-2}\right) + O_p\left(T^{-\frac{1}{2}}\lambda_T^{-(q-q^*)-1}\alpha_T^{-1}\delta_T^{-2}\right) + O_p\left(T^{-\frac{1}{2}}\lambda_T^{-(q-q^*)}\alpha_T^{-1}\delta_T^{-3}\right)$$
$$+ O_p\left(\lambda_T^r\alpha_T^{-1}\delta_T^{-3}\right) + O_p\left(T^{-\frac{1}{2}}\lambda_T^{-(q-q^*)}\alpha_T^{-1}\right) + O_p\left(T^{-\frac{1}{2}}\lambda_T^{-q}\alpha_T^{-2}\right) + O_p\left(\lambda_T^r\alpha_T^{-2}\right).$$

Lemma 10. *Let the assumptions in lemma 5 hold, and let* $\xi_{1T}(v)$ *be the sequence of functions in lemma 9. Then, for all* $\theta \in \Theta$ *and* $j = 1, \cdots, p_\theta$*,*

$$\sup_{(z,v)\in A_T\times B_T} \left| \frac{\nabla_{\theta_j}\tilde{\pi}_T(z|\, v;\theta_0)E\left[\pi_T(z,v)\right]\xi_{1T}(v)\,\pi_T(v)}{\pi_T^1(v;\theta_0)\,\pi_T(z,v)} - \nabla_{\theta_j}\pi(z|\, v;\theta_0)\xi_1(v) \right|$$
$$= O_p\left(T^{-\frac{1}{2}}\lambda_T^{-q-1}\alpha_T^{-1}\delta_T^{-3}\right) + O_p\left(T^{-\frac{1}{2}}\lambda_T^{-q}\alpha_T^{-2}\delta_T^{-1}\right) + O_p\left(T^{-\frac{1}{2}}\lambda_T^{-(q-q^*)}\alpha_T^{-1}\delta_T^{-4}\right)$$
$$+ O_p\left(\lambda_T^r\alpha_T^{-1}\delta_T^{-4}\right) + O_p\left(\lambda_T^r\alpha_T^{-2}\delta_T^{-1}\right) + O_p\left(\lambda_T^r\delta_T^{-1}\alpha_T^{-1}\right).$$

B. Proof of theorem 1

B.1 Consistency

Proposition 1. *Let assumptions 1, 2 and 3-(a) hold. Then* $\forall\theta \in \Theta$, $L_T(\theta) \xrightarrow{p} L(\theta)$ *as* $T \to \infty$.

According to a well-known result (see Newey (1991, thm. 2.1 p. 1162)), the following conditions are equivalent:

C1: $\lim_{T\to\infty} P\left(\sup_{\theta\in\Theta} |L_T(\theta) - L(\theta)| > \epsilon\right) = 0$.

C2: $\forall\theta \in \Theta$, $L_T(\theta) \xrightarrow{p} L(\theta)$, *and* $L_T(\theta)$ *is stochastically equicontinuous.*

By Newey and McFadden (1994, lemma 2.9 p. 2138), assumption 3-(b) guarantees that $L_T(\theta)$ is stochastically equicontinuous, and so weak consistency follows from the equivalence of C1 and C2 above, assumption 3-(a,b), compactness of Θ, and a classical argument (e.g., White (1994, theorem 3.4)). So we are only left to prove proposition 1.

Proof of proposition 1. We have:

$$|L_T(\theta) - L(\theta)| \leq \int |g_T(x;\theta)|\, dx,$$

where

$$\begin{aligned}
g_T(x;\theta) &\equiv \sigma_{1T}(x;\theta) + [|\tilde{\pi}_T(x;\theta) - \pi_T(x)| - |m(x;\theta) - m(x;\theta_0)|] \cdot [\rho_T(x;\theta) + \rho(x;\theta)] \\
&\leq \sigma_{1T}(x;\theta) + |[\tilde{\pi}_T(x;\theta) - m(x;\theta)] - [\pi_T(x) - m(x;\theta_0)]| \cdot [\rho_T(x;\theta) + \rho(x;\theta)] \\
&\equiv \sigma_{1T}(x;\theta) + \sigma_{2T}(x;\theta) \\
\sigma_{1T}(x;\theta) &\equiv |\tilde{\pi}_T(x;\theta) - \pi_T(x)| \cdot |m(x;\theta) - m(x;\theta_0)| \cdot |w_T(x) - w(x)| \\
\rho_T(x;\theta) &\equiv |\tilde{\pi}_T(x;\theta) - \pi_T(x)| \cdot w_T(x) \\
\rho(x;\theta) &\equiv |m(x;\theta) - m(x;\theta_0)| \cdot w(x)
\end{aligned}$$

We claim that for all $\theta \in \Theta$, $\int \sigma_{1T} \xrightarrow{p} 0$. Indeed, for fixed θ, σ_{1T} is clearly bounded by integrable functions independent of T. As $T \to \infty$, $\sigma_{1T}(x;\theta) \xrightarrow{p} 0$, x-pointwise. By dominated convergence, $\lim_{T\to\infty} E[\sigma_{1T}(x;\theta)] = E[\lim_{T\to\infty} \sigma_{1T}(x;\theta)] = 0$ all $(x,\theta) \in X \times \Theta$. By Fubini, $E\left[\int \sigma_{1T}(x;\theta)\,dx\right] = \int E[\sigma_{1T}(x;\theta)]\,dx$ all $\theta \in \Theta$. Again by dominated convergence,

$$\lim_{T\to\infty} E\left[\int \sigma_{1T}(x;\theta)\,dx\right] = \lim_{T\to\infty} \int E[\sigma_{1T}(x;\theta)]\,dx = \int \lim_{T\to\infty} E[\sigma_{1T}(x;\theta)]\,dx = 0,\ \forall \theta \in \Theta.$$

By Markov's inequality:

$$\forall \epsilon > 0,\ \ P\left\{\int \sigma_{1T}(x;\theta)\,dx > \epsilon\right\} \leq \frac{E\left[\int \sigma_{1T}(x;\theta)\,dx\right]}{\epsilon},\ \ \forall \theta \in \Theta.$$

Hence, for all $\theta \in \Theta$, $\int \sigma_{1T} \xrightarrow{p} 0$. The proof for the σ_{2T} term is similar. The additional argument is the observation that for all $x,\theta \in X \times \Theta$, $\max[m(x;\theta) - \tilde{\pi}_T(x;\theta), 0] \leq m(x;\theta)$, which is clearly integrable, and so by $\tilde{\pi}_T(x;\theta) \xrightarrow{p} m(x;\theta)$, x-pointwise, and dominated convergence,

$$\text{for all } \theta \in \Theta, \quad \int |m(x;\theta) - \tilde{\pi}_T(x;\theta)|\,dx = 2\int \max[m(x;\theta) - \tilde{\pi}_T(x;\theta), 0]\,dx \xrightarrow{p} 0.$$

Hence by arguments nearly identical to the ones leading to $\int \sigma_{1T} \xrightarrow{p} 0$, we also have that for all $\theta \in \Theta$, $\int \sigma_{2T} \xrightarrow{p} 0$, and the proof is complete. The case $\lambda \equiv \lambda_T \downarrow 0$ is identical. ■

B.2 Asymptotic normality

Let $\mathbf{0}_n$ denote a column vector of n zeros. By assumption 4-(a), the order of derivation and integration in $\nabla_\theta L_T(\theta)$ may be interchanged (see Newey and McFadden (1994, lemma 3.6 p. 2152-2153)), and the first order conditions satisfied by the SNE are,

$$\mathbf{0}_{p_\theta} = \int [\tilde{\pi}_T(x;\theta_{T,S}) - \pi_T(x)]\,\nabla_\theta \tilde{\pi}_T(x;\theta_{T,S})\,w_T(x)dx.$$

Let $\theta(c) \equiv c \circ (\theta_0 - \theta_{T,S}) + \theta_{T,S}$, where, for any $c \in (0,1)^{p_\theta}$ and $\theta \in \Theta$, $c \circ \theta$ denotes the vector in Θ whose i-th element is $c^{(i)}\theta^{(i)}$. By assumption 4-(a), there exists a c^* in $(0,1)^{p_\theta}$ such that:

$$\begin{aligned}\mathbf{0}_{p_\theta} &= \sqrt{T}\int \left[\tilde{\pi}_T(x;\theta_0) - \pi_T(x)\right] \nabla_\theta \tilde{\pi}_T\left(x;\theta_0\right) w_T(x)dx \\ &+ \left[\int \left|\nabla_\theta \tilde{\pi}_T(x;\bar{\theta})\right|_2 w_T(x)dx + \left(\bar{\theta} - \theta_0\right) \cdot k_{1T}\left(\bar{\theta}\right) + k_{2T}\left(\bar{\theta}\right)\right] \cdot \sqrt{T}(\theta_{T,S} - \theta_0), \qquad \text{(B1)}\end{aligned}$$

where $\bar{\theta} \equiv \theta(c^*)$, $|b|_2$ denotes the outer product $b \cdot b^\top$ of a column vector b, and for some θ^*,

$$\begin{aligned}\left|k_{1T}\left(\bar{\theta}\right)\right| &\leq \int \left|\nabla_\theta \tilde{\pi}_T(x;\theta^*)\right| \left|\nabla_{\theta\theta} \tilde{\pi}_T\left(x;\bar{\theta}\right)\right| w_T(x)dx \\ \left|k_{2T}\left(\bar{\theta}\right)\right| &\leq \int \left|\tilde{\pi}_T(x;\theta_0) - \pi_T(x)\right| \left|\nabla_{\theta\theta} \tilde{\pi}_T\left(x;\bar{\theta}\right)\right| w_T(x)dx\end{aligned}$$

By assumption 4-(a), the term $\nabla_{\theta\theta}\tilde{\pi}_T\left(x;\bar{\theta}\right)$ is bounded in probability as T becomes large. Hence a) so is $\left|k_{1T}\left(\bar{\theta}\right)\right|$; and b) by lemma 1, $\left|k_{2T}\left(\bar{\theta}\right)\right| \stackrel{p}{\to} \mathbf{0}_{p_\theta \times p_\theta}$. Moreover,

$$\int \left|\nabla_\theta \tilde{\pi}_T(x;\bar{\theta})\right|_2 w_T(x)dx = \int \left|\nabla_\theta \tilde{\pi}_T(x;\theta_0)\right|_2 w_T(x)dx + R_T\left(\bar{\theta}\right),$$

where

$$\left|R_T\left(\bar{\theta}\right)\right|_{i,j} \leq \int \left|\left|\nabla_\theta \tilde{\pi}_T(x;\bar{\theta})\right|_2 - \left|\nabla_\theta \tilde{\pi}_T(x;\theta_0)\right|_2\right|_{i,j} w_T(x)dx.$$

Since $\int |w_T - w| \stackrel{p}{\to} 0$ and $\bar{\theta} \stackrel{p}{\to} \theta_0$, then by lemma 5, $\left|R_T\left(\bar{\theta}\right)\right|_{i,j} \stackrel{p}{\to} 0$ for all i,j. Hence,

$$\int \left|\nabla_\theta \tilde{\pi}_T(x;\bar{\theta})\right|_2 w_T(x)dx + \left(\bar{\theta} - \theta_0\right) k_{1T}\left(\bar{\theta}\right) + k_{2T}\left(\bar{\theta}\right) \stackrel{p}{\to} \int \left|\nabla_\theta \pi(x;\theta_0)\right|_2 w(x)dx. \qquad \text{(B2)}$$

Next, consider the first term in (B1). For all $x \in X$ and fixed T, $E\left[\pi_T^i(x;\theta_0)\right] = E\left[\pi_T(x)\right]$ $(i = 1, \cdots, S)$. Hence,

$$\begin{aligned}&\sqrt{T}\int \left[\tilde{\pi}_T(x;\theta_0) - \pi_T(x)\right] \nabla_\theta \tilde{\pi}_T(x;\theta_0) w_T(x)dx \\ &= \int \sqrt{T}\left[\tilde{\pi}_T(x;\theta_0) - E(\tilde{\pi}_T(x;\theta_0))\right] \nabla_\theta \tilde{\pi}_T(x;\theta_0) w_T(x)dx \\ &- \int \sqrt{T}\left[\pi_T(x) - E(\pi_T(x))\right] \nabla_\theta \tilde{\pi}_T(x;\theta_0) w_T(x)dx. \qquad \text{(B3)}\end{aligned}$$

Let $\mathbb{G}$ be a measurable V-C subgraph class of uniformly bounded functions (see, e.g., Arcones and Yu (1994, definition 2.2 p. 51)). By Arcones and Yu (1994, corollary 2.1 p. 59-60), for each $G \in \mathbb{G}$, $T^{-1/2}\sum_{t=t_l}^{T}\left[G(x_t) - EG\right]$ converges in law to a Gaussian process under assumption 2. Now $\lambda_T^{-q} K\left((x_t - x)/\lambda_T\right) \in \mathbb{G}$. Let $F(x;\theta) = \int_0^x \pi(v;\theta)dv$, $F_T(x) = \int_0^x \pi_T(v)dv$ and $F(x) =$

$\int_0^x \pi_0(v)dv$. Under the theorem's assumptions,

$$A_T \equiv \sqrt{T}(F_T(x) - E(F_T(x))) \Rightarrow \omega^0 \left(F\left(x\right)\right),$$

where $\omega^0 \left(F\right)$ is a Generalized Brownian Bridge with covariance kernel,

$$\min\left(F\left(x\right), F\left(y\right)\right)\left[1 - F\left(y\right)\right] + \sum_{k=1}^{\infty} \left[F^k\left(x,y\right) + F^k\left(y,x\right) - 2F\left(x\right)F\left(y\right)\right],$$

and $F^k\left(x,y\right) \equiv P\left(x_0 \leq x, x_k \leq y\right)$. We have,

$$\begin{aligned} J_T &\equiv \sqrt{T} \int \left[\pi_T(x) - E\left(\pi_T(x)\right)\right] \nabla_\theta \tilde{\pi}_T(x;\theta_0) w_T(x) dx \\ &= \int \left[w_T(x) - w(x)\right]\left[\nabla_\theta \tilde{\pi}_T(x;\theta_0) - \nabla_\theta \pi(x;\theta_0)\right] dA_T\left(x\right) \\ &+ \int \left[\nabla_\theta \tilde{\pi}_T(x;\theta_0) - \nabla_\theta \pi(x;\theta_0)\right] w\left(x\right) dA_T\left(x\right) \\ &+ \int \left[w_T(x) - w(x)\right] \nabla_\theta \pi(x;\theta_0) dA_T\left(x\right) + \int \nabla_\theta \pi(x;\theta_0) w\left(x\right) dA_T(x) \\ &\equiv J_{1T} + J_{2T} + J_{3T} + J_{4T}. \end{aligned}$$

By the continuous mapping theorem,

$$J_{4T} \xrightarrow{d} J_4 \equiv \int \nabla_\theta \pi\left(x;\theta_0\right) w(x) d\omega^0\left(F\left(x\right)\right).$$

By w_T and w bounded, and lemma 5, $J_{iT} = [O_p(T^{-\frac{1}{2}}\lambda_T^{-q-1}) + O_p\left(\lambda_T^r\right)]\mathbf{1}_{p_\theta}$, $i = 1, 2$. By assumption 4-(b), $J_{3T} = [O_p(T^{-\frac{1}{2}}\lambda_T^{-q}) + O_p\left(\lambda_T^r\right)]\mathbf{1}_{p_\theta}$. By the theorem's conditions, therefore, $J_T \xrightarrow{d} N\left(0, V_J\right)$, $V_J \equiv var\left(J_4\right)$. By the same computations in Aït-Sahalia (1994) (proof of thm. 1 p. 21-22) and Aït-Sahalia (1996) (proof of eq. (12), p. 420-421),

$$\begin{aligned} V_J = var\left[\nabla_\theta \pi\left(x_1;\theta_0\right) w(x_1)\right] + \sum_{k=1}^{\infty} \{cov\left[\nabla_\theta \pi\left(x_1;\theta_0\right) w\left(x_1\right), \nabla_\theta \pi\left(x_{1+k};\theta_0\right) w\left(x_{1+k}\right)\right] \\ + cov\left[\nabla_\theta \pi\left(x_{1+k};\theta_0\right) w\left(x_{1+k}\right), \nabla_\theta \pi\left(x_1;\theta_0\right) w\left(x_1\right)\right]\}. \end{aligned} \quad \text{(B4)}$$

Finally, let $F_T^i(x;\theta) \equiv \int_0^x \pi_T^i(v;\theta)dv$, $i = 1, \cdots, S$. As for A_T, $A_T^i(x;\theta_0) \equiv \sqrt{T}[F_T^i(x;\theta_0) - E(F_T^i(x;\theta_0))] \Rightarrow \omega_i^0\left(F\left(x\right)\right)$, where $\omega_i^0\left(F\right)$ are independent Generalized Brownian Bridges. Hence,

$$\sqrt{T}\sum_{i=1}^{S}\left[F_T^i(x;\theta_0) - E(F_T^i(x;\theta_0))\right] \Rightarrow \sum_{i=1}^{S} \omega_i^0\left(F\left(x\right)\right).$$

Since $E(F_T^i(x;\theta_0)) = E(F_T^j(x;\theta_0))$ for all $i,j = 1,\cdots,S$, we have, similarly as for the J_T term,

$$\int [\sqrt{T}\,(\tilde{\pi}_T(x;\theta_0) - E(\tilde{\pi}_T(x;\theta_0)))]\nabla_\theta \tilde{\pi}_T(x;\theta_0) w_T(x)dx \xrightarrow{d} N\left(0, \frac{1}{S}V_J\right),$$

where V_J is as in (B4). Finally, A and A_T^i, $i = 1,\cdots,S$, are all independent. Therefore, by (B3),

$$\sqrt{T}\int \left[\tilde{\pi}_T(x;\theta_0) - \pi_T(x)\right] \nabla_\theta \tilde{\pi}_T(x;\theta_0) w_T(x)dx \xrightarrow{d} N\left(0, \left(1+\frac{1}{S}\right)V_J\right). \tag{B5}$$

Hence by (B1), (B2), (B5) and Slutzky's theorem, $\sqrt{T}\,(\theta_{T,S} - \theta_0) \xrightarrow{d} N\left(0, \left(1+\frac{1}{S}\right)V\right)$, where

$$V \equiv \left[\int |\nabla_\theta \pi(x;\theta_0)|_2\, w(x)dx\right]^{-1} \cdot V_J \cdot \left[\int |\nabla_\theta \pi(x;\theta_0)|_2\, w(x)dx\right]^{\top -1}.$$

C. Proof of theorem 2

The following assumption contains one set of regularity conditions mentioned in the statement of theorem 2:

Assumption T1. *We have,*

(a) $\delta_T \to 0$ *and* $T^{\frac{1}{2}}\delta_T^2 \to \infty$.

(b) *In addition to assumption T1-(a),* $\lambda \equiv \lambda_T \to 0$, $T^{\frac{1}{2}}\lambda_T^{q-q^*}\delta_T^5 \to \infty$, $T^{\frac{1}{2}}\lambda_T^{q+1}\delta_T^4 \to \infty$, *and* $\delta_T^5\lambda_T^{-r} \to \infty$.

C.1 Consistency

Similarly as for the SNE, the objective function of the CD-SNE $\bar{L}_T$ satisfies $\left|\bar{L}_T(\theta) - \bar{L}(\theta)\right| \leq \iint \left(s_{1T}(z,v;\theta) + s_{2T}(z,v;\theta)\right) dzdv$, where

$$\begin{aligned}
s_{1T}(z,v;\theta) &\equiv |\tilde{\pi}_T(z|\,v;\theta) - \pi_T(z|\,v)|\,\mathbb{T}_{T,\delta}(v;\theta)\cdot |n(z,v;\theta) - n(z,v;\theta_0)| \cdot |w_T(z,v) - w(z,v)|;\\
s_{2T}(z,v;\theta) &\equiv |[\tilde{\pi}_T(z|\,v;\theta)\mathbb{T}_{T,\delta}(v;\theta) - n(z,v;\theta)] - [\pi_T(z|\,v)\mathbb{T}_{T,\delta}(v;\theta) - n(z,v;\theta_0)]|\\
&\quad \times [r_T(z,v;\theta) + r(z,v;\theta)];\\
r_T(z,v;\theta) &\equiv |\tilde{\pi}_T(z|\,v;\theta) - \pi_T(z|\,v)|\,\mathbb{T}_{T,\delta}(v;\theta)\cdot w_T(z,v);\\
r(z,v;\theta) &\equiv |n(z,v;\theta) - n(z,v;\theta_0)| \cdot w(z,v).
\end{aligned}$$

We now show that $\iint (s_{1T} + s_{2T}) \xrightarrow{p} 0$ for all $\theta \in \Theta$. We study the two integrals separately.

- For all $(z,v,\theta)\in Z\times V\times\Theta$, $s_{1T}(z,v;\theta)\le \ell_T(z,v;\theta)\cdot r_{2T}(z,v;\theta)$, where

$$\begin{aligned}\ell_T(z,v;\theta) &\equiv |n(z,v;\theta)-n(z,v;\theta_0)|\cdot|w_T(z,v)-w(z,v)| \\ r_{2T}(z,v;\theta) &\equiv \frac{1}{S}\sum_{i=1}^{S}\left|\pi_T^i(z|\,v;\theta)-n(z,v;\theta)\right|\mathbb{T}_{T,\delta}(v;\theta)+|\pi_T(z|\,v)-n(z,v;\theta_0)|\,\mathbb{T}_{T,\delta}(v;\theta) \\ &\quad +|n(z,v;\theta)-n(z,v;\theta_0)|\,\mathbb{T}_{T,\delta}(v;\theta).\end{aligned}$$

For each $\theta\in\Theta$, function ℓ_T is bounded by integrable functions independent of T, and $\ell_T\overset{p}{\to}0$ (z,v)-pointwise. Moreover,

$$\sup_{(z,v)\in Z\times V}\left|\pi_T^i(z|\,v;\theta)-n(z,v;\theta)\right|\mathbb{T}_{T,\delta}(v;\theta)\overset{p}{\to}0,\quad i=1,\cdots,S,$$

as a consequence of lemma 3-(a), and the conditions in the theorem. This result clearly holds for the second term of r_{2T} as well. Finally, $|n(\cdot,\cdot;\theta)-n(\cdot,\cdot;\theta_0)|$ is bounded. Therefore, $\iint s_{1T}(z,v;\theta)\overset{p}{\to}0$ for all $\theta\in\Theta$.

- For all $(z,v,\theta)\in Z\times V\times\Theta$,

$$\begin{aligned}&s_{2T}(z,v;\theta)\\ &\le\left[\frac{1}{S}\sum_{i=1}^{S}\left|\pi_T^i(z|\,v;\theta)\mathbb{T}_{T,\delta}(v;\theta)-n(z,v;\theta)\right|+|\pi_T(z|\,v)\mathbb{T}_{T,\delta}(v;\theta)-n(z,v;\theta_0)|\right]r_{3T}(z,v;\theta),\end{aligned}\tag{C1}$$

where $r_{3T}(z,v;\theta)\equiv r(z,v;\theta)+r_T(z,v;\theta)\le r(z,v;\theta)+r_{2T}(z,v;\theta)\,w_T(z,v)$. For each $i=1,\cdots,S$, and $(z,v,\theta)\in Z\times V\times\Theta$,

$$\begin{aligned}&\left|\pi_T^i(z|\,v;\theta)\mathbb{T}_{T,\delta}(v;\theta)-n(z,v;\theta)\right|r_{3T}(z,v;\theta)\\ &\le n(z,v;\theta)\left[1-\mathbb{T}_{T,\delta}(v;\theta)\right]\cdot\left[r(z,v;\theta)+r_{2T}(z,v;\theta)\,w_T(z,v)\right]\\ &+\left|\pi_T^i(z|\,v;\theta)-n(z,v;\theta)\right|\mathbb{T}_{T,\delta}(v;\theta)\cdot\left[r(z,v;\theta)+r_{2T}(z,v;\theta)\,w_T(z,v)\right]\\ &\equiv s_{21T}(z,v;\theta)+s_{22T}(z,v;\theta),\end{aligned}$$

where the inequality holds by the triangle inequality. Since w_T, r and n are bounded, and w_T and r are also integrable, $\iint s_{22T}(z,v;\theta)\overset{p}{\to}0$ for all $\theta\in\Theta$ by lemma 3-(a). As for the s_{21T} term, clearly $|1-\mathbb{T}_{T,\delta}(v;\theta)|\le 1$. Moreover, $1-\mathbb{T}_{T,\delta}(v;\theta)\overset{p}{\to}P_\pi\{\pi_0(v_1)<\lim_{T\to\infty}\delta_T\}-\int_{v:\pi_0(v)\in(\lim_{T\to\infty}\delta_T,2\lim_{T\to\infty}\delta_T)}\mathbb{T}_{T,\delta}(v;\theta)\,P_\pi(dv)$, where P_π is the stationary measure of v. Hence, by the conditions in the theorem and again lemma 3-(a), $\iint s_{21T}(z,v;\theta)\overset{p}{\to}0$ for all $\theta\in\Theta$. By reiterating the previous arguments, one shows that the same result holds for the second term in (C1) and therefore, $\iint s_{2T}(z,v;\theta)\overset{p}{\to}0$ for all $\theta\in\Theta$.

The case $\lambda \equiv \lambda_T \downarrow 0$ is dealt with similarly through lemma 3-(b) instead of lemma 3-(a), and the proof of consistency is complete by the same arguments in appendix B.1.

C.2 Asymptotic normality

By assumption 4-(a), the CD-SNE satisfies the following first order conditions,

$$\begin{aligned}\mathbf{0}_{p_\theta} &= \frac{1}{S}\sum_{i=1}^{S}\iint \left[\frac{\pi_T^i(z,v;\theta_{T,S})}{\pi_T^i(v;\theta_{T,S})} - \frac{\pi_T(z,v)}{\pi_T(v)}\right]\nabla_\theta\tilde{\pi}_T(z|\,v;\theta_{T,S})\,w_T(z,v)\mathbb{T}_{T,\delta}^2(v;\theta_{T,S})\,dzdv \\ &\quad + \iint \left[\tilde{\pi}_T(z|\,v;\theta_{T,S}) - \pi_T(z|\,v)\right]^2 w_T(z,v)\mathbb{T}_{T,\delta}(v;\theta_{T,S})\,\nabla_\theta\mathbb{T}_{T,\delta}(v;\theta_{T,S})\,dzdv.\end{aligned}$$

In Al-M05, we demonstrate that the trimming effects are asymptotically negligible under the conditions in the theorem. More precisely, an expansion of the first order conditions satisfied by the CD-SNE around θ_0 leaves:

$$\begin{aligned}\mathbf{0}_{p_\theta} &= \frac{1}{S}\sum_{i=1}^{S}\sqrt{T}\iint \left[\frac{\pi_T^i(z,v;\theta_0)}{\pi_T^i(v;\theta_0)} - \frac{\pi_T(z,v)}{\pi_T(v)}\right]\nabla_\theta\tilde{\pi}_T(z|\,v;\theta_0)w_T(z,v)\mathbb{T}_{T,\delta}^2(v;\theta_0)\,dzdv + o_p(1)\mathbf{1}_{p_\theta} \\ &\quad + \left[\iint \left|\nabla_\theta\tilde{\pi}_T(z|\,v;\theta_0)\mathbb{T}_{T,\delta}(v;\theta_0)\right|_2 w_T(z,v)dzdv + o_p(1)\mathbf{1}_{p_\theta\times p_\theta}\right]\cdot\sqrt{T}\left(\theta_{T,S}-\theta_0\right),\end{aligned}$$

where the $o_p(1)$ term arises for arguments similar to ones leading to (B2) in appendix B.2 (see Al-M05 for more details). Tedious computations (also detailed in Al-M05) then lead to,

$$\mathbf{0}_{p_\theta} = \frac{1}{S}\sum_{i=1}^{S}\left(D_{1T}^i + D_{2T}^i\right) - D_{1T}^0 + \left[D_{3T} + o_p(1)\mathbf{1}_{p_\theta\times p_\theta}\right]\cdot\sqrt{T}\left(\theta_{T,S}-\theta_0\right),$$

where

$$\begin{aligned}D_{1T}^i &\equiv \iint \frac{\nabla_\theta\tilde{\pi}_T(z|\,v;\theta_0)w_T(z,v)}{\pi_T^i(v;\theta_0)}\mathbb{T}_{T,\delta}^2(v;\theta_0)\,dA_T^i(z,v;\theta_0); \\ D_{1T}^0 &\equiv \iint \frac{\nabla_\theta\tilde{\pi}_T(z|\,v;\theta_0)w_T(z,v)}{\pi_T(v)}\mathbb{T}_{T,\delta}^2(v;\theta_0)\,dA_T(z,v); \\ D_{2T}^i &\equiv \iint \frac{\nabla_\theta\tilde{\pi}_T(z|\,v;\theta_0)E\left[\pi_T(z,v)\right]w_T(z,v)}{\pi_T^i(v;\theta_0)\cdot\pi_T(v)}\mathbb{T}_{T,\delta}^2(v;\theta_0)\,dz\left[dA_T(v) - dA_T^i(v;\theta_0)\right]; \\ D_{3T} &\equiv \iint \left|\nabla_\theta\tilde{\pi}_T(z|\,v;\theta_0)\mathbb{T}_{T,\delta}(v;\theta_0)\right|_2 w_T(z,v)dzdv;\end{aligned}$$

and $A_T^i(z,v;\theta_0)$, $A_T(z,v)$, $A_T(v)$ and $A_T^i(v;\theta_0)$ are defined similarly as in appendix B.2. One may now make use of the same strategy of proof in appendix B.2, and lemmas 6-8, and show that (see Al-M05 for details),

$$
\begin{aligned}
D_{1T}^{i} &\overset{d}{\to} D_{1}^{i} && \equiv \iint \frac{\nabla_{\theta}\pi\left(z|\,v;\theta_{0}\right)w\left(z,v\right)}{\pi\left(v;\theta_{0}\right)}d\omega_{i}^{0}\left(F(z,v)\right), && i=0,1,\cdots,S;\\
D_{2T}^{i} &\overset{d}{\to} D_{2}^{0}-D_{2}^{i} && \equiv \int_{V}\gamma\left(v\right)d\hat{\omega}_{0}^{0}\left(F(v)\right)-\int_{V}\gamma\left(v\right)d\hat{\omega}_{i}^{0}\left(F(v)\right), && i=1,\cdots,S;\\
D_{3T} &\overset{p}{\to} D_{3} && \equiv \iint \left|\nabla_{\theta}\pi\left(z|\,v;\theta_{0}\right)\right|_{2}w(z,v)dzdv;
\end{aligned}
$$

where $\omega_{i}^{0},\ i=0,1,\cdots,S$, are independent Generalized Brownian Bridges; and $\hat{\omega}_{i}^{0},\ i=0,1,\cdots,S$, are also independent Brownian Bridges; and

$$\gamma\left(v\right)\equiv\int_{Z}\frac{\nabla_{\theta}\pi(z|\,v;\theta_{0})\pi_{0}\left(z,v\right)w(z,v)}{\pi\left(v;\theta_{0}\right)^{2}}dz. \tag{C2}$$

As in appendix B.2, the terms $D_{1}^{i},\ i=0,1,\cdots,S$, are all independent and asymptotically centered Gaussian. Therefore, $\sqrt{T}\left(\theta_{T,S}-\theta_{0}\right)$ is asymptotically centered Normally distributed with variance $V\equiv D_{3}^{-1}\cdot var[\frac{1}{S}\sum_{i=1}^{S}(D_{1}^{i}-D_{2}^{i})-(D_{1}^{0}-D_{2}^{0})]\cdot D_{3}^{\top-1}$.

D. Proof of theorem 3

The following assumption contains one set of regularity conditions mentioned in the statement of theorem 3:

Assumption T2. *We have,*

(a) $\alpha_{T}\to 0$, $\delta_{T}\to 0$, $T^{\frac{1}{2}}\alpha_{T}^{3}\to\infty$, $T^{\frac{1}{2}}\delta_{T}^{2}\to\infty$, *and* $T^{\frac{1}{2}}\alpha_{T}^{2}\delta_{T}^{2}\to\infty$.

(b) *In addition to assumption T2-(a),* $\lambda\equiv\lambda_{T}\to 0$, $T^{\frac{1}{2}}\lambda_{T}^{q}\alpha_{T}^{3}\to\infty$, $T^{\frac{1}{2}}\lambda_{T}^{q}\alpha_{T}^{2}\delta_{T}\to\infty$, $T^{\frac{1}{2}}\lambda_{T}^{q+1}\alpha_{T}\delta_{T}^{3}\to\infty$, $T^{\frac{1}{2}}\lambda_{T}^{q-q^{*}}\alpha_{T}^{2}\delta_{T}^{2}\to\infty$, $T^{\frac{1}{2}}\lambda_{T}^{q-q^{*}}\alpha_{T}\delta_{T}^{4}\to\infty$, $\alpha_{T}^{3}\lambda_{T}^{-r}\to\infty$, $\delta_{T}^{2}\lambda_{T}^{-r}\to\infty$, $\alpha_{T}^{2}\delta_{T}^{2}\lambda_{T}^{-r}\to\infty$, *and* $\alpha_{T}\delta_{T}^{4}\lambda_{T}^{-r}\to\infty$.

D.1 Consistency

By appendixes B.1 and C.1, we only have to show that for all $\theta\in\Theta$, $\iint s_{iT}\left(z,v;\theta\right)dzdv\overset{p}{\to}0$, $i=1,2$, where s_{iT} are defined in appendix C.1, with $w_{T}\left(z,v\right)=\left[\pi_{T}\left(v\right)/\,\pi_{T}\left(z|\,v\right)\right]\mathbb{T}_{T,\alpha}\left(z,v\right)$, $w\left(z,v\right)=m_{0}\left(v\right)/\,n_{0}\left(z,v\right)$. We proceed as in appendix C.1, and study these two integrals separately.

- For all $(z,v,\theta) \in Z \times V \times \Theta$,

$$
\begin{aligned}
& s_{1T}(z,v;\theta) \\
& \leq |\tilde{\pi}_T(z|\,v;\theta) - \pi_T(z|\,v)|\,|n(z,v;\theta) - n(z,v;\theta_0)| \left| \frac{\pi_T(v)}{\pi_T(z|\,v)} - \frac{m_0(v)}{n_0(z,v)} \right| \mathbb{T}_{T,\delta}(v;\theta)\,\mathbb{T}_{T,\alpha}(z,v) \\
& + |\tilde{\pi}_T(z|\,v;\theta) - \pi_T(z|\,v)|\,\mathbb{T}_{T,\delta}(v;\theta)\,|n(z,v;\theta) - n(z,v;\theta_0)| \frac{m_0(v)}{n_0(z,v)} [1 - \mathbb{T}_{T,\alpha}(z,v)] \\
& \leq \ell_{1T}(z,v;\theta) \cdot \ell_{2T}(z,v;\theta) \cdot m_0(z,v)\,|n(z,v;\theta) - n(z,v;\theta_0)| \\
& + \ell_{2T}(z,v;\theta) \cdot [n(z,v;\theta) - n(z,v;\theta_0)]^2\, m_0(z,v)\,\mathbb{T}_{T,\delta}(v;\theta) \\
& + \ell_{3T}(z,v;\theta)\,|n(z,v;\theta) - n(z,v;\theta_0)|\, m_0(z,v)\, m_0(v)\,[1 - \mathbb{T}_{T,\alpha}(z,v)] \\
& + \frac{m_0(v)}{n_0(z,v)} [n(z,v;\theta) - n(z,v;\theta_0)]^2\,\mathbb{T}_{T,\delta}(v;\theta)\,[1 - \mathbb{T}_{T,\alpha}(z,v)] \\
& \equiv s_{11T}(z,v;\theta) + s_{12T}(z,v;\theta) + s_{13T}(z,v;\theta) + s_{14T}(z,v;\theta),
\end{aligned}
$$

where

$$
\begin{aligned}
\ell_{1T}(z,v;\theta) & \equiv \left[\frac{1}{S} \sum_{i=1}^{S} \left| \pi_T^i(z|\,v;\theta) - n(z,v;\theta) \right| + |\pi_T(z|\,v) - n(z,v;\theta_0)| \right] \mathbb{T}_{T,\delta}(v;\theta) \\
\ell_{2T}(z,v;\theta) & \equiv \frac{1}{m_0(z,v)} \left| \frac{\pi_T(v)}{\pi_T(z|\,v)} - \frac{m_0(v)}{n_0(z,v)} \right| \mathbb{T}_{T,\alpha}(z,v) \\
\ell_{3T}(z,v;\theta) & \equiv \frac{\ell_{1T}(z,v;\theta)}{m_0(z,v)\, n_0(z,v)}
\end{aligned}
$$

By lemmas 2-(a), 3-(a) and 4-(a), $\iint s_{1jT} \xrightarrow{p} 0$ for all $\theta \in \Theta$ and $j = 1,2,3$. As regards the s_{14T} term, notice that function $n(z,v;\theta_0)^{-1} [n(z,v;\theta) - n(z,v;\theta_0)]^2 m_0(v)$ is the integrand of the asymptotic objective function, which is bounded and integrable by assumption. Moreover, $|\mathbb{T}_{T,\delta}(v;\theta)[1 - \mathbb{T}_{T,\alpha}(z,v)]| \leq 1$, and $[1 - \mathbb{T}_{T,\alpha}(z,v)] \xrightarrow{p} P_\pi \{\pi_0(z_1,v_1) < \lim_{T\to\infty} \alpha_T\} - \iint_{(z,v):\pi_0(z,v)\in(\lim_{T\to\infty}\alpha_T, 2\lim_{T\to\infty}\alpha_T)} \mathbb{T}_{T,\alpha}(z,v)\, P_\pi(dz,dv)$, where P_π is now the stationary measure of (z,v). Hence $\iint s_{14T} \xrightarrow{p} 0$ for all $\theta \in \Theta$, and so $\iint s_{1T} \xrightarrow{p} 0$ for all $\theta \in \Theta$ as well.

- For all $(z,v,\theta) \in Z \times V \times \Theta$,

$$
\begin{aligned}
& s_{2T}(z,v;\theta) \\
& \leq [r_T(z,v;\theta) + r(z,v;\theta)]\,[|\tilde{\pi}_T(z|\,v;\theta) - n(z,v;\theta)| + |\pi_T(z|\,v) - n(z,v;\theta_0)|]\,\mathbb{T}_{T,\delta}(v;\theta) \\
& + [r_T(z,v;\theta) + r(z,v;\theta)]\,[n(z,v;\theta) - n(z,v;\theta_0)]\,[1 - \mathbb{T}_{T,\delta}(v;\theta)] \\
& \equiv s_{21T}(z,v;\theta) + s_{22T}(z,v;\theta).
\end{aligned}
$$

For all $(z, v, \theta) \in Z \times V \times \Theta$,

$$
\begin{aligned}
& s_{21T}(z, v; \theta) \\
& \leq |\tilde{\pi}_T(z|\, v; \theta) - n(z, v; \theta)|\, \mathbb{T}_{T,\delta}(v; \theta) \frac{1}{m_0(z, v)} \left| \frac{\pi_T(v)}{\pi_T(z|\, v)} - \frac{m_0(v)}{n_0(z, v)} \right| \mathbb{T}_{T,\alpha}(z, v)\, m_0(z, v) \\
& \times [|\tilde{\pi}_T(z|\, v; \theta) - n(z, v; \theta)| + |\pi_T(z|\, v) - n(z, v; \theta_0)|]\, \mathbb{T}_{T,\delta}(v; \theta) \\
& + |\pi_T(z|\, v) - n_0(z, v)|\, \mathbb{T}_{T,\delta}(v; \theta) \frac{1}{m_0(z, v)} \left| \frac{\pi_T(v)}{\pi_T(z|\, v)} - \frac{m_0(v)}{n_0(z, v)} \right| \mathbb{T}_{T,\alpha}(z, v)\, m_0(z, v) \\
& \times [|\tilde{\pi}_T(z|\, v; \theta) - n(z, v; \theta)| + |\pi_T(z|\, v) - n(z, v; \theta_0)|]\, \mathbb{T}_{T,\delta}(v; \theta) \\
& + |n(z, v; \theta) - n_0(z, v)|\, \mathbb{T}_{T,\delta}(v; \theta) \frac{1}{m_0(z, v)} \left| \frac{\pi_T(v)}{\pi_T(z|\, v)} - \frac{m_0(v)}{n_0(z, v)} \right| \mathbb{T}_{T,\alpha}(z, v)\, m_0(z, v) \\
& \times [|\tilde{\pi}_T(z|\, v; \theta) - n(z, v; \theta)| + |\pi_T(z|\, v) - n(z, v; \theta_0)|]\, \mathbb{T}_{T,\delta}(v; \theta) \\
& + |n(z, v; \theta) - n(z, v; \theta_0)|\, m_0(v)\, m_0(z, v) \\
& \times \frac{1}{n_0(z, v)\, m_0(z, v)} [|\tilde{\pi}_T(z|\, v; \theta) - n(z, v; \theta)| + |\pi_T(z|\, v) - n(z, v; \theta_0)|]\, \mathbb{T}_{T,\delta}(v; \theta) \\
& + |\tilde{\pi}_T(z|\, v; \theta) - \pi_T(z|\, v)|\, \mathbb{T}_{T,\delta}(v; \theta) \frac{m_0(v)}{n_0(z, v)} \mathbb{T}_{T,\alpha}(z, v) \\
& \times [|\tilde{\pi}_T(z|\, v; \theta) - n(z, v; \theta)| + |\pi_T(z|\, v) - n(z, v; \theta_0)|]\, \mathbb{T}_{T,\delta}(v; \theta) \\
& \equiv s_{211T}(z, v; \theta) + s_{212T}(z, v; \theta) + s_{213T}(z, v; \theta) + s_{214T}(z, v; \theta) + s_{215T}(z, v; \theta)
\end{aligned}
$$

By lemmas 2-(a), 3-(a) and 4-(a), $\iint s_{21jT} \xrightarrow{p} 0$ (all $\theta \in \Theta$ and $j = 1, 2, 3, 4$). Similar results for the s_{215T} term lead to $\iint s_{215T} \xrightarrow{p} 0$ (all $\theta \in \Theta$), and so $\iint s_{21T} \xrightarrow{p} 0$ for all $\theta \in \Theta$.

Next, for all $(z, v, \theta) \in Z \times V \times \Theta$,

$$
s_{22T}(z, v; \theta) \leq \left\{ s^*_{22T}(z, v; \theta)\, m_0(z, v) + [n(z, v; \theta) - n_0(z, v)]^2 \frac{m_0(v)}{n_0(z, v)} \right\} [1 - \mathbb{T}_{T,\delta}(v; \theta)],
$$

where

$$
\begin{aligned}
& s^*_{22T}(z, v; \theta) \\
& \equiv [|\tilde{\pi}_T(z|\, v; \theta) - n(z, v; \theta)| + |\pi_T(z|\, v) - n_0(z, v)| + |n(z, v; \theta) - n_0(z, v)|]\, \mathbb{T}_{T,\delta}(v; \theta) \\
& \times \frac{1}{m_0(z, v)} \left| \frac{\pi_T(v; \theta)}{\pi_T(z|\, v)} - \frac{m_0(v)}{n_0(z, v)} \right| \mathbb{T}_{T,\alpha}(z, v) \cdot |n(z, v; \theta) - n_0(z, v)| \\
& + \frac{1}{m_0(z, v)\, n_0(z, v)} |\tilde{\pi}_T(z|\, v; \theta) - \pi_T(z|\, v)|\, \mathbb{T}_{T,\delta}(v; \theta)\, m_0(v) \cdot |n(z, v; \theta) - n_0(z, v)|\, \mathbb{T}_{T,\alpha}(z, v).
\end{aligned}
$$

Similarly as in the previous appendixes, $1 - \mathbb{T}_{T,\delta}(v; \theta) \xrightarrow{p} 0$ and since $1 - \mathbb{T}_{T,\delta}(v; \theta) \leq 1$, and both $m_0(z, v)$ and $n(z, v; \theta_0)^{-1} [n(z, v; \theta) - n(z, v; \theta_0)]^2 m_0(v)$ are bounded and integrable,

$\iint s_{22T} \xrightarrow{p} 0$ for all $\theta \in \Theta$. Hence, $\iint s_{2T} \xrightarrow{p} 0$ for all $\theta \in \Theta$.

The case $\lambda \equiv \lambda_T \downarrow 0$ is dealt with similarly through lemmas 2-(b), 3-(b) and 4-(b).

D.2 Asymptotic normality

Let $\xi(z,v) \equiv \pi_0(z,v)\, w(z,v)/\, \pi_0(v)^2$, and consider the definition of γ in appendix C.2 (see (C2)). In terms of this new function ξ, γ is

$$\gamma(v) = \int_Z \nabla_\theta \pi(z|\, v;\theta_0)\, \xi(z,v) dz. \tag{D1}$$

Next, let

$$W_T^\xi \equiv \left\{ w_T(z,v) : w_T(z,v) = \xi_{1T}(v) \cdot \frac{\pi_T(v)^2}{\pi_T(z,v)} \mathbb{T}_{T,\alpha}(z,v) \right\},$$

where function ξ_{1T} satisfies the conditions in lemma 9. We study the asymptotic behavior of the CD-SNE for weighting functions $w_T \in W_T^\xi$. Consider the terms D_{jT}^i and D_{3T} in appendix C.2, and let $w_T \in W_T^\xi$. By lemmas 9 and 10, and assumption T2,

$$\begin{aligned}
D_{1T}^i &\xrightarrow{d} D_1^i \equiv \iint \frac{\nabla_\theta \pi(z|\, v;\theta_0)}{\pi(z|\, v;\theta_0)} \xi(z,v) d\omega_i^0(F(z,v)), && i = 0,1,\cdots,S; \\
D_{2T}^i &\xrightarrow{d} D_2^i \equiv \int_V \gamma(v)\, d\hat{\omega}_0^0(F(v)) - \int_V \gamma(v)\, d\hat{\omega}_i^0(F(v)), && i = 1,\cdots,S; \\
D_{3T} &\xrightarrow{p} D_3 \equiv \iint \left| \frac{\nabla_\theta \pi(z|\, v;\theta_0)}{\pi(z|\, v;\theta_0)} \right|_2 \xi(z,v) \pi_0(z,v)\, dz dv.
\end{aligned}$$

Moreover, for any $w_T \in W_T^\xi$, the limiting function in (D1) $\xi(z,v) = \xi_1(v)$. But for all $v \in V$, $\int_Z \nabla_\theta \pi(z|\, v;\theta_0)\, dz = 0$. Hence $\gamma(v) = 0$ for all $v \in V$, and then $D_2^i \equiv 0$. So we have shown the following result:

Proposition 2. *Under the assumptions of theorem 2 and assumption T2, CD-SNEs with weighting functions $w_T \in W_T^\xi$ are consistent and asymptotically normal with variance/covariance matrix*

$$V \equiv \left(1 + \frac{1}{S}\right) \cdot \left\{ var(\Psi_1) + \sum_{k=1}^{\infty} \left[cov(\Psi_1, \Psi_{1+k}) + cov(\Psi_{1+k}, \Psi_1) \right] \right\}$$

(provided it exists finitely), where $\Psi_i \equiv \Psi(z_i, v_i)$ and,

$$\Psi(z,v) \equiv \left[\iint \left| \frac{\nabla_\theta \pi(u_1|\, u_2;\theta_0)}{\pi(u_1|\, u_2;\theta_0)} \right|_2 \xi_1(u_2)\, \pi_0(u_1,u_2)\, du_1 du_2 \right]^{-1} \frac{\nabla_\theta \pi(z|\, v;\theta_0)}{\pi(z|\, v;\theta_0)} \xi_1(v). \tag{D2}$$

Theorem 3 is a special case of proposition 2 with $\xi_1(\cdot) = \xi_{1T}(\cdot) \equiv 1$ and $(z,v) = (y_2, y_1)$. The efficiency claim follows by the standard score martingale difference argument.

References

Aït-Sahalia, Y., 1994, "The Delta Method for Nonparametric Kernel Functionals," working paper, Princeton University.

Aït-Sahalia, Y., 1996, "Testing Continuous-Time Models of the Spot Interest Rate," *Review of Financial Studies*, 9, 385–426.

Aït-Sahalia, Y., 2002, "Maximum Likelihood Estimation of Discretely Sampled Diffusions: a Closed-Form Approximation Approach," *Econometrica*, 70, 223-262.

Aït-Sahalia, Y., 2003, "Closed-Form Likelihood Expansions for Multivariate Diffusions," working paper, Princeton University.

Altissimo, F. and A. Mele, 2005, "Simulated Nonparametric Estimation of Dynamic Models with Applications to Finance (Unabridged version)," forthcoming as Financial Markets Group working paper, London School of Economics.

Andrews, D.W.K., 1995, "Nonparametric Kernel Estimation for Semiparametric Models," *Econometric Theory*, 11, 560-596.

Arcones, M.A. and B. Yu, 1994, "Central Limit Theorems for Empirical and U-Processes of Stationary Mixing Sequences," *Journal of Theoretical Probability*, 7, 47-71.

Billio, M. and A. Monfort, 2003, "Kernel-Based Indirect Inference," *Journal of Financial Econometrics*, 1, 297-326.

Basu, A. and B. G. Lindsay, 1994, "Minimum Disparity Estimation for Continuous Models: Efficiency, Distributions and Robustness," *Annals of the Institute of Statistical Mathematics*, 46, 683-705.

Bickel, P.J., 1982, "On Adaptive Estimation," *Annals of Statistics*, 70, 647-671.

Bickel, P.J. and M. Rosenblatt, 1973, "On Some Global Measures of the Deviations of Density Function Estimates," *Annals of Statistics*, 1, 1071-1095.

Carrasco, M., L. P. Hansen and X. Chen, 1999, "Time Deformation and Dependence," working paper, University of Rochester.

Carrasco, M., M. Chernov, J.-P. Florens and E. Ghysels, 2004, "Efficient Estimation of Jump-Diffusions and General Dynamic Models with a Continuum of Moment Conditions," working paper, University of Rochester.

Chen, X., L. P. Hansen and M. Carrasco, 1999, "Nonlinearity and Temporal Dependence," working paper, University of Rochester.

Chen, X, O. Linton, and P. M. Robinson, 2001, "The Estimation of Conditional Densities," *Journal of Statistical Planning and Inference*, 71-84.

Corradi, V. and N. R. Swanson, 2005, "Bootstrap Specification Tests for Diffusion Processes," *Journal of Econometrics*, 124, 117-148.

Duffie, D. and K.J. Singleton, 1993, "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929-952.

Elerian, O., S. Chib and N. Shephard, 2001, "Likelihood Inference for Discretely Observed Nonlinear Diffusions," *Econometrica*, 69, 959-993.

Fan, Y., 1994, "Testing the Goodness-of-Fit of a Parametric Density Function by Kernel Method," *Econometric Theory*, 10, 316-356.

Fermanian, J.-D. and B. Salanié, 2004, "A Nonparametric Simulated Maximum Likelihood Estimation Method," *Econometric Theory*, 20, 701-734.

Gallant, A. R., 2001, "Effective Calibration," working paper, University of North Carolina.

Gallant, A. R. and H. White, 1988, *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, Oxford, Basil Blackwell.

Gallant, A. R. and G. Tauchen, 1996, "Which Moments to Match ?," *Econometric Theory*, 12, 657-681.

Gouriéroux, C., A. Monfort and E. Renault, 1993, "Indirect Inference," *Journal of Applied Econometrics*, 8, S85-S118.

Hajivassiliou, V. and D. McFadden, 1998, "The Method of Simulated Scores for the Estimation of Limited-Dependent Variable Models," *Econometrica,* 66, 863-896.

Hansen, L. and J. A. Scheinkman, 1995, "Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes," *Econometrica*, 63, 767-804.

Härdle, W. and E. Mammen, 1993, "Comparing Nonparametric versus Parametric Regression Fits," *Annals of Statistics*, 21, 1926-1947.

Karatzas, I. and S. Shreve, 1991, *Brownian Motion and Stochastic Calculus*, Berlin: Springer Verlag.

Kloeden, P.E. and E. Platen, 1999, *Numerical Solutions of Stochastic Differential Equations*, Berlin: Springer Verlag.

Laroque, G. and B. Salanié, 1989, "Estimation of Multimarket Fix-Price Models: An Application of Pseudo-Maximum Likelihood Methods," *Econometrica,* 57, 831-860.

Laroque, G. and B. Salanié, 1993, "Simulation-Based Estimation of Models with Lagged Latent Variables," *Journal of Applied Econometrics*, 8, S119-S133.

Laroque, G. and B. Salanié, 1994, "Estimating the Canonical Disequilibrium Model: Asymptotic Theory and Finite Sample Properties," *Journal of Econometrics*, 62, 165-210.

Lee, B-S. and B. F. Ingram, 1991, "Simulation Estimation of Time-Series Models," *Journal of Econometrics*, 47, 197-207.

Lee, L. F., 1995, "Asymptotic Bias in Simulated Maximum Likelihood Estimation of Discrete Choice Models," *Econometric Theory*, 11, 437-483.

Lindsay, B. G., 1994, "Efficiency versus Robustness: The Case for Minimum Hellinger Distance and Related Methods," *Annals of Statistics*, 22, 1081-1114.

Linton, O. and Z. Xiao, 2000, "Second Order Approximation for Adaptive Regression Estimators," working paper, London School of Economics.

McFadden, D., 1989, "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57, 995-1026.

Menzly L., T. Santos and P. Veronesi, 2004, "Understanding Predictability," *Journal of Political Economy*, 112, 1-47.

Meyn, S.P. and R. L. Tweedie, 1993, "Stability of Markovian Processes III: Foster-Lyapunov Criteria for Continuous-Time Processes," *Advances in Applied Probability*, 25, 518-548.

Newey, W. K., 1991, "Uniform Convergence in Probability and Stochastic Equicontinuity," *Econometrica*, 59, 1161-1167.

Newey, W. K. and D. L. McFadden, 1994, "Large Sample Estimation and Hypothesis Testing," in Engle, R.F. and D. L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, chapter 36, 2111-2245. Amsterdam: Elsevier.

Pagan, A. and A. Ullah, 1999, *Nonparametric Econometrics*, Cambridge: Cambridge University Press.

Pakes, A. and D. Pollard, 1989, "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027-1057.

Pastorello, S., V. Patilea, and E. Renault, 2003, "Iterative and Recursive Estimation in Structural Non Adaptive Models," *Journal of Business and Economic Statistics*, 21, 449-509.

Pedersen, A.R., 1995, "A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations," *Scandinavian Journal of Statistics*, 22, 55-71.

Rao, C.R., 1962, "Efficient Estimates and Optimum Inference Procedures in Large Samples," *Journal of The Royal Statistical Society*, Series B, 24, 46-63.

Santa-Clara, P. 1995, "Simulated Likelihood Estimation of Diffusions With an Application to the Short Term Interest Rate," Ph.D. dissertation, INSEAD.

Singleton, K.J., 2001, "Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function," *Journal of Econometrics*, 102, 111-141.

Smith, A., 1993, "Estimating Nonlinear Time Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8, S63-S84.

Stone, C., 1975, "Adaptive Maximum Likelihood Estimation of a Location Parameter," *Annals of Statistics*, 3, 267-284.

Tjøstheim D., 1990, "Non-Linear Time Series and Markov Chains," *Advances in Applied Probability,* 22, 587-611.

White, H., 1994, *Estimation, Inference and Specification Analysis*, Cambridge: Cambridge University Press.

Tables 2 through 5

Table 2 - Monte Carlo experiments. (Vasicek model (13).) True parameter values are: $b_1 = 3.00$, $b_2 = 0.50$ and $a_1 = 3.00$.

Sample	Estimators		b_1	b_2	a_1
T=1000	CD-SNE	Mean	2.87	0.49	3.08
		Median	2.89	0.47	3.10
		Sample std. dev.	0.97	0.17	0.29
		Asymptotic std. dev.	1.10	0.19	0.23
		Coverage rate 90% conf. interval	0.95	0.92	0.82
	CD-SNE - Double bandwidth	Mean	2.65	0.43	3.23
		Median	2.56	0.44	3.16
		Sample std. dev.	0.84	0.17	0.28
	CD-SNE - Half bandwidth	Mean	2.98	0.54	2.97
		Median	2.93	0.56	3.04
		Sample std. dev.	1.06	0.23	0.40
	SNE	Mean	3.20	0.55	2.89
		Median	3.07	0.52	2.76
		Sample std. dev.	1.11	0.25	0.41
		Asymptotic std. dev.	1.24	0.22	0.31
		Coverage rate 90% conf. interval	0.95	0.85	0.81
	Analytical-NE	Mean	3.47	0.57	3.55
		Median	3.20	0.47	3.46
		Sample std. dev.	2.09	0.64	0.62
	MLE	Mean	3.74	0.62	3.01
		Median	3.93	0.63	2.99
		Sample std. dev.	1.21	0.20	0.07
T=500	CD-SNE	Mean	2.95	0.48	3.14
		Median	2.95	0.48	3.12
		Sample std. dev.	1.03	0.24	0.42
		Asymptotic std. dev.	1.36	0.26	0.32
		Coverage rate 90% conf. interval	0.94	0.94	0.83
	SNE	Mean	3.06	0.58	2.58
		Median	3.03	0.51	2.51
		Sample std. dev.	1.41	0.35	0.71
		Asymptotic std. dev.	1.65	0.31	0.57
		Coverage rate 90% conf. interval	0.97	0.84	0.76
	MLE	Mean	3.99	0.70	2.99
		Median	4.01	0.69	3.00
		Sample std. dev.	1.36	0.27	0.10

Table 3 - Monte Carlo experiments. (Continuous-time stochastic volatility model (15).) True parameter values are: $b_1 = 3.00$, $b_2 = 0.50$, $a_1 = 3.00$, $b_3 = 1.00$ and $a_2 = 0.30$.

Sample	Estimator		b_1	b_2	a_1	b_3	a_2
T=1000	CD-SNE	Mean	3.03	0.48	3.05	1.11	0.34
		Median	3.07	0.49	3.04	0.98	0.32
		Sample std. dev.	0.93	0.22	0.40	0.59	0.20
		Asymptotic std. dev.	1.17	0.22	0.32	0.45	0.16
		Coverage rate 90% conf. interval	0.95	0.88	0.83	0.82	0.83
	SNE	Mean	2.91	0.48	2.97	1.10	0.38
		Median	2.95	0.49	2.91	1.05	0.33
		Sample std. dev.	1.15	0.22	0.50	0.52	0.20
		Asymptotic std. dev.	1.20	0.23	0.31	0.50	0.18
		Coverage rate 90% conf. interval	0.91	0.91	0.78	0.84	0.88
T=500	CD-SNE	Mean	2.94	0.49	3.12	1.30	0.34
		Median	2.99	0.49	3.07	1.11	0.31
		Sample std. dev.	1.41	0.30	0.62	0.77	0.27
		Asymptotic std. dev.	1.69	0.31	0.44	0.63	0.22
		Coverage rate 90% conf. interval	0.95	0.89	0.80	0.83	0.85
	SNE	Mean	2.96	0.46	2.92	1.29	0.33
		Median	3.01	0.47	2.87	1.12	0.29
		Sample std. dev.	1.52	0.29	0.61	0.75	0.25
		Asymptotic std. dev.	1.75	0.32	0.43	0.70	0.25
		Coverage rate 90% conf. interval	0.94	0.92	0.81	0.87	0.89

Table 4 - Monte Carlo experiments. (Univariate discrete-time stochastic volatility model (16).) True parameter values are: $\phi = 0.95$, $\sigma_b = 0.025$ and $\sigma_e = 0.260$. Sample size: $T = 500$.

Estimator		ϕ	σ_b	σ_e
CD-SNE	Mean	0.909	0.024	0.229
	Median	0.939	0.023	0.210
	Sample std. dev.	0.102	0.003	0.131
	Asymptotic std. dev.	0.115	0.004	0.089
	Coverage rate 90% conf. interval	0.92	0.93	0.74
SNE	Mean	0.942	0.027	0.297
	Median	0.960	0.026	0.274
	Sample std. dev.	0.095	0.005	0.144
	Asymptotic std. dev.	0.121	0.005	0.093
	Coverage rate 90% conf. interval	0.94	0.89	0.72
QML*	Mean	0.906	···	0.302
	Sample std. dev.	0.18	···	0.17
MCL*	Mean	0.930	···	0.233
	Sample std. dev.	0.10	···	0.07
NPSML*	Mean	0.913	0.022	0.318
	Sample std. dev.	0.10	0.003	0.17

* QML stands for Quasi Maximum Likelihood; MCL for Monte Carlo Maximum Likelihood; and NPSML for Nonparametric Simulated Maximum Likelihood.

Table 5 - Monte Carlo experiments. (Bivariate discrete-time stochastic volatility model (17).) True parameter values are: $\phi = 0.95$, $\sigma_{b1} = 0.025$, $\sigma_{b2} = 0.025$ and $\sigma_e = 0.260$. Sample size: $T = 500$.

Estimator		ϕ	σ_{b1}	σ_{b2}	σ_e
CD-SNE	Mean	0.916	0.025	0.026	0.289
	Median	0.919	0.026	0.027	0.287
	Sample std. dev.	0.072	0.004	0.004	0.101
	Asymptotic std. dev.	0.080	0.004	0.004	0.113
	Coverage rate 90% conf. interval	0.92	0.83	0.88	0.91
SNE	Mean	0.913	0.027	0.027	0.365
	Median	0.938	0.026	0.027	0.331
	Sample std. dev.	0.084	0.004	0.004	0.164
	Asymptotic std. dev.	0.085	0.005	0.005	0.154
	Coverage rate 90% conf. interval	0.88	0.92	0.93	0.88