A Scrambled Method of Moments (Preliminary: do not cite, do not circulate)

Jean-Jacques Forneron*

September 9, 2019

Abstract

Quasi-Monte Carlo (qMC) methods are a powerful alternative to classical Monte-Carlo (MC) integration. Under certain conditions, they can approximate the desired integral at a faster rate than the usual Central Limit Theorem, resulting in more accurate estimates. This paper explores these methods in a simulation-based estimation setting with an emphasis on the scramble of Owen (1997). For cross-sections and short-panels, the resulting Scrambled Method of Moments simply replaces the random number generator with the scramble (available in most software) to reduce simulation noise. Scrambled Indirect Inference estimation is also considered. For time-series, qMC may not apply directly because a curse of dimensionality over the time dimension. A simulation approach and a class of moments which circumvent this issue are described. Asymptotic results are given for each algorithm. Monte-Carlo examples illustrate these results in finite samples, including an income process with "lots of heterogeneity."

JEL Classification: C11, C12, C13, C32, C36.
Keywords: Simulated Method of Moments, Indirect Inference, quasi-Monte Carlo, Scramble.

*Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215 USA.
Email: jjmf@bu.edu, Website: http://jjforneron.com.
Comments are welcome. All errors are my own.

1 Introduction

Simulation-based estimation is a popular approach to estimate complex economic models. The econometrician simply matches a set of sample with simulated moments drawn from a model of interest. The resulting Simulated Method of Moments (SMM) or Indirect Inference (II) estimator makes estimation feasible even though the likelihood or the moments' expectation, required for MLE and GMM, may be impossible or impractical to compute. However, using simulations rather than analytical computations introduces simulation noise, which increases the variance of the estimates. Also, the resulting objective function is typically non-smooth and hence more difficult to minimize numerically than using analytical computations. In theory, simulating many samples reduces simulation noise and smoothes the objective. In practice, this may not be feasible because of the increased computational cost. Alternatively, variance reduction methods such as antithetic draws[1] can reduce simulation noise with negligible computational overhead.

This paper investigates quasi-Monte Carlo (qMC) integration, another variance reduction approach, for simulation-based estimation - with an emphasis on the scramble of Owen (1997). Under certain conditions, qMC can approximate an expectation at a faster rate than the usual Monte-Carlo (MC) Central Limit Theorem. This suggests that a Scrambled Method of Moments could outperform conventional SMM estimates using as many simulated samples. This is shown to be the case for a large class of models in cross-sections and short-panels with potentially non-smooth moments as in McFadden (1989); Pakes & Pollard (1989) or auxiliary parameters as in Gouriéroux et al. (1993). For time-series, qMC may not apply directly because a curse of dimensionality over the time dimension. A class of models and moments which circumvent this issue are described.

Using the scramble in an estimation setting poses several practical and theoretical challenges. These sequences are designed to approximate a fixed integral of an *iid* sequence. Improper use of the scramble under dependence or with covariates may result in inconsistent estimators. Hence, the first and main contribution of the paper is methodological. The second contribution is theoretical. Uniform Laws of Large Numbers (ULLN) and Central Limit Theorems (CLT) are provided to handle smooth moments in cross-sections and short-panels. Scrambled draws are random and identically distributed *but not independent*. This makes it more challenging to handle time-series and non-smooth moments. A stochastic equicontinuity result for cross-sections and short-panels is established by re-writing the scrambled

[1] Antithetic draws are a popular approach to reduce simulation noise. See Section 2.1 for a brief overview.

empirical process as the sum of a non-identically distributed but independent array with a standard qMC sequence. This allows to invoke existing results for each term separately. In the time-series setting, a similar strategy allows to invoke results for dependent heterogeneous arrays found in e.g. White (1984).

The finite sample properties of the Scrambled Method of Moments are illustrated using several simple Monte-Carlo examples and an income process with "lots of heterogeneity" (Browning et al., 2010) in the heterogeneous agents literature where simulation-based estimation is commonly used (see e.g. Guvenen, 2011). In this example, the scramble improves on SMM with random and antithetic draws in terms of variance. Furthermore, optimization over the 2,000 replications was on average 15% faster with the scramble than SMM using as many simulated samples - most likely because scrambled moments are smoother than MC moments.

Structure of the Paper

After a review of the literature, Section 2 provides an overview of (quasi)-Monte Carlo integration which is lesser known in economics. Section 3 shows how to implement the Scrambled Method of Moments in various settings. Section 4 illustrates its finite sample properties with Monte-Carlo simulations. Since the main interest of the paper is methodological, the asymptotic theory is differed to Appendix A and the proofs to Appendix B. Section 5 concludes.

Related Literatures

There are two related literatures: simulation-based estimation and variance reduction techniques. In economics, simulation-based estimation includes the Simulated Method of Moments (McFadden, 1989; Pakes & Pollard, 1989; Duffie & Singleton, 1993), Indirect Inference (Gouriéroux et al., 1993) and the Efficient Method of Moments (Gallant & Tauchen, 1996). See Smith (2008) for an overview of simulation-based estimation in economics and common empirical applications. In statistics, Bayesian methods such as Approximate Bayesian Computation (also known as ABC; Marin et al., 2012) and Synthetic Likelihood (Wood, 2010) are more common. See Forneron & Ng (2018) for an overview and comparisons of these frequentist and Bayesian methods.

As discussed in the introduction, Monte-Carlo methods introduce simulation noise in the estimation which increases the variance of the estimator. There is large number of variance reduction techniques, the following summary will only cover some that are most relevant to simulation-based estimation. One approach is to use low-discrepancy sequences - this is more commonly known as quasi-Monte Carlo integration (see e.g. Lemieux, 2009; Dick & Pillichshammer, 2010, for an overview). qMC integration for *iid* sequences, is very similar to MC integration but with a specific sequence of numbers rather than random draws. Under certain conditions, qMC integration achieves faster than $\sqrt{n}$-rate convergence which generally translates into efficiency gains. qMC integration has been extended to non-linear state-space particle filtering (Gerber & Chopin, 2015, 2017), MCMC sampling (Owen & Tribble, 2005) and importance sampling for ABC (Buchholz & Chopin, 2017), among others. A key takeaway from these papers is that a lot of care is required in implementing qMC integration in non *iid* settings (MCMC or filtering) where naive implementations may be inconsistent. This may explain why it is only rarely used in empirical economics, even though their appeal has been known for some time (Judd, 1998). qMC typically requires the ability to transform the shocks from a uniform to the desired distribution by the Rosenblatt transform (inverse CDF). When this is not feasible, an alternative could be to rely on low-energy methods (Mak & Joseph, 2018). In Economics, antithetic draws are a popular variance reduction technique. However, they can lead to either efficiency gains or losses depending on the integrand as discussed in Section 2.1. Another variance reduction method, which is more popular in Statistics, is the control variates approach (see e.g. Robert & Casella, 2004).[2] The main idea is to augment the estimating sample and simulated moments with analytically tractable moments for the shocks themselves. This additional information can help reduce the uncertainty due to simulation noise.[3] Beyond variance reduction, the control functional approach (Oates et al., 2017), which uses all the information about the distribution of the shocks, can result in faster than $\sqrt{n}$-rate convergence. Efficiency gains require the moments in the control variates to span a sufficiently large space which could lead to a curse of dimensionality. For instance, the model of Section 4.2 has shocks with dimension $d = 30$ so that spanning polynomials of order up to 2 or 3 would require introducing 496 or $5,456$ additional moments respectively. The number of moments quickly becomes greater than the sample size itself.

[2]Note that, despite the similarity in names, this is not related to control variable estimation used in structural econometric estimation.

[3]See Davis et al. (2019) for an application of control variates to Indirect Inference. Control variates for qMC integration were considered in Hickernell et al. (2005).

2 (quasi)-Monte Carlo Integration and the Scramble

The following provides a brief overview of Monte-Carlo (MC) and quasi-Monte Carlo (qMC) integration methods.[4] Throughout, we are interested in evaluating the following integral:

$$I = \int_{[0,1]^d} f(u)du, \tag{1}$$

with a fixed or random sequence of points $u_1, \ldots, u_n$ in $[0,1]^d$:

$$\hat{I}_n = \frac{1}{n}\sum_{i=1}^{n} f(u_i). \tag{2}$$

2.1 Monte-Carlo Integration and Antithetic Draws

A widely applicable approach is MC integration. Take *iid* uniform draws $u_i \sim \mathcal{U}_{[0,1]^d}$ and compute the sample analog $\hat{I}_n^{MC} = \frac{1}{n}\sum_{i=1}^n f(u_i)$. Assuming $f(u_i)$ has finite variance, $\hat{I}_n^{MC}$ is unbiased and the approximation error $|\hat{I}_n^{MC} - I|$ is of order $\sqrt{\text{var}[f(u_i)]/n}$. This implies that in order to reduce the approximation error tenfold, the number of draws must be a hundred times greater: the computational cost increases faster than the approximation error declines.

A popular approach to reduce the variance is to use antithetic draws. For n even,

$$\hat{I}_n^{Anti} = \frac{1}{n}\sum_{i=1}^{n/2}[f(u_i) + f(1-u_i)], u_i \overset{iid}{\sim} \mathcal{U}_{[0,1]^d}.$$

This approach is only valid if $f(u_i)$ and $f(1-u_i)$ have the same distribution; for instance, $e_i = \Phi^{-1}(u_i) \sim \mathcal{N}(0,1)$ and $-e_i = \Phi^{-1}(1-u_i) \sim \mathcal{N}(0,1)$ as well. Without this property, when the distribution is asymmetric, $\hat{I}_n^{Anti}$ may not be consistent for I.

Assuming $f(u_i)$ and $f(1-u_i)$ have the same distribution, $\hat{I}_n^{Anti}$ is unbiased and $\text{var}(\hat{I}_n^{Anti}) = (\text{var}[f(u_i)] + \text{cov}[f(u_i), f(1-u_i)])/n$. If $\text{corr}[f(u_i), f(1-u_i)] = -1$, then $\text{var}(\hat{I}_n^{Anti}) = 0$; the estimator is exact as soon as $n = 2$. This improves significantly on MC integration. However, if $\text{corr}[f(u_i), f(1-u_i)] = +1$ then $\text{var}(\hat{I}_n^{Anti}) = 2\text{var}(\hat{I}_n^{MC})$. Now, $\hat{I}_n^{Anti}$ performs worse than $\hat{I}_n^{MC}$.

The performance of antithetic draws relative to simple MC draws will typically depend on both the parameter of interest and the choice of estimating moments. To illustrate, consider the following two examples. First, suppose $I = \mathbb{E}(e_i)$ where $e_i = \Phi^{-1}(u_i) \sim \mathcal{N}(0,1)$. Note

[4] For further reading, Lemieux (2009) provides a non-technical introduction to MC and qMC integration; Dick & Pillichshammer (2010) provide the underlying theory.

that $-e_i \sim \mathcal{N}(0,1)$ and $\hat{I}_n^{Anti}$ is consistent for I in this example. Since $\text{corr}(e_i, -e_i) = -1$, $\text{var}(\hat{I}_n^{Anti}) = 0$. The estimator is exact as soon as $n = 2$. Second, suppose $I = \mathbb{E}(e_i^2)$ with e_i as above. Now, $\text{corr}(e_i^2, [-e_i]^2) = +1$ and $\text{var}(\hat{I}_n^{Anti}) = 2\text{var}(\hat{I}_n^{MC})$. These examples suggest that the moments need to have some symmetry properties in the shocks in order to produce efficiency gains. This can be hard to check for intractable non-linear models.

2.2 quasi-Monte Carlo Integration

The discussion above shows that some sequences can outperform MC integration. For instance, for f smooth and $u_i \in [0,1]^d$ with $d = 1$ the lattice sequence $u_i = i/(n-1), i = 0, \dots, n-1$, the estimator

$$\hat{I}_n^{Lattice} = \frac{1}{n}\sum_{i=1}^n f(u_i), \quad u_i = i/(n-1),\, i \in \{0, \dots, n-1\}$$

has an approximation error of order $O(\|\partial_u f\|_\infty / n)$. The approximation error declines linearly with the computational cost. However for $d \geq 2$, lattice sequence has approximation errors of order $n^{-1/d}$ which is worse than MC as soon as $d \geq 3$. As a result, lattice sequences are rarely used in practice.

It is possible to break this curse of dimensionality. To achieve this, the qMC literature relies on two pivotal inequalities. The first is the Koksma-Hlawka inequality:

$$\Big|\frac{1}{n}\sum_{i=1}^n f(u_i) - \int_{[0,1]^d} f(u)du\Big| \leq \|f\|_{TV} \times D_n^\star(u_1, \dots, u_n), \tag{3}$$

where $\|f\|_{TV}$ is the total variance norm of f in the sense of Hardy and Krause:

$$\|f\|_{TV} = \sum_{\mathfrak{u} \subseteq \mathcal{I}_d} \int_{[0,1]^{|\mathfrak{u}|}} \Big|\frac{\partial^{|\mathfrak{u}|} f(\mathfrak{u})}{\partial \mathfrak{u}}\Big| d\mathfrak{u}, \tag{4}$$

$\partial^{|\mathfrak{u}|} f(\mathfrak{u})/\partial \mathfrak{u}$ consists of all univariate derivatives $\partial_{u_1} f(u), \dots, \partial_{u_d} f(u)$ and partial cross-derivatives $\partial^2_{u_1,u_2} f(u), \partial^2_{u_1,u_3} f(u), \partial^2_{u_2,u_3} f(u), \dots, \partial^2_{u_{d-1},u_d} f(u)$ up to order d with $\partial^d_{u_1,\dots,u_d} f(u)$. It *does not include repeated derivatives* such as $\partial^2_{u_1,u_1} f(u)$. What matters is the smoothness of f across the co-ordinates $u_1, \dots, u_d$. As a result, integrating over larger dimensions d typically requires additional smoothness in f over these cross-derivatives.

The other term in the Koksma-Hlawka inequality is $D_n^\star(u_1, \dots, u_n)$ which corresponds to the star discrepancy of the sequence $(u_1, \dots, u_n)$, defined as:

$$D_n^\star(u_1, \dots, u_n) = \sup_{u \in [0,1]^d} \Big|\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{u_i \in [0,u)} - \int_{[0,u)} 1 du\Big|. \tag{5}$$

In Statistics, this is the Kolmogorov-Smirnov distance between the empirical CDF of $(u_1, \ldots, u_n)$ and the population CDF of a uniform $\mathcal{U}_{[0,1]^d}$ distribution.

For a given function f, the user looking to reduce the approximation error in (3) will seek sequences with smaller star discrepancies (5). For *iid* random draws, $D_n^* = O_p(n^{-1/2})$ by Donsker theorems (van der Vaart & Wellner, 1996). Lattice sequences have $D_n^* = O(n^{-1/d})$ for $d \geq 1$.

For any sequence $(u_1, \ldots, u_n)$, the second pivotal inequality initially due to Roth (1954) and generalized by Schmidt (1970) provides a lower bound on its star discrepancy:

$$D_n^\star(u_1, \ldots, u_n) \geq C_d \times \frac{\log(n)^{d-1}}{n}, \tag{6}$$

where C_d is a universal constant which only depends on the dimension d. For any $d \geq 1$ fixed, the lower bound suggests a faster rate than MC is possible. There are a number of deterministic sequences which are close in rate to the bound (6); for instance the Sobol, Halton, van der Corput and Hammersley sequences. Most of these are readily available in statistical software.[5] There is, however, a curse of dimensionality when d becomes large. For instance, $C_d = 2^d$ for the Sobol sequence which increases very rapidly with d. As a result, in finite samples MC integration may outperform qMC integration when d is large. Under additional smoothness conditions, so-called higher-order sequences can achieve even faster rates of order $n^{-\alpha}$ (up to log-terms) for some $\alpha > 1$ which depends on the smoothness of f and higher-order properties of the sequence.

2.3 Randomized quasi-Monte Carlo and the Scramble

These results above are restrictive since they require f smooth, otherwise $\|f\|_{TV} = +\infty$ and the Koksma-Hlawka inequality (3) is uninformative. Also, $\hat{I}_n^{qMC}$ computed with deterministic sequences is biased and the approximation error is hard to evaluate numerically. This would make it difficult to compute standard errors in an estimation setting.

One solution is to use randomized quasi-Monte Carlo (RqMC) methods. A simple randomizer is the digital shift. Take one random draw $u \sim \mathcal{U}_{[0,1]^d}$, a qMC sequence $u_1, \ldots, u_n$ (e.g. Sobol, Halton) and compute $\tilde{u}_i = [u_i + u]$ modulo 1. The modulo operator is applied one dimension at a time. This shifts all the co-ordinates of $u_1, \ldots, u_n$ by the same random quantity u and preserves the order of magnitude of its star discrepancy $D_n^\star$. The randomized

[5] The R package *randtoolbox*, the *Sobol* module in Julia and *SOBOL* in C++ generate the Sobol sequence, Matlab has *quasirandomset* tools to generate several sequences and Gerber & Chopin (2017) provide C code to generate low-discrepancy sequences.

$\tilde{u}_i$ are identically distributed $\mathcal{U}_{[0,1]^d}$ but not independent. The estimator $\hat{I}_n^{RqMC}$ is unbiased. To approximate its variance, apply the digital shift with different draws u to approximate the integral several times and then compute the variance across these estimates (Lemieux, 2009).

A more popular approach, which will be the main focus of this paper, is the scramble introduced by Owen (1997). Similarly to the random shift above, it transforms a deterministic low-discrepancy sequence into random identically but not independent distributed uniform $\mathcal{U}_{[0,1]^d}$ draws. The main idea is to expand qMC sequence in the basis b under which it was constructed and apply a random permutation to its digits. For instance, Sobol sequences are constructed in basis $b = 2$, which is binary. As a result, scrambling amounts to randomly shuffling the zeros and ones in the binary representation and re-writing the result in base 10, the usual decimal system. The underlying theory is quite involved since it relies of Walsh expansions, an approach similar to Fourier expansions but in a digital basis b. See Dick & Pillichshammer (2010) for more details. The scramble approximates I under the same conditions as the classical CLT as shown in Theorem 1.

Theorem 1 (Owen, 1997)**.** *Let $u_1, \dots, u_n$ be a scrambled sequence using the algorithm of Owen (1997). Suppose that f is measurable and $f(u_i)$ has finite variance then:*

$$\frac{1}{n}\sum_{i=1}^{n} f(u_i) - \int_{[0,1]^d} f(u)du = o_p(n^{-1/2}).$$

Under additional smoothness conditions $\hat{I}_n^{scramble}$ approximates I at a near $n^{-3/2}$-rate; which is faster than deterministic qMC sequences. The scrambled estimator $\hat{I}_n^{Scramble}$ is unbiased. Its variance can be approximated the same way as for $\hat{I}_n^{RqMC}$. Note that these results assume $d \geq 1$ is fixed. In practice, MC may outperform the scramble for d large. Other scrambles have also been proposed by Hickernell (1996) and Matoušek (1998), among others. See Lemieux (2009) and Dick & Pillichshammer (2010) for additional references.

3 A Scrambled Method of Moments

This section introduces the main algorithms to implement the Scrambled Method of Moments and Scrambled Indirect Inference. The data generating process (DGP) is the same as in Gouriéroux et al. (1993):

$$y_{i,t} = g_{obs}(y_{i,t-1}, x_{i,t}, z_{i,t}; \theta) \tag{7}$$

$$z_{i,t} = g_{latent}(z_{i,t-1}, u_{i,t}; \theta) \text{ where } u_{i,t} \overset{iid}{\sim} \mathcal{U}_{[0,1]^d}. \tag{8}$$

A simple transformation allows to replace $u_{i,t}$ with $e_{i,t} = \Phi^{-1}(u_{i,t}) \overset{iid}{\sim} \mathcal{N}(0,1)$ or other distributions by the Rosenblatt transform. $i = 1,\dots,n$ indexes individuals and $t = 1,\dots,T$ the time-dimension. $y_{i,t}$ is the vector of observed outcome variables. $x_{i,t}$ is a vector of strictly exogenous covariates and $z_{i,t}$ a vector of unobserved latent variables. The functions g_{obs} and g_{latent} are assumed to be known up to a finite dimensional parameter θ to be estimated.

3.1 Static Models

For static models, which corresponds to cross-sections and short-panels, the t index will be omitted to re-write (7)-(8) and the moments as:

$$y_i = g(x_i, u_i; \theta), \quad \hat{\psi}_n = \frac{1}{n}\sum_{i=1}^n \psi(y_i, x_i), \tag{9}$$

where $y_i = (y_{i,1},\dots,y_{i,T})$ and $u_i = (u_{i,1},\dots,u_{i,T}) \in [0,1]^{T\times d_{u_{i,t}}}$. The dimension d of u_i is $T \times \dim(u_{i,t})$, using the notation in (7)-(8). Given a vector of moments $\hat{\psi}_n$ and a weighting matrix W_n, a simple SMM estimator is given in Algorithm 1.

Algorithm 1 Simulated Method of Moments for Static Models

Draw a random sequence $u_i^s \overset{iid}{\sim} \mathcal{U}_{[0,1]^d}, i = 1,\dots,n$; and $s = 1,\dots,S$
Simulate: $y_i^s(\theta) = g_{obs}(x_i, u_i^s; \theta)$
Compute: $\hat{\psi}_n^S(\theta) = \frac{1}{n\times S}\sum_{s=1}^S \sum_{i=1}^n \psi(y_i^s(\theta), x_i)$
Find: $\hat{\theta}_n^S = \text{argmin}_{\theta\in\Theta} \|\hat{\psi}_n - \hat{\psi}_n^S(\theta)\|_{W_n}$

Without covariates x_i, the expectation $\mathbb{E}[\hat{\psi}_n^S(\theta)]$ has the same form as (1). The scramble can be applied if the moments have finite variance. The resulting Algorithm 2 is thus very similar to SMM. In practice, one samples an $(nS)\times d$ matrix of scrambled shocks rather than

Algorithm 2 Scrambled Method of Moments for Static Models without Covariates

Draw a scrambled sequence $\tilde{u}_i \overset{iid}{\sim} \mathcal{U}_{[0,1]^d}, i = 1,\dots,n \times S$
Simulate: $\tilde{y}_i(\theta) = g_{obs}(\tilde{u}_i; \theta)$
Compute: $\hat{\psi}_n^S(\theta) = \frac{1}{n\times S}\sum_{i=1}^{n\times S} \psi(\tilde{y}_i(\theta))$
Find: $\hat{\theta}_n^S = \text{argmin}_{\theta\in\Theta} \|\hat{\psi}_n - \hat{\psi}_n^S(\theta)\|_{W_n}$

S different $n \times d$ matrices of random numbers. This is may be useful because using a large simulated sample of $n \times S$ observations implies a reduction in variance greater than S, as a consequence of the faster rate in Theorem 1, compared to using S independent simulated samples. Asymptotic results for Algorithm 2 are provided in Proposition A1 in the Appendix assuming the moments are smooth in θ. The assumptions are comparable to those required for SMM.

When $\hat{\psi}_n$ is a vector of auxiliary moments as (Gouriéroux et al., 1993), the results from Proposition A1 can be extended for the scramble as shown in Proposition A3. Again, the assumptions are comparable to those required for Indirect Inference. These Indirect Inference results could also be extended to non-smooth moments and time-series given appropriate changes to the assumptions.

In the presence of covariates, the expectation $\mathbb{E}[\hat{\psi}_n^S(\theta)]$ does not have the same form as (1):

$$\mathbb{E}[\hat{\psi}_n^S(\theta)] = \int_{[0,1]^d \times \mathcal{X}} \psi\left(g(x,u;\theta),x\right) f_x(x) dx du,$$

where f_x is joint density of the covariates x.[6] Without further assumptions, it is typically not possible to sample from the population f_x directly so that qMC sequence with $n \times S$ elements for (x_i, u_i) cannot be constructed. Taking the covariates as given, Algorithm 3 relies on S independent scrambled sequences of size n rather than a large sequence of size $n \times S$ as in Algorithm 2.[7]

Algorithm 3 Scrambled Method of Moments for Static Models with Covariates

Draw S independently scrambled sequences $\tilde{u}_i^s \overset{iid}{\sim} \mathcal{U}_{[0,1]^d}, i = 1, \dots, n$
Simulate: $\tilde{y}_i^s(\theta) = g_{obs}(x_i, \tilde{u}_i^s; \theta)$
Compute: $\hat{\psi}_n^S(\theta) = \frac{1}{n \times S} \sum_{i=1}^{n \times S} \psi(\tilde{y}_i^s(\theta), x_i)$
Find: $\hat{\theta}_n^S = \text{argmin}_{\theta \in \Theta} \|\hat{\psi}_n - \hat{\psi}_n^S(\theta)\|_{W_n}$

For each $s \in \{1, \dots, S\}$, the function $\mathbb{E}[\psi(y_i^s(\theta), x_i)|\tilde{u}_i^s = u]$ does not depend on x so that Theorem 1 can be applied to this conditional expectation, assuming it has finite variance. This insight was used to derive CLTs for moments based on hybrid sequences which combine MC draws with qMC sequences in Ökten et al. (2006) and Buchholz & Chopin (2017) for bounded ψ. The results in Proposition A2 extend these results to unbounded empirical processes over $\theta \in \Theta$, allowing $\hat{\psi}_n^s$ to be non-smooth in θ. The assumptions are more demanding than for SMM, although they could be weakened for smooth moments with covariates. The conditional expectation $\mathbb{E}[\hat{\psi}_n^S(\cdot)|\tilde{u}_1, \tilde{u}_2, \dots]$ itself is required to be smooth in

[6]The results could be extended to allow some components of x to be discrete. However, the assumptions in the Appendix imply that at least one of the covariates should have a continuous density.

[7]It is implicitly assumed that $(x_1, \dots, x_n)$ is a random sample. If the ordering is deterministic, then x_i and u_i) are not independent. Randomly shuffling the covariates without replacement solves this issue.

θ, *i.e.* integrating out the covariates smoothes out the sample and simulated moments. This implies that at least one of the covariates has a continuous density.

3.2 Dynamic Models

For dynamic models, which corresponds to time-series observations, the i index will be omitted to re-write (7)-(8) and the moments as:

$$y_t = g_{obs}(y_{t-1}, z_t; \theta), \quad z_t = g_{latent}(z_{t-1}, u_t; \theta), \quad u_t \overset{iid}{\sim} \mathcal{U}_{[0,1]^d} \tag{10}$$

$$\hat{\psi}_T = \frac{1}{T} \sum_{t=L+1}^{T} \psi(y_t, \dots, y_{t-L}). \tag{11}$$

Covariates are omitted to simplify the theoretical results. Only moments involving a fixed and finite number of lags L will be considered as explained below. Algorithm 4 details the SMM procedure to estimate (10)-(11).

Algorithm 4 Simulated Method of Moments for Dynamic Models

Draw a random sequence $u_t^s \overset{iid}{\sim} \mathcal{U}_{[0,1]^d}, t = 1, \dots, T;\ s = 1, \dots, S$
Set $(y_0^s, z_0^s) = (y_0, z_0)$, a fixed initial value
Simulate: $z_t^s(\theta) = g_{latent}(z_{t-1}^s, u_t^s; \theta)$ and $y_t^s(\theta) = g_{obs}(y_{t-1}^s(\theta), z_t^s(\theta); \theta)$
Compute: $\hat{\psi}_T^S(\theta) = \frac{1}{T \times S} \sum_{s=1}^{S} \sum_{t=L+1}^{n} \psi(y_t^s(\theta), \dots, y_{t-L}^s(\theta))$
Find: $\hat{\theta}_T^S = \text{argmin}_{\theta \in \Theta} \|\hat{\psi}_T - \hat{\psi}_T^S(\theta)\|_{W_T}$

To understand the issues caused by the dynamics for the scramble and qMC integration, note that for any initial value (y_0, z_0), y_t can be re-written as:

$$y_t = g_t(u_t, \dots, u_1, y_0, z_0; \theta),$$

for some function g_t which can be expressed in terms of g_{obs} and g_{latent}. Using this notation, the expected value of $\hat{\psi}_T$ can be re-written as:

$$\mathbb{E}(\hat{\psi}_T) = \frac{1}{T} \sum_{t=L+1}^{T} \int_{[0,1]^{t \times d}} \psi \circ (g_t, \dots, g_{t-L})(u_t, \dots, u_1, y_0, z_0) du_t \dots du_1.$$

The expectation above differs from the qMC setting in (1) in several ways. First, the function to be integrated involves g_t which varies with t unlike the function in (1). Second, the integral is computed over $u_1, \dots, u_t$ which has a dimension t that increases with the sample size. This implies a curse of dimensionality for qMC which requires the dimension d to be fixed. Third, both randomized and non-randomized qMC sequences are identically but not independently distributed. A naive implementation of the scramble could introduce spurious dependence in the simulated data and the resulting estimator may not be consistent as a result.

Implementing qMC integration in a dynamic setting without additional structure comes at a cost. In finance, qMC sequences are used to simulate long time-series and price financial derivatives (see e.g. Paskov & Traub, 1995; Lemieux, 2009). This is done by setting $d = T$ and sampling a very large number n of financial series. In the present setting, this amounts to picking S very large and $d = T$ which is not computationally attractive compared to standard SMM.[8] For state-space filtering, Gerber & Chopin (2015, 2017) propose a Hilbert sorting step to re-sample draws into a low-discrepancy sequence using the fractal map from $[0, 1]$ to $[0, 1]^d$ developed by Hilbert in Mathematics. This Hilbert map can be challenging to implement in practice and suffers from a curse of dimensionality.

3.2.1 qMC-only Approach

The class of moments described in (11) where the number of lags L is fixed and finite allows to circumvent these issues. To get some intuition, suppose that it is possible to draw $(y_t^1, z_t^1) = F_{y,z}^{(-1)}(v_t^1)$ from the stationary distribution directly using the Rosenblatt transform with $v_t^1 \sim \mathcal{U}_{[0,1]^{\dim(v_t^1)}}$. Then, using additional shocks $u_t^2, \dots, u_t^L$ one could simulate a short time-series consisting of $L \geq 1$ observations for each $t = 1, \dots, T \times S$:

$$
\begin{aligned}
(y_t^1, z_t^1) &= F_{y,z}^{(-1)}(v_t^1) \\
(y_t^2, z_t^2) &= \big(g_{obs}(y_t^1, z_t^2; \theta),\ g_{latent}(z_t^1, u_t^2; \theta)\big) \\
&\vdots \\
(y_t^L, z_t^L) &= \big(g_{obs}(y_t^{L-1}, z_t^L; \theta),\ g_{latent}(z_t^{L-1}, u_t^L; \theta)\big)\,.
\end{aligned}
$$

The resulting draws $(y_t^1, \dots, y_t^L)$ are *iid* over $t = 1, \dots, T \times S$ from the stationary distribution by construction.[9] This is now within the setting of (1). Algorithm 5 describes a Scrambled Method of Moments for models where simulating as described above is feasible. The main idea is to simulate the $(y_t^1, \dots, y_t^L)$ $T \times S$ times with scrambled shocks $(v_t^1, u_t^2, \dots, u_t^L)_{t=1,\dots,T\times S} \in [0, 1]^d$ with dimension $d = \dim(v_t^1, u_t^2, \dots, u_t^L)$. which depends on the dimension of the shocks and the numbers of lags L. Note that while Algorithm 4 requires $T \times S$ draws, Algorithm 5 effectively requires $n \times S \times L$ draws. However, the latter Algorithm

[8] Recall that for the Sobol sequence $C_d = 2^d$ so that the error would be of the order of $2^T/S$. Consistency of the qMC integral would require $S \gg 2^T$, *i.e.* S needs to grow exponentially fast with the sample size T.

[9] This idea was also used in Davis et al. (2019) but as a variance reduction method with MC draws.

is massively parallel over t so that for some models it may run faster than the former in a parallel environment. Proposition A4 provides the asymptotic results for Algorithm 5.[10]

Algorithm 5 Scrambled Method of Moments for Dynamic Models - qMC-only Approach

Draw a scrambled sequence $\tilde{u}_t = (\tilde{v}_t, \tilde{u}_t^2, \dots, \tilde{u}_t^L) \in [0,1]^{d\times(L-1)+\tilde{d}}$, $t = 1, \dots, T \times S$
Compute $(\tilde{y}_t^1(\theta), \tilde{z}_t^1(\theta)) = F^{-1}(\tilde{v}; \theta)$ for $t = 1, \dots, T \times S$
Simulate: $\tilde{z}_t^\ell(\theta) = g_{latent}(\tilde{z}_t^{\ell-1}, \tilde{u}_t^\ell; \theta)$ and $\tilde{y}_t^\ell(\theta) = g_{obs}(\tilde{y}_t^{\ell-1}(\theta), \tilde{z}_t^\ell(\theta); \theta)$ for $\ell = 2, \dots, L$
Compute: $\hat{\psi}_T^S(\theta) = \frac{1}{T\times S}\sum_{t=1}^{T\times S} \psi(\tilde{y}_t^L(\theta), \dots, \tilde{y}_t^1(\theta))$
Find: $\hat{\theta}_T^S = \text{argmin}_{\theta\in\Theta} \|\hat{\psi}_T - \hat{\psi}_T^S(\theta)\|_{W_T}$

Sampling from the stationary distribution directly is feasible for some DGPs such as the Gaussian ARMA model (see the Monte-Carlo example in Section 4.1.3) or the following stochastic volatility process:

$$\log(\sigma_t) = \mu_\sigma + \rho_\sigma \log(\sigma_{t-1}) + \kappa_\sigma e_{t,1}, \quad y_t = \sigma_t e_{t,2}, \quad (e_{t,1}, e_{t,2}) \overset{iid}{\sim} \mathcal{N}(0, I_2).$$

Since the log-volatility follows a Gaussian AR(1) process, one can simply draw $\log(\sigma_t^1) \sim \mathcal{N}(\mu_\sigma/(1-\rho_\sigma, \kappa_\sigma^2/[1-\rho_\sigma^2])$ and $y_t^1 = \sigma_t^1 e_{t,2}^1$ where $e_{t,2}^1$ to simulate (y_t^1, σ_t^1) from their stationary distribution. For more complex DGPs this may not be feasible, however.

3.2.2 Hybrid MC-qMC Approach

When the direct approach in Algorithm 5 is not feasible, an alternative is to sample the initial draws (y_t^1, z_t^1) by MC methods and then simulate $(y_t^2, z_t^2), \dots, (y_t^L, z_t^L)$ using the scramble. This hybrid MC-qMC approach allows to sample from intractable distributions while retaining some of the the features of qMC integration.

The resulting Algorithm 6 combines elements from Algorithms 4 and 5. It requires an additional loop compared to the latter, which is more computationally demanding. Because the estimation combines MC with qMC, the variance of the estimates will typically be greater than a qMC only approach. Note that once the $(z_t^1(\theta), y_t^1(\theta))$ are drawn by MC simulations, the remaining $(z_t^\ell(\theta), y_t^\ell(\theta))$ can be simulated in parallel which might be computationally attractive.

Proposition A5 provides asymptotic results for Algorithm 6 with conditions similar to Duffie & Singleton (1993) but assuming bounded moments. Relaxing this assumption would

[10] When simulating the initial draw with the Rosenblatt transform is not possible, one may consider using a fixed starting value and a *burn-in* period assuming some decay conditions hold. This is only considered for the hybrid MC-qMC method, theoretical investigations for qMC-only draws is left to future research.

Algorithm 6 Scrambled Method of Moments for Dynamic Models - Hybrid MC-qMC Approach

Draw a random sequence $u_t^1 \overset{iid}{\sim} \mathcal{U}_{[0,1]^d}, t = 1, \dots, T \times S$
Set $(y_0^s, z_0^s) = (y_0, z_0)$, a fixed initial value
Simulate: $z_t^1(\theta) = g_{latent}(z_{t-1}^1, u_t^1; \theta)$ and $y_t^1(\theta) = g_{obs}(y_{t-1}^1(\theta), z_t^1(\theta); \theta)$
Draw a scrambled sequence $\tilde{u}_t = (\tilde{u}_t^2, \dots, \tilde{u}_t^L) \in [0,1]^{d \times (L-1)}$, $t = 1, \dots, T \times S$
Simulate: $\tilde{z}_t^\ell(\theta) = g_{latent}(\tilde{z}_t^{\ell-1}, \tilde{u}_t^\ell; \theta)$ and $\tilde{y}_t^\ell(\theta) = g_{obs}(\tilde{y}_t^{\ell-1}(\theta), \tilde{z}_t^\ell(\theta); \theta)$ for $\ell = 2, \dots, L$
Compute: $\hat{\psi}_T^S(\theta) = \frac{1}{T \times S} \sum_{t=1}^{T \times S} \psi(\tilde{y}_t^L(\theta), \dots, \tilde{y}_t^1(\theta))$
Find: $\hat{\theta}_T^S = \text{argmin}_{\theta \in \Theta} \|\hat{\psi}_T - \hat{\psi}_T^S(\theta)\|_{W_T}$

require to extend existing CLTs for dependent heterogeneous arrays (see e.g. White, 1984, Theorme 5.10) which goes beyond the scope of this paper. The simulations in Section 4.1.3 suggest that the estimator performs well with unbounded moments in practice.

3.3 Computing Standard Errors for the Simulated and Scrambled Method of Moments

Given that the scramble is different from standard Monte-Carlo methods, the following shows how to compute standard errors for $\hat{\theta}_n^S$ for SMM, antithetic draws and the scramble.

As shown in the literature for SMM and in the Appendix for the scramble, under regularity conditions, the estimators satisfy the following asymptotic expansion:

$$\hat{\theta}_n^S - \theta_0 = -\left(G'W_nG\right)^{-1} G'W_n \left[\hat{\psi}_n - \hat{\psi}_n^S(\theta_0)\right] + o_p(n^{-1/2}),$$

where $G = \partial_\theta \mathbb{E}\left[\hat{\psi}_n^S(\theta_0)\right]$, the usual Jacobian matrix. Under a Central Limit Theorem, the asymptotic variance is given by the usual sandwich formula. Given that W_n is chosen by the user, only two terms need to be approximated: the Jacobian G and the asymptotic variance of $[\hat{\psi}_n - \hat{\psi}_n^S(\theta_0)]$.

When the moments are smooth, the plug-in Jacobian estimator $\hat{G}_n = \partial_\theta \hat{\psi}_n^S(\hat{\theta}_n^S)$ is consistent for G under a ULLN. For non-smooth moments, there are several possibilities. The more computationally demanding approach is to Bootstrap the estimator $\hat{\theta}_n^S$ directly. Alternatively, Bruins et al. (2018) propose to smooth the draws $y_{i,t}^s$ in dynamic discrete choice models using a kernel; this transforms non-smooth and unbiased into smooth but biased simulated moments. Frazier et al. (2019) rely on a change of variable argument to compute analytical Jacobians in a class of discrete choice models. The quasi-Jacobian matrix in Forneron (2019) smoothes the moments themselves to consistently approximate G. It is also possible to use MCMC methods to sample from a quasi-posterior distributions which approximates the frequentist distribution of $\hat{\theta}_n^S$ (see e.g. Chernozhukov & Hong, 2003; Wood, 2010).

For cross-sections and short panels, the asymptotic variance of $[\hat{\psi}_n - \hat{\psi}_n^S(\theta_0)]$ in SMM can be approximated with the cross-sectional variance of $[\psi(y_i, x_i) - \frac{1}{S}\sum_{s=1}^S \psi(y_i^s(\hat{\theta}_n^S), x_i)]$. Pooling all the simulated samples that way ensures that the estimator is consistent for both standard and antithetic draws.[11] For time-series, under appropriate conditions, a HAC estimator is consistent for the long-run variance of $\hat{\psi}_T$ and the averaged $\hat{\psi}_T^S(\theta_n^S)$ respectively. Computing the long-run variance for the averaged $\sum_{s=1}^S \psi(y_t^s, \dots, y_{t-L}^s)/S$ ensures that the estimate is consistent for both standard and antithetic draws. As before, an estimate for the non-averaged moment may not be consistent for antithetic draws because of the dependence between simulated moments.

For the Scrambled Method of Moments, the variance should not be computed as above because scrambled draws are not independent from one another. Theorem 1 implies that the asymptotic variance only involves $\hat{\psi}_n$ in most cases; because simulation noise is asymptotically negligible.[12] One approach is to only compute the variance of $\hat{\psi}_n$. However, as illustrated in Section 4, even though the simulation noise is small in finite samples, it may not be completely negligible for some DGPs. In these cases, one would want to account for the variance due to $\hat{\psi}_n^S$. As discussed in Section 2.3, to consistently estimates the variance of $\hat{\psi}_n^S$ one can evaluate $\hat{\psi}_n^S$ several times with different seeds for the scramble and compute the variance across these estimates.

4 Monte-Carlo Illustrations

The following illustrates the finite sample properties of the Scrambled Method of Moments and Scrambled Indirect Inference in several simple settings and one example drawn from the heterogeneous agents literature. All simulations were carried out in R and C++ using the *Rcpp* package . Scrambled sequences were generated using the *fOptions* package.

[11] Another approach is to use use the variance of $\psi(y_i^s(\hat{\theta}_n^S), x_i)$ divided by S as an estimate for $\hat{\psi}_n^S(\theta_0)$. Although commonly used, this may actually not be consistent in the presence of antithetic draws. Depending on the correlation described in Section 2.1 it may either under or over-estimate the variance.

[12] See e.g. Proposition A1 in the Appendix.

4.1 Simple Examples

4.1.1 Mean-Variance

The first example, drawn from Gouriéroux et al. (1993), considers the estimation of a sample mean and variance of for an *iid* Gaussian sample:

$$y_i = \mu + \sigma e_i, \quad e_i \sim \mathcal{N}(0,1).$$

This example illustrates Algorithms 1, 2 and Proposition A3. As in the original paper, the auxiliary parameters $\hat{\psi}_n$ are the sample mean and variance of $(y_1, \dots, y_n)$:

$$\hat{\psi}_n = (\hat{\mu}_n, \hat{\sigma}_n^2)' = \frac{1}{n}\sum_{i=1}^{n}(y_i, [y_i - \hat{\mu}_n]^2)'.$$

In the 5,000 Monte-Carlo replications, the sample size is $n = 100$ and $\theta_0 = (\mu_0, \sigma_0^2) = (0,1)$. The number of simulated samples is $S = 1, 2, 4$ and 20. For SMM, e_i^s is drawn using the random number generator *rnorm* in *R* and antithetic draws are generated for $S = 2, 4$ and 20 by taking $e_i^{s+S/2} = -e_i^s$ for each $s = 1, \dots, S/2$. The *fOptions* package generates the scrambled Gaussian shocks directly. Table 1 summarizes the biases and standard deviations of the estimators.

Table 1: Mean and Variance Estimation

	MM	SMM				Antithetic				Scramble			
coef./S		1	2	4	20	1	2	4	20	1	2	4	20
$\sqrt{n} \times \text{std}(\hat{\mu})$	0.99	1.44	1.22	1.10	1.01	-	0.99	1.00	0.98	1.00	1.00	1.00	1.00
$\sqrt{n} \times \text{std}(\hat{\sigma}^2)$	1.41	2.07	1.76	1.60	1.47	-	2.03	1.75	1.50	1.44	1.44	1.41	1.41
$100 \times \text{bias}(\hat{\sigma}^2)$	-0.93	2.38	0.89	0.49	0.25	-	1.93	0.98	0.44	-0.43	-0.87	-1.07	-0.80

Because it has no simulation noise, the Method of Moments (MM) estimator has the smallest variance. SMM has a bias correction property for $\hat{\sigma}^2$ (Gouriéroux et al., 1993). For $\hat{\mu}_n$, antithetic draws and the scramble perform equally well for $S = 2$ and $S = 1$, respectively. For $\hat{\sigma}_n^2$, antithetic draws perform worse than SMM and the scramble. This is in line with the discussion in Section 2.1. The scramble performs similarly to the MM while SMM requires $S = 20$ to perform similarly. SMM and antithetic draws reduce the bias while the scramble does not. This reflects the fact that the scrambled $\hat{\psi}_n^S(\theta_0)$ approximates the asymptotic binding function $\psi_\infty(\theta_0) = \lim_{n\to\infty} \mathbb{E}(\hat{\psi}_n)$ while SMM and antithetic draws approximate the binding function $\psi(\theta_0) = \mathbb{E}(\hat{\psi}_n)$ which provides some finite sample bias correction.

4.1.2 Probit Model

The second example illustrates Algorithm 3 with non-smooth moments and covariates. The DGP is a Probit model:

$$y_i = \mathbb{1}\left\{\theta_1 + \theta_2 x_i \theta + e_i \geq 0\right\}, \quad e_i \overset{iid}{\sim} \mathcal{N}(0,1), \quad x_i \sim \mathcal{N}(0,1).$$

The moments $\hat{\psi}_n$ consist of the intercept and the slope in an OLS regression of y_i on x_i. In the $5,000$ Monte-Carlo replications, the sample size is $n = 1,000$ and $\theta_0 = (\theta_{1,0}, \theta_{2,0}) = (1,1)$. The number of simulated samples is $S = 1, 2, 4$ and 20. The standard deviations of the estimators are reported in Table 2.

Table 2: Probit Models: $\sqrt{n} \times \text{std}(\hat{\theta}_n^S)$

	SMM				Antithetic				Scramble			
coef./S	1	2	4	10	1	2	4	10	1	2	4	10
$\hat{\theta}_{1,n}$	2.38	2.24	2.11	1.91	-	2.17	2.06	1.91	2.14	2.09	2.01	1.90
$\hat{\theta}_{2,n}$	2.76	2.57	2.42	2.22	-	2.47	2.35	2.21	2.68	2.52	2.39	2.19

The Scrambled Method of Moments outperforms SMM for $S = 1$ and above. For $S \geq 2$, the scramble performs similarly to antithetic draws for estimating θ_0 and performs slightly better for θ_1. The gains are less substantial than in the previous example.

4.1.3 ARMA Model

To illustrate Algorithms 5 and 6 consider the following ARMA(1,1) model:[13]

$$y_t = \rho y_{t-1} + \sigma[e_t + \vartheta e_{t-1}], \quad e_t \overset{iid}{\sim} \mathcal{N}(0,1),$$

In the $5,000$ Monte-Carlo replications, the sample size is $T = 200$ and $\theta_0 = (\vartheta_0, \rho_0, \sigma_0^2) = (0.5, 0.5, 1)$. The number of simulated samples is $S = 1, 2$. The moments are the OLS coefficients from regressing y_t on its first $L = 4$ lags and the variance of the OLS residuals. Using auto-covariances as moments instead yields similar results.

Algorithm 5 requires sampling (y_t^1, e_t^1) from its stationary distribution directly. The marginals are known since $e_t^1 \sim \mathcal{N}(0,1)$ by assumption and $y_t^1 \sim \mathcal{N}(0, [1 + \vartheta^2 + 2\rho\vartheta]/[1 -$

[13] In the notation of Section A.2, the model can be written as: $y_t = \rho y_{t-1} + \sigma[z_{t,1} + \vartheta z_{t,2}]$, $(z_{t,1}, z_{t,2})' = (e_t, z_{t-1,1})'$, with $z_t = (z_{t,1}, z_{t,2})'$.

$\rho^2]\sigma^2)$. Since they are jointly Gaussian, it is sufficient to compute their covariance, $\text{cov}(e_t^1, y_t^1) = \rho\vartheta\sigma$, to find their joint distribution:

$$\begin{pmatrix} y_t^1 \\ e_t^1 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1+\vartheta^2+2\rho\vartheta}{1-\rho^2}\sigma^2 & \rho\vartheta\sigma \\ \rho\vartheta\sigma & 1 \end{pmatrix} \right).$$

Transforming independent bivariate scrambled Gaussian shocks into draws from the joint distribution above is then straightforward. For Algorithm 6, the (y_t^1, e_t^1) need to be sampled using MC methods. First, the initial value $(y_0^1, e_0^1) = (0, 0)$ is set and a path (y_t^1, e_t^1) is simulated with random MC draws. Once these (y_t^1, e_t^1) are simulated, the remaining $(y_t^2, e_t^2, \dots, y_t^5, e_t^5)$ are computed using scrambled Gaussian shocks.[14]

Table 3 compares the Maximum Likelihood Estimator (MLE) with SMM and antithetic draws, the qMC-only scramble from Algorithm 5 (reported in the Scramble column) and the hybrid MC-qMC scrambled from Algorithm 6 (reported in the Scramble-MC column).

Table 3: ARMA(1,1): $\sqrt{n} \times \text{std}(\hat{\theta}_n^S)$

	MLE	SMM		Antithetic		Scramble		Scramble-MC	
coef./S		1	2	1	2	1	2	1	2
$\hat{\rho}_n$	1.10	1.64	1.44	-	1.66	1.20	1.17	1.39	1.28
$\hat{\theta}_n$	1.13	1.86	1.57	-	1.87	1.33	1.28	1.53	1.36
$\hat{\sigma}_n$	0.72	1.05	0.90	-	1.04	0.76	0.72	0.94	0.84

MLE corresponds to the lower bound for variance of the estimators. The qMC-only scramble from Algorithm 5 outperforms SMM and antithetic draws. Antithetic draws perform worse than SMM using the same S which further illustrates the discussion in Section 2.1. The hybrid MC-qMC Algorithm 6 performs better than SMM and antithetic draws and, as expected, worse than the qMC-only approach.

[14] Here the number of lags used to compute the moments is $L = 5$ because y_t^5 is regressed on its 4 lags $y_t^4, \dots, y_t^1$.

4.2 An Income Process with "Lots of Heterogeneity"

The last example is a more substantial model borrowed from Browning et al. (2010).[15] Simulation-based estimation is commonly used in this heterogeneous agents literature due to the complexity and intractability of the models.[16] The baseline data generating process is an ARMA(1,1) at the individual level:

$$y_{i,t} = \delta_i \times \left([1-\omega_i^t] + \beta_i \times [1-\omega_i^{t-1}]\right) + \alpha_i\beta_i + \beta_i y_{i,t-1} + \alpha_i \times [1-\beta_i] \times t + \varepsilon_{i,t} + \theta_i \varepsilon_{i,t-1}$$

where the drift α_i, long-run mean δ_i, AR and MA coefficients β_i, θ_i as well as the persistence coefficient ω_i all vary at the individual level. The Gaussian shocks to log-income in the time dimension are denoted by ε_i while the Gaussian shocks to the ARMA coefficients $\alpha_i, \beta_i, \dots$ will be denoted by η_i. The initial value for log-income $y_{i,0}$ is drawn as:

$$y_{i,0} = \exp(\tau) \times \eta_{i,0}.$$

The heterogenous ARMA coefficients are then drawn using:

$$\begin{aligned}
\nu_{i,0} &= \exp(\phi_{11} + \phi_{12} \times y_{i,0} + \psi_{11} \times \eta_{i,1}) \\
\theta_i &= \text{logit}\left(\phi_{21} + \phi_{22} \times y_{i,0} + \psi_{21} \times \eta_{i,1} + \psi_{22} \times \eta_{i,2}\right) - 1/2 \\
\alpha_i &= \phi_{31} + \phi_{32} \times y_{i,0} + \psi_{3,1} \times \eta_{i,1} \\
\beta_i &= \text{logit}\left(\phi_{41} + \phi_{42} \times y_{i,0} + \psi_{4,1} \times \eta_{i,1} + \psi_{4,2} \times \eta_{i,2}\right) \\
\delta_i &= \phi_{51} + \phi_{52} \times y_{i,0} + \psi_{5,1} \times \eta_{i,1} + \psi_{5,2} \times \eta_{i,2} \\
\omega_i &= \text{logit}(\phi_{61} + \psi_{62} \times \eta_{i,2})
\end{aligned}$$

where logit is the usual logistic transformation $\text{logit}(x) = 1/[1+\exp(-x)]$. $\eta_{i,0}, \dots, \eta_{i,2} \overset{iid}{\sim} \mathcal{N}(0,1)$. For a discussion of the parameters and the role of the transformations, see Browning et al. (2010). $\nu_{i,0}$ is the initial value for the ARCH-type heteroskedasticity in the shocks $\varepsilon_{i,t}$:

$$\begin{aligned}
\sigma_{i,1}^2 &= \nu_{i,0}, & \varepsilon_{i,1} &= \sigma_{i,1} \times e_{i,1} \\
\sigma_{i,t}^2 &= \nu_{i,0} + \text{logit}(\varphi) \times \varepsilon_{i,t-1}^2, & \varepsilon_{i,t} &= \sigma_{i,t} \times e_{i,t}
\end{aligned}$$

where $e_{i,0}, \dots, e_{i,T} \overset{iid}{\sim} \mathcal{N}(0,1)$. In the simulations, the number of households is $n = 1,000$; the number of time periods is $T = 30$. As in the original paper, a burn-in period of $T_{burn} = 3$

[15] The data generating process considered here involves all the coefficients found in Browning et al. (2010), Table 2 minus the measurement errors and the time-trend in the ARCH component which are not considered in this Monte-Carlo exercise.

[16] See e.g. Guvenen (2011) for an overview of the computation and estimation of heterogeneous agents models.

periods is used to reduce the effect of the initial conditions. The parameter values are taken from Table 2 in Browning et al. (2010) and the moments are those described in their Appendix A.2 except the ones involving year of birth which are not considered in these simulations. In a nutshell, the moments involve the aggregation of individual-level OLS regression coefficients, moments of the residuals from these regressions, as well as autocorrelations and measures of social mobility.

The implementation of SMM is standard and described in Appendix A.4 of the original paper. For the scramble, a $(n \times S) \times (T + T_{burn} + 3) = (1,000 \times S) \times 36$ matrix of scrambled standard gaussian shocks is drawn. The integration dimension $d = 36$ is sufficiently large to illustrate the finite sample performance of the scrambled method of moments with a relatively large number of shocks. The first three dimensions, or columns of the matrix, correspond to $\eta_{i,0}, \dots, \eta_{i,2}$, the remaining dimensions correspond to time dimensions $e_{i,1}, \dots, e_{i,T+T_{burn}}$. The rows correspond to the cross-sectional dimension of the shocks, *i.e.* the $i = 1, \dots, n \times S$ index.

The results from the $2,000$ Monte-Carlo replications are presented in Table 4 for $S = 1, 2, 4$. SMM and antithetic draws are used as a benchmark for the scramble with either a large sample of $n \times S$ individuals (as in Algorithm 2 or S samples of n individuals (as in Algorithm 2). The scramble generally outperform SMM and antithetic draws. Antithetic draws either under or over-performs SMM depending on the parameter of interest which is in line with previous discussions. Both implementations of the Scrambled Method of Moments perform similarly. For some coefficients, there is little to no improvement in increasing S from 1 to 2 or 4. For most coefficients, the scramble with $S = 2$ outperforms SMM with $S = 4$.

Furthermore, using the same $S = 4$, the replications were computed about 15% faster for the scramble than SMM. Since the only difference between the two is the shocks used in the simulations, this reflects faster convergence of the optimizer. Possibly because the scramble are smoother (less noisy) than the MC moments which makes the objective function easier to minimize.

Table 4: Income Process with Heterogeneity: $\sqrt{n} \times \text{std}(\hat{\theta}_n^S)$

	SMM			Antithetic			Scramble					
							S samples of size n			1 sample of size nS		
coef./S	1	2	4	1	2	4	1	2	4	1	2	4
τ	1.27	1.20	1.20	-	1.25	1.35	1.06	1.12	1.11	1.06	1.13	1.13
ϕ_{11}	1.29	1.22	1.18	-	1.06	1.23	1.00	0.99	1.00	1.00	1.03	1.04
ϕ_{12}	4.32	3.70	3.28	-	3.70	4.47	3.39	3.07	2.95	3.39	3.09	3.20
ϕ_{21}	1.62	1.44	1.39	-	1.53	1.73	1.30	1.31	1.26	1.30	1.20	1.28
ϕ_{31}	0.18	0.15	0.14	-	0.14	0.15	0.15	0.14	0.13	0.15	0.14	0.14
ϕ_{32}	0.17	0.16	0.14	-	0.15	0.15	0.15	0.15	0.14	0.15	0.14	0.14
ϕ_{41}	1.90	1.66	1.68	-	1.72	1.84	1.60	1.57	1.54	1.60	1.60	1.59
ϕ_{51}	2.98	2.84	2.70	-	2.68	2.79	2.94	2.78	2.59	2.94	2.76	2.61
ϕ_{52}	8.54	8.17	7.58	-	7.75	7.73	8.00	7.59	7.23	8.00	7.54	7.31
ϕ_{61}	4.18	4.20	3.29	-	3.36	3.32	3.27	3.16	2.87	3.27	3.01	3.04
ψ_{11}	1.25	1.29	1.20	-	1.29	1.28	0.89	0.88	0.96	0.98	0.97	1.01
ψ_{22}	2.55	2.42	2.39	-	2.47	2.89	2.13	2.08	2.15	2.13	2.09	2.21
ψ_{31}	0.08	0.07	0.07	-	0.07	0.07	0.06	0.07	0.06	0.06	0.06	0.06
ψ_{41}	2.68	2.36	2.30	-	2.40	2.61	2.27	2.22	2.10	2.27	2.22	2.19
ψ_{42}	1.90	1.76	1.74	-	1.75	1.87	1.56	1.53	1.55	1.56	1.63	1.58
ψ_{51}	3.30	3.13	2.83	-	2.90	3.28	2.64	2.89	2.56	2.64	2.64	2.70
ψ_{52}	2.14	2.03	2.20	-	2.14	2.45	1.78	1.95	1.83	1.78	1.84	1.96
ψ_{62}	3.53	3.53	3.65	-	3.55	4.01	3.34	3.28	3.36	3.34	3.44	3.56
φ	2.24	2.10	2.39	-	2.76	3.06	1.99	1.94	2.21	1.99	2.30	2.56

5 Conclusion

This paper proposes several Algorithms that implement Owen's scramble in a simulation-based estimation framework. Since the method is designed for computing integrals of *iid* sequences, some care is needed when simulating data with covariates or time-series. Large sample results are provided in Appendix A to support the proposed Algorithms. The results for dynamic models could be extended to non-smooth bounded moments through additional stochastic equicontinuity results using the inequality in Andrews & Pollard (1994), for instance. The simulations illustrate the finite performance of the Scrambled Methods of Moments and Scrambled Indirect Inference compared to other commonly used methods. The last example suggests the scramble could be useful in larger scale problems found in the heterogenous agents literature where SMM is commonly used.

References

Andrews, D. W. K. & Pollard, D. (1994). An Introduction to Functional Central Limit Theorems for Dependent Stochastic Processes. *International Statistical Review / Revue Internationale de Statistique*, 62(1), 119.

Browning, M., Ejrnaes, M., & Alvarez, J. (2010). Modelling Income Processes with Lots of Heterogeneity. *Review of Economic Studies*, 77(4), 1353–1381.

Bruins, M., Duffy, J. A., Keane, M. P., & Smith, A. A. (2018). Generalized indirect inference for discrete choice models. *Journal of Econometrics*, 205(1), 177–203.

Buchholz, A. & Chopin, N. (2017). Improving approximate Bayesian computation via quasi-Monte Carlo.

Chernozhukov, V. & Hong, H. (2003). An MCMC approach to classical estimation. *Journal of Econometrics*, 115(2), 293–346.

Davis, R. A., Sousa, T. d. R., & Klüppelberg, C. (2019). Indirect Inference for Time Series Using the Empirical Characteristic Function and Control Variates.

Davydov, Y. A. (1968). Convergence of Distributions Generated by Stationary Stochastic Processes. *Theory of Probability & Its Applications*, 13(4), 691–696.

Dick, J. & Pillichshammer, F. (2010). *Digital Nets and Sequences*. Cambridge: Cambridge University Press.

Duffie, D. & Singleton, K. J. (1993). Simulated Moments Estimation of Markov Models of Asset Prices. *Econometrica*, 61(4), 929.

Forneron, J.-J. (2019). Detecting Identification Failure in Moment Condition Models.

Forneron, J.-J. & Ng, S. (2018). The ABC of simulation estimation with auxiliary statistics. *Journal of Econometrics*, 205(1).

Frazier, D. T., Oka, T., & Zhu, D. (2019). Indirect Inference with a Non-Smooth Criterion Function. *Forthcoming in the Journal of Econometrics*.

Gallant, a. R. & Tauchen, G. (1996). Which Moments to Match? *Econometric Theory*, 12(04), 657.

Gerber, M. & Chopin, N. (2015). Sequential quasi Monte Carlo. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(3), 509–579.

Gerber, M. & Chopin, N. (2017). Convergence of sequential quasi-Monte Carlo smoothing algorithms. *Bernoulli*, 23(4B), 2951–2987.

Gouriéroux, C., Monfort, A., & Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8(S1), S85—-S118.

Guvenen, F. (2011). Macroeconomics with Heterogeneity: A Practical Guide. *Federal Re-*

serve Bank of Richmond Economic Quarterly, 97(3), 255–327.

Hickernell, F. J. (1996). The mean square discrepancy of randomized nets. *ACM Transactions on Modeling and Computer Simulation*, 6(4), 274–296.

Hickernell, F. J., Lemieux, C., & Owen, A. B. (2005). Control Variates for Quasi-Monte Carlo. *Statistical Science*, 20(1), 1–31.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, 40(2), 633–643.

Judd, K. L. (1998). *Numerical methods in economics*. MIT press.

Lemieux, C. (2009). *Monte Carlo and Quasi-Monte Carlo Sampling*. Springer Series in Statistics. New York, NY: Springer New York.

Mak, S. & Joseph, V. R. (2018). Support points. *The Annals of Statistics*, 46(6A), 2562–2592.

Marin, J. M., Pudlo, P., Robert, C. P., & Ryder, R. J. (2012). Approximate Bayesian computational methods. *Statistics and Computing*, 22(6), 1167–1180.

Matoušek, J. (1998). On the L2-Discrepancy for Anchored Boxes. *Journal of Complexity*, 14(4), 527–556.

McFadden, D. (1989). A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration. *Econometrica*, 57(5), 995.

Newey, W. K. & McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, 4, 2111–2245.

Oates, C. J., Girolami, M., & Chopin, N. (2017). Control functionals for Monte Carlo integration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3), 695–718.

Ökten, G., Tuffin, B., & Burago, V. (2006). A central limit theorem and improved error bounds for a hybrid-Monte Carlo sequence with applications in computational finance. *Journal of Complexity*, 22(4), 435–458.

Owen, A. B. (1997). Scrambled net variance for integrals of smooth functions. *The Annals of Statistics*, 25(4), 1541–1562.

Owen, A. B. & Tribble, S. D. (2005). A quasi-Monte Carlo Metropolis algorithm. *Proceedings of the National Academy of Sciences*, 102(25), 8844–8849.

Pakes, A. & Pollard, D. (1989). Simulation and the Asymptotics of Optimization Estimators. *Econometrica*, 57(5), 1027.

Paskov, S. H. & Traub, J. F. (1995). Faster Valuation of Financial Derivatives. *The Journal of Portfolio Management*, 22(1), 113–123.

Robert, C. & Casella, G. (2004). Monte Carlo Statistical Methods. *Textbooks in Statistics,* *second edn, Springer*.

Roth, K. F. (1954). On irregularities of distribution. *Mathematika*, 1(2), 73–79.

Schmidt, W. M. (1970). Irregularities of distribution. V. *Proceedings of the American Mathematical Society*, 25(3), 608–608.

Smith, A. A. (2008). Indirect Inference. *in S. Durlauf L. Blume (eds), The New Palgrave Dictionary of Economics*, 2.

van der Vaart, A. W. & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics. New York, NY: Springer New York.

White, H. (1984). *Asymptotic Theory for Econometricians*. Elsevier.

Wood, S. N. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466(7310), 1102–1104.

Appendix A Asymptotic Theory

In the following $\hat{\theta}_n^S$ and $\hat{\theta}_T^S$ will denote the scrambled estimator for static and dynamic models respectively.

Assumption A1 (Identification, Regularity, Sample Moments)**.** *Suppose the following holds:*

i. (Identification) $\mathbb{E}[\hat{\psi}_n] = \mathbb{E}[\hat{\psi}_n^S(\theta)] \Leftrightarrow \theta = \theta_0$.

ii. (Regularity) $\theta_0 \in$ *interior*(Θ) *where* Θ *is a compact and convex subset of* $\mathbb{R}^{d_\theta}$, $1 \leq d_\theta < +\infty$ *fixed.* $\mathbb{E}[\hat{\psi}_n^S(\cdot)]$ *is continuously differentiable around* θ_0 *and* $\partial_\theta \mathbb{E}[\hat{\psi}_n^S(\theta_0)]$ *has full rank.*

iii. (Sample Moments) $\hat{\psi}_n$ *satisfies a Law of Large Numbers and a Central Limit Theorem:*

$$\sqrt{n}\left[\hat{\psi}_n - \mathbb{E}(\hat{\psi}_n)\right] \xrightarrow{d} \mathcal{N}(0, V).$$

iv. (Weighting Matrix) $W_n \xrightarrow{p} W$ *positive definite*

A.1 Static Models

To simplify notation, let:

$$\tilde{\psi}(x_i, \theta, u_i) \overset{def}{=} \psi(g(x_i, \theta, u_i), x_i)$$

A.1.1 Results for Estimation with Smooth Moments and no Covariates

Assumption A2 (Scrambled Smooth Moments without Covariates)**.** *Suppose that the following holds:*

i. For all $\theta \in \Theta$,

$$\mathbb{E}\left(\left\|\tilde{\psi}(\theta, u_i)\right\|^2\right) < +\infty$$

and

$$\|\tilde{\psi}(\theta_1, u_i) - \tilde{\psi}(\theta_2, u_i)\| \leq C_1(u_i) \times \|\theta_1 - \theta_2\|,$$

where $\mathbb{E}[C_1(u_i)^2] < +\infty$.

ii. For all $\theta \in \Theta$, $\tilde{\psi}$ *is continuously differentiable in* θ *around* θ_0 *and:*

$$\mathbb{E}\left(\left\|\partial_\theta \tilde{\psi}(\theta, u_i)\right\|^2\right) < +\infty,$$

and

$$\|\partial_\theta \tilde{\psi}(\theta_1, u_i) - \partial_\theta \tilde{\psi}(\theta_2, u_i)\| \leq C_2(u_i) \times \|\theta_1 - \theta_2\|,$$

where $\mathbb{E}[C_2(u_i)^2] < +\infty$.

Proposition A1 (Consistency and Asymptotic Normality without Covariates)**.** *Suppose Assumptions A1 and A2 hold, then* $\hat{\theta}_n^S \xrightarrow{p} \theta_0$ *and*

$$\sqrt{n}\left(\hat{\theta}_n^S - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

where

$$\Sigma = (G'WG)^{-1} G'WVWG\, (G'WG)^{-1},$$

$G = \partial_\theta \mathbb{E}[\hat{\psi}_n^S(\theta_0)]$, $V = \lim_{n\to\infty} n \times var(\hat{\psi}_n)$.

A.1.2 Results for Estimation with Non-Smooth Moments and Covariates

Assumption A3 (Scrambled Non-Smooth Moments with Covariates)**.** *Suppose that for some* $\delta > 0$ *the following holds:*

i. $\mathbb{E}\left[var\left(\psi(y_i, x_i) - \tilde{\psi}(x_i, u_i; \theta_0)|u_i\right)\right]$ *is positive definite and finite, also* $\mathbb{E}\left[\|var\left(\psi(y_i, x_i) - \tilde{\psi}(x_i, u_i; \theta_0)|u_i\right)\|^{2+\delta}\right] < +\infty$.

ii. There exists an envelope function $\bar{\psi}$ *such that for all* $\theta \in \Theta$, $\|\tilde{\psi}(x_i, u_i; \theta)\| \leq \bar{\psi}(x_i, u_i)$ *with* $\mathbb{E}\left[var\left(\bar{\psi}(x_i, u_i)|u_i\right)\right] > 0$ *and* $\mathbb{E}\left[var\left(\bar{\psi}(x_i, u_i)|u_i\right)^{2+\delta}\right) < +\infty$.

iii. There exists $\tilde{C}_1(\cdot)$ *such that* $\theta_1, \theta_2 \in \Theta$, $\mathbb{E}(\|\tilde{\psi}(x_i, u_i; \theta_1) - \tilde{\psi}(x_i, u_i; \theta_2)\|^2|u_i) \leq \tilde{C}(u_i)^2 \times \|\theta_1 - \theta_2\|^2$ *with* $\mathbb{E}\left(\tilde{C}(u_i)^4\right) < +\infty$.

iv. $\mathbb{E}(\tilde{\psi}(x_i, u_i; \cdot)|u_i)$ *is continuously differentiable in* $\theta \in \Theta$, u_i *almost surely. There exists* $\tilde{C}_2(\cdot)$ *such that for all* $\theta_1, \theta_2 \in \Theta$, $\|\mathbb{E}\left(\tilde{\psi}(x_i, u_i; \theta_1) - \tilde{\psi}(x_i, u_i; \theta_2)|u_i\right) - \partial_\theta \mathbb{E}\left(\tilde{\psi}(x_i, u_i; \theta_2)|u_i\right)(\theta_1 - \theta_2)\| \leq \tilde{C}_2(u_i) \times \|\theta_1 - \theta_2\|^2$. *There exists* $\tilde{C}_3(\cdot)$ *such that for all* $\theta \in \Theta$, $\mathbb{E}\left[\|\partial_\theta \mathbb{E}\left(\tilde{\psi}(x_i, u_i; \theta)|u_i\right)\|^2\right] < +\infty$, $\|\partial_\theta \mathbb{E}\left[\tilde{\psi}(\theta_1, u_i)|u_i\right] - \partial_\theta \mathbb{E}\left[\tilde{\psi}(\theta_2, u_i)|u_i\right]\| \leq \tilde{C}_3(u_i) \times \|\theta_1 - \theta_2\|$, *where* $\mathbb{E}[\tilde{C}_3(u_i)^2] < +\infty$.

Proposition A2 (Consistency and Asymptotic Normality with Covariates)**.** *For* $S \geq 1$, *suppose that* $\|\hat{\psi}_n - \hat{\psi}_n^S(\hat{\theta}_n^S)\|_{W_n} \leq o_p(n^{-1/2})$ *and that Assumptions A1, A3 hold then* $\hat{\theta}_n^S \xrightarrow{p} \theta_0$ *and*

$$\sqrt{n}\left(\hat{\theta}_n^S - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

where

$$\Sigma = (G'WG)^{-1} G'W\tilde{V}WG (G'WG)^{-1},$$

$G = \partial_\theta \mathbb{E}[\hat{\psi}_n^S(\theta_0)]$, $\tilde{V} = \mathbb{E}\left[var\Big(\psi(y_i, x_i) - \tilde{\psi}(x_i, u_i; \theta_0)|u_i\Big)\right]$.

A.1.3 Results for the Scrambled Indirect Inference Estimator

The following extends the results from Proposition A1 to the Indirect Inference estimator of Gouriéroux et al. (1993). The moments $\hat{\psi}_n, \hat{\psi}_n^S(\theta)$ are now defined as solutions to the sample and simulated M-estimation problems:

$$\hat{\psi}_n = \text{argmin}_{\psi\in\Psi} M_n(\psi), \qquad \text{where } M_n(\psi) = \frac{1}{n}\sum_{i=1}^n m(y_i;\psi)$$
$$\hat{\psi}_n^S(\theta) = \text{argmin}_{\psi\in\Psi} M_n^S(\theta;\psi), \quad \text{where } M_n^S(\theta;\psi) = \frac{1}{nS}\sum_{i=1}^{nS} m(y_i^s(\theta);\psi).$$

Again, to simplify notation consider:

$$\tilde{m}(u_i, \theta; \psi) \stackrel{def}{=} m(y_i^s(\theta); \psi).$$

As in Gouriéroux et al. (1993), the binding function $\psi_\infty(.\cdot)$ is defined as:

$$\psi_\infty(\theta) \stackrel{def}{=} \text{argmin}_{\psi\in\Psi} \mathbb{E}\left[\tilde{m}(u_i, \theta; \psi)\right].$$

Assumption A4 (Scrambled Indirect Inference)**.** *Suppose that the following holds:*

i. The mapping $\theta \to \psi_\infty(\theta) \in \Psi$ is continuous differentiable and injective. Ψ is a compact and convex subset of $\mathbb{R}^{d_\psi}$, finite-dimensional and $\psi_\infty(\theta_0) \in interior(\Psi)$.

ii. For all $(\theta, \psi) \in \Theta \times \Psi$,

$$\mathbb{E}\left[\|\tilde{m}(u_i, \theta; \psi)\|^2\right] < +\infty,$$

and there exists $C_1(\cdot,\cdot)$ such that for all $\theta \in \Theta$ and $\psi_1, \psi_2 \in \Psi$:

$$\|\tilde{m}(u_i, \theta; \psi_1) - \tilde{m}(u_i, \theta; \psi_2)\| \leq C_1(u_i, \theta) \times \|\psi_1 - \psi_2\|,$$

with $\mathbb{E}[C_1(u_i, \theta)^2] < +\infty$ for all $\theta \in \Theta$.

iii. $\tilde{m}$ is twice continuously differentiable in (θ, ψ), u_i almost surely. For all $(\theta, \psi) \in \Theta \times \Psi$,

$$\mathbb{E}\left[\|\partial_\psi \tilde{m}(u_i, \theta; \psi)\|^2\right] < +\infty, \quad \mathbb{E}\left[\|\partial^2_{\psi,\psi'} \tilde{m}(u_i, \theta; \psi)\|^2\right] < +\infty, \quad \mathbb{E}\left[\|\partial^2_{\psi,\theta'} \tilde{m}(u_i, \theta; \psi)\|^2\right] < +\infty,$$

and there exists $C_2(\cdot), C_3(\cdot), C_4(\cdot)$ such that for all $\theta_1, \theta_2 \in \Theta$ and $\psi_1, \psi_2 \in \Psi$:

$$\|\partial_\psi \tilde{m}(u_i, \theta_1; \psi_1) - \partial_\psi \tilde{m}(u_i, \theta_2; \psi_2)\| \leq C_2(u_i) \times (\|\theta_1 - \theta_2\| + \|\psi_1 - \psi_2\|),$$
$$\|\partial^2_{\psi,\psi'} \tilde{m}(u_i, \theta_1; \psi_1) - \partial^2_{\psi,\psi'} \tilde{m}(u_i, \theta_2; \psi_2)\| \leq C_3(u_i) \times (\|\theta_1 - \theta_2\| + \|\psi_1 - \psi_2\|),$$
$$\|\partial^2_{\psi,\theta'} \tilde{m}(u_i, \theta_1; \psi_1) - \partial^2_{\psi,\theta'} \tilde{m}(u_i, \theta_2; \psi_2)\| \leq C_4(u_i) \times (\|\theta_1 - \theta_2\| + \|\psi_1 - \psi_2\|),$$

with $\mathbb{E}[C_2(u_i)^2]$, $\mathbb{E}[C_3(u_i)^2]$ and $\mathbb{E}[C_4(u_i)^2] < +\infty$.

iv. The Hessian $\partial^2_{\psi,\psi'} \mathbb{E}[\tilde{m}(u_i, \theta; \psi)]$ is positive definite for all $\theta \in \Theta$ and all $\psi \in \Psi$ with

$$0 < \inf_{(\theta,\psi)\in\Theta\times\Psi} \lambda_{\min}(\partial^2_{\psi,\psi'} \mathbb{E}[\tilde{m}(u_i, \theta; \psi)]) \leq \sup_{(\theta,\psi)\in\Theta\times\Psi} \lambda_{\max}(\partial^2_{\psi,\psi'} \mathbb{E}[\tilde{m}(u_i, \theta; \psi)]) < +\infty.$$

Also, $\sup_{(\theta,\psi)\in\Theta\times\Psi} \|\partial^2_{\psi,\theta'} \mathbb{E}[\tilde{m}(u_i, \theta; \psi)])\| < +\infty$.

Proposition A3. *Consistency and Asymptotic Normality with Auxiliary Parameters Suppose Assumption A1 and A4 hold, then $\hat{\theta}^S_n \xrightarrow{p} \theta_0$ and*

$$\sqrt{n}\left(\hat{\theta}^S_n - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

where

$$\Sigma = (G'WG)^{-1} G'WVWG\,(G'WG)^{-1},$$

$G = \partial_\theta \mathbb{E}[\hat{\psi}^S_n(\theta_0)]$, $V = \lim_{n\to\infty} n \times var(\hat{\psi}_n)$.

A.2 Dynamic Models

A.2.1 qMC-only Estimator

For simplicity, write:

$$\hat{\psi}^S_T(\theta) = \frac{1}{TS} \sum_{t=1}^{TS} \tilde{\psi}(u_t; \theta),$$

where u_t has the appropriate dimension d given in Section 3.2.1.

Proposition A4 (Consistency and Asymptotic Normality - qMC only)**.** *Suppose Assumptions A1 and A2 hold and the draws are generate as in Algorithm 5 then $\hat{\theta}^S_n \xrightarrow{p} \theta_0$ and:*

$$\sqrt{T}\left(\hat{\theta}^S_T - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

where

$$\Sigma = (G'WG)^{-1} G'WVWG\,(G'WG)^{-1},$$

$G = \partial_\theta \mathbb{E}\left[\hat{\psi}^S_T(\theta_0)\right]$, $V = \lim_{T\to\infty} T \times var(\hat{\psi}_T)$.

A.2.2 MC-qMC Hybrid Estimator

For simplicity, write:

$$\hat{\psi}_T^S(\theta) = \frac{1}{TS} \sum_{t=1}^{TS} \tilde{\psi}(y_t^1, z_t^1, u_t; \theta),$$

where y_t^1, z_t^1 are simulated using MC methods as in Algorithm 6.

Assumption A5 (Dynamic Models - MC-qMC)**.** *Suppose there exists a constant $K > 0$ such that:*

i. *For all $\theta \in \Theta$, (y_t^1, z_t^1) is geometrically ergodic: $\|f_t(y_t^1, z_t^1; \theta) - f_\infty(y_t^\infty, z_t^\infty; \theta)\|_{TV} \leq C_1 \times \rho^t$, for some $\rho \in [0, 1)$ and $0 \leq C_1 < +\infty$, where f_∞ is the ergodic distribution of y_t^1, z_t^1 and f_t its non-stationary distribution with fixed starting value.*

ii. *For all $\theta \in \Theta$, $\mathbb{E}[\|\tilde{\psi}(y_t^1, z_t^1, u_t; \theta)\|^4 | u_t] \leq K < +\infty$, $\mathbb{E}[\|\tilde{\psi}(y_t^\infty, z_t^\infty, u_t; \theta)\|^4 | u_t] \leq K < +\infty$.*

iii. *For any $\|\theta_1 - \theta_2\|$ small, $\|\tilde{\psi}(y_t^1, z_t^1, u_t; \theta_1) - \tilde{\psi}(y_t^1, z_t^1, u_t; \theta_2)\| \leq C_2(y_t^1, z_t^1, u_t; \theta_1) \times \|\theta_1 - \theta_2\|$ with $\mathbb{E}[\|C_2(y_t^1, z_t^1, u_t; \theta_1)\|^4 | u_t] \leq K < +\infty$ and $\mathbb{E}[\|C_2(y_t^\infty, z_t^\infty, u_t; \theta_1)\|^4 | u_t] \leq K < +\infty$.*

iv. *For all $\theta \in \Theta$, $\mathbb{E}[\|\partial_\theta \tilde{\psi}(y_t^1, z_t^1, u_t; \theta)\|^4 | u_t] \leq K < +\infty$, $\mathbb{E}[\|\partial_\theta \tilde{\psi}(y_t^\infty, z_t^\infty, u_t; \theta)\|^4 | u_t] \leq K < +\infty$.*

v. *For any $\|\theta_1 - \theta_2\|$ small, $\|\partial_\theta \tilde{\psi}(y_t^1, z_t^1, u_t; \theta_1) - \partial_\theta \tilde{\psi}(y_t^1, z_t^1, u_t; \theta_2)\| \leq C_3(y_t^1, z_t^1, u_t; \theta_1) \times \|\theta_1 - \theta_2\|$ with $\mathbb{E}[\|C_3(y_t^1, z_t^1, u_t; \theta_1)\|^4 | u_t] \leq K < +\infty$ and $\mathbb{E}[\|C_3(y_t^\infty, z_t^\infty, u_t; \theta_1)\|^4 | u_t] \leq K < +\infty$.*

vi. *$\lim_{T \to \infty} T \times var(\hat{\psi}_T^S(\theta_0) | u_1, \ldots, u_{TS})$ is positive definite and finite.*

Proposition A5 (Consistency and Asymptotic Normality - MC-qMC)**.** *Suppose Assumptions A1 and A5 hold, then $\hat{\theta}_T^S \xrightarrow{p} \theta_0$ and*

$$\sqrt{T}(\hat{\theta}_T^S - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

where

$$\Sigma = (G'WG)^{-1} G'WVWG (G'WG)^{-1},$$

$G = \lim_{T \to \infty} \mathbb{E}[\partial_\theta \hat{\psi}_T^S(\theta_0)]$, $V = \lim_{T \to \infty} T \times var(\hat{\psi}_T^S(\theta_0) | u_1, \ldots, u_{TS})$.

Appendix B Proofs

B.1 Static Models

B.1.1 Smooth Moments with No Covariates

Lemma B1 (ULLN and CLT for Smooth Moments without Covariates)**.** *Suppose the conditions in Assumption A2 hold, then:*

i. $\sup_{\theta\in\Theta}\|\hat{\psi}_n^s(\theta)-\mathbb{E}(\hat{\psi}_n^s(\theta))\| = o_p(1)$,

ii. $\sup_{\theta\in\Theta}\|\partial_\theta\hat{\psi}_n^s(\theta)-\partial_\theta\mathbb{E}(\hat{\psi}_n^s(\theta))\| = o_p(1)$,

iii. $\|\hat{\psi}_n^s(\theta_0)-\mathbb{E}[\hat{\psi}_n^s(\theta_0)]\| = o_p(n^{-1/2})$.

Proof of Lemma B1.
Part i. ULLN for $\hat{\psi}_n^S(\theta)$
Assumption A2 implies $\|\hat{\psi}_n^S(\theta)-\mathbb{E}[\hat{\psi}_n^S(\theta)]\| = o_p(1)$ pointwise. Using the same steps as in Jennrich (1969), the Lipschitz condition implies that for a finite cover $(\theta_1,\dots,\theta_J)$ of Θ:

$$\begin{aligned}&\sup_{\theta\in\Theta}\|\hat{\psi}_n^S(\theta)-\mathbb{E}(\hat{\psi}_n^S(\theta))\|\\ &\quad\le \max_{j\in\{1,\dots,J\}}\|\hat{\psi}_n^S(\theta_j)-\mathbb{E}(\hat{\psi}_n^S(\theta_j))\| + \sup_{\theta\in\Theta}\min_{j\in\{1,\dots,J\}}\Big\|[\hat{\psi}_n^S(\theta)-\hat{\psi}_n^S(\theta_j)]-[\mathbb{E}(\hat{\psi}_n^S(\theta))-\mathbb{E}(\hat{\psi}_n^S(\theta_j))]\Big\|.\end{aligned}$$

Using the Lipschitz condition for $\tilde{\psi}$, we have:

$$\begin{aligned}&\sup_{\theta\in\Theta}\min_{j\in\{1,\dots J\}}\Big\|[\hat{\psi}_n^S(\theta)-\hat{\psi}_n^S(\theta_j)]-[\mathbb{E}(\hat{\psi}_n^S(\theta))-\mathbb{E}(\hat{\psi}_n^S(\theta_j))]\Big\|\\ &\quad\le\Big[\frac{1}{n}\sum_{i=1}^n C_1(u_i)+\mathbb{E}(C_1(u_i))\Big]\times\sup_{\theta\in\Theta}\min_{j\in\{1,\dots,J\}}\|\theta-\theta_j\|\end{aligned}$$

Since C_1 is square integrable, Theorem 1 applies to $C_1(u_i)$ so that $\Big[\frac{1}{n}\sum_{i=1}^n C_1(u_i)+\mathbb{E}(C_1(u_i))\Big] = 2\times\mathbb{E}[C_1(u_i)]+o_p(1)$. For $J\ge 1$ large enough and an appropriate cover, $\sup_{\theta\in\Theta}\min_{j\in\{1,\dots,J\}}\|\theta-\theta_j\|\le\frac{\varepsilon}{4\mathbb{E}[C_1(u_i)]}$. Similarly, for any given $J\ge 1$ fixed, $\max_{j\in\{1,\dots,J\}}\|\hat{\psi}_n^S(\theta_j)-\mathbb{E}(\hat{\psi}_n^S(\theta_j))\|\le\varepsilon/2$ with probability going to 1 as $n\to\infty$. Overall, this implies that:

$$\mathbb{P}(\sup_{\theta\in\Theta}\|\hat{\psi}_n^S(\theta)-\mathbb{E}(\hat{\psi}_n^S(\theta))\|>\varepsilon)\to 0,$$

this provides a ULLN with scrambled draws.

Part ii. ULLN for $\hat{\psi}_n^S(\theta)$
The ULLN can be directly applied to $\partial_\theta \hat{\psi}_n^s(\theta)$ under the stated assumptions.

Part iii. Convergence rate for $\hat{\psi}_n^S(\theta_0) - \mathbb{E}[\hat{\psi}_n^S(\theta_0)]$
This is a direct application of Theorem 1 which concludes the proof. $\square$

Proof of Proposition A1. Combining Assumption A1 with the ULLN in Lemma B1 imply that the consitency Theorem 2.1 in Newey & McFadden (1994) applies; *i.e.* $\hat{\theta}_n^S \overset{p}{\to} \theta_0$. Then, the ULLN for the Jacobian with a mean value expansion argument imply:

$$\begin{aligned}
\sqrt{n}\left(\hat{\theta}_n^S - \theta_0\right) &= -\left(G'WG\right)^{-1} G'W\sqrt{n}\left[\underbrace{\hat{\psi}_n - \mathbb{E}[\hat{\psi}_n^S(\theta_0)]}_{=O_p(n^{-1/2})} + \underbrace{\mathbb{E}[\hat{\psi}_n^S(\theta_0)] - \hat{\psi}_n^S(\theta_0)}_{=o_p(n^{-1/2})}\right] + o_p(1) \\
&= -\left(G'WG\right)^{-1} G'W\sqrt{n}\left[\hat{\psi}_n - \mathbb{E}[\hat{\psi}_n^S(\theta_0)]\right] + o_p(1) \\
&\overset{d}{\to} \mathcal{N}(0, \Sigma),
\end{aligned}$$

where Σ is defined in the Proposition. This concludes the proof. $\square$

B.1.2 Non-Smooth Moments with Covariates

Lemma B2 (Stochastic Equicontinuity and CLT with Covariates)**.** *Suppose that Assumptions A1 and A3 hold and $S = 1$, then:*

i. $\sqrt{n}\left[\hat{\psi}_n - \hat{\psi}_n^S(\theta_0)\right] \overset{d}{\to} \mathcal{N}(0, \tilde{V})$ *where* $\tilde{V} = \mathbb{E}\left[var\left(\psi(y_i, x_i) - \tilde{\psi}(x_i, u_i; \theta)|u_i\right)\right]$

ii. $\sup_{\|\theta_1-\theta_2\|\leq\delta_n} \sqrt{n}\|[\hat{\psi}_n^S(\theta_1) - \hat{\psi}_n^S(\theta_2)] - \mathbb{E}[\hat{\psi}_n^S(\theta_1) - \hat{\psi}_n^S(\theta_2)|u_1, \dots, u_n]\| = o_p(1)$, $\forall \delta_n \searrow 0$

iii. $\sup_{\|\theta_1-\theta_2\|\leq\delta_n} \|\mathbb{E}[\hat{\psi}_n^S(\theta_1) - \hat{\psi}_n^S(\theta_2)|u_1, \dots, u_n] - \partial_\theta \mathbb{E}[\hat{\psi}_n^S(\theta_2)](\theta_1 - \theta_2)\| \leq O_p(\delta_n^2)$, $\forall \delta_n \searrow 0$

Proof of Lemma B2.
Part i. CLT for $\hat{\psi}_n - \hat{\psi}_n^S(\theta_0)$
Similarly to Ökten et al. (2006), the main idea is to verify the conditions for an independent non-identically distributed CLT hold holding the qMC draws $u_1, \dots, u_n$ fixed. Note that:

$$\hat{\psi}_n - \hat{\psi}_n^S(\theta_0) = \underbrace{\hat{\psi}_n - \mathbb{E}[\hat{\psi}_n^S(\theta_0)|u_1, \dots, u_n]}_{\text{independent non-identically distributed}} + \underbrace{\hat{\psi}_n^S(\theta_0) - \mathbb{E}[\hat{\psi}_n^S(\theta_0)|u_1, \dots, u_n]}_{\text{scrambled sequence}}.$$

For the second term, Theorem 1 can be applied given that $\mathbb{E}[\hat{\psi}_i^S(\theta_0)|u_i]$ has finite variance. For the first term, Assumption A3 i. implies a Lyapunov condition holds. As a result, the

CLT for independent non-identically distributed arrays can be applied (White, 1984, Theorem 5.10). Note that similar arguments implies that for each $\theta \in \Theta$, $(\hat{\psi}_n^S(\theta) - \mathbb{E}[\hat{\psi}_n^S(\theta)]) = O_p(n^{-1/2})$, *i.e.* pointwise convergence holds.

Part ii. Stochastic Equicontinuity Result for $\hat{\psi}_n^S(\theta) - \mathbb{E}[\hat{\psi}_n^S(\theta)|u_1, \dots, u_n]$
As in Part i., Assumption A3 i. implies a Lyapunov condition holds for the envelope $\bar{\psi}$. This implies a Lindeberg condition for the envelope holds. Further, Assumption A3 iii. implies that:

$$\begin{aligned}
\sup_{\|\theta_1-\theta_2\|\le\delta_n} \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\|[\tilde{\psi}(x_i,u_i;\theta_1) - \tilde{\psi}(x_i,u_i;\theta_2)] - \mathbb{E}[\tilde{\psi}(x_i,u_i;\theta_1) - \tilde{\psi}(x_i,u_i;\theta_2)|u_i]\|^2|u_i\right] \\
\le \left(\frac{1}{n}\sum_{i=1}^n \tilde{C}_1(u_i)^2 + \mathbb{E}[\tilde{C}_1(u_i)^2]\right) \times \delta_n^2 \\
= \left(2\mathbb{E}[\tilde{C}_1(u_i)^2] + o_p(1)\right) \times \delta_n^2,
\end{aligned}$$

which goes to 0 for all sequences $\delta_n \to 0$. The last equality comes from applying Theorem 1 to $\tilde{C}(u_i)^2$ which has finite variance by assumption. Θ is a compact and convex subset of $\mathbb{R}^{d_\theta}$ which is finite dimensional. Given the Lindeberg condition, pointwise convergence in Part i. and the L^2-smoothness result above holds, the Jain-Markus Theorem can be applied[17] which implies the desired stochastic equicontinuity result.

Part iii. Taylor Expansion of $\mathbb{E}[\hat{\psi}_n^S(\theta)|u_1, \dots, u_n]$
For all θ_1, θ_2, Assumption A3 iv. implies:

$$\begin{aligned}
\|\frac{1}{n}\sum_{i=1}^n \{\mathbb{E}[\tilde{\psi}(x_i,u_i;\theta_1) - \tilde{\psi}(x_i,u_i;\theta_2)|u_i] - \partial_\theta \mathbb{E}\left[\tilde{\psi}(x_i,u_i;\theta_2)|u_i\right](\theta_1-\theta_2)\}\| \\
\le \frac{1}{n}\sum_{i=1}^n \tilde{C}_3(u_i) \times \|\theta_1-\theta_2\|^2 \\
= (\mathbb{E}[\tilde{C}_3(u_i)] + o_p(1)) \times \|\theta_1-\theta_2\|^2,
\end{aligned}$$

which implies the desired result. The last equality follows from Theorem 1 applied to $\tilde{C}_3(u_i)$ which has finite variance. Also note, that the conditions imply that the ULLN of Lemma B1 applies to $\partial_\theta \mathbb{E}[\hat{\psi}_n^S(\theta)|u_1, \dots, u_n]$ so that $\partial_\theta \mathbb{E}[\hat{\psi}_n^S(\theta)|u_1, \dots, u_n] = \partial_\theta \mathbb{E}[\hat{\psi}_n^S(\theta)] + o_p(1)$ uniformly in $\theta \in \Theta$. This concludes the proof. $\square$

[17]See Example 2.11.13 and Theorem 2.11.9 in van der Vaart & Wellner (1996).

Proof of Proposition A2. By Lemma B2, $\hat{\psi}_n - \hat{\psi}_n^S(\theta)$ is stochastically equicontinuous which, together with Assumption A1, implies that $\hat{\theta}_n^S \xrightarrow{p} \theta_0$ by Theorem 2.1 in Newey & McFadden (1994). Then, using Lemma B2 and standard arguments, we have:

$$
\begin{aligned}
0 &= G'W\mathbb{E}\left[\hat{\psi}_n - \hat{\psi}_n^S(\theta_0)\right] \\
&= G'W\left(\mathbb{E}\left[\hat{\psi}_n - \hat{\psi}_n^S(\theta_0)\right] - \left[\hat{\psi}_n - \hat{\psi}_n^S(\theta_0)\right] + \left[\hat{\psi}_n - \hat{\psi}_n^S(\theta_0)\right]\right) \\
&= G'W\left(\mathbb{E}\left[\hat{\psi}_n - \hat{\psi}_n^S(\hat{\theta}_n^S)|u_1,\dots,u_n\right] - \left[\hat{\psi}_n - \hat{\psi}_n^S(\hat{\theta}_n^S)\right] + \left[\hat{\psi}_n - \hat{\psi}_n^S(\theta_0)\right]\right) + o_p(n^{-1/2}) \\
&= G'W\left(\mathbb{E}\left[\hat{\psi}_n - \hat{\psi}_n^S(\hat{\theta}_n^S)|u_1,\dots,u_n\right] + \left[\hat{\psi}_n - \hat{\psi}_n^S(\theta_0)\right]\right) + o_p(n^{-1/2}) \\
&= G'W\left(\mathbb{E}\left[\hat{\psi}_n^S(\theta_0) - \hat{\psi}_n^S(\hat{\theta}_n^S)|u_1,\dots,u_n\right] + \left[\hat{\psi}_n - \hat{\psi}_n^S(\theta_0)\right]\right) + o_p(n^{-1/2}).
\end{aligned}
$$

The stochastic equicontinuity result can then be applied:

$$
\begin{aligned}
&\mathbb{E}\left[\hat{\psi}_n - \hat{\psi}_n^S(\theta_0)|u_1,\dots,u_n\right] - \left[\hat{\psi}_n - \hat{\psi}_n^S(\theta_0)\right] \\
&= \mathbb{E}\left[\hat{\psi}_n - \hat{\psi}_n^S(\hat{\theta}_n^S)|u_1,\dots,u_n\right] - \left[\hat{\psi}_n - \hat{\psi}_n^S(\hat{\theta}_n^S)\right] + o_p(n^{-1/2}).
\end{aligned}
$$

Then, by Theorem 1, $\|\mathbb{E}\left[\hat{\psi}_n - \hat{\psi}_n^S(\theta_0)|u_1,\dots,u_n\right] - \mathbb{E}\left[\hat{\psi}_n - \hat{\psi}_n^S(\theta_0)\right]\| = o_p(n^{-1/2})$ which allows to substitute $\mathbb{E}\left[\hat{\psi}_n - \hat{\psi}_n^S(\theta_0)\right]$ with the desired quantity. Using the CLT and stochastic equicontinuity result in Lemma B2:

$$
\begin{aligned}
\sqrt{n}\left(\hat{\theta}_n^S - \theta\right) &= -\left(G'WG\right)^{-1} G'W\sqrt{n}[\hat{\psi}_n - \hat{\psi}_n^S(\theta_0)] + o_p(1) \\
&\xrightarrow{d} \mathcal{N}(0,\Sigma),
\end{aligned}
$$

where

$$
\Sigma = \left(G'WG\right)^{-1} G'W\tilde{V}WG\left(G'WG\right)^{-1},
$$

$G = \partial_\theta \mathbb{E}[\hat{\psi}_n^S(\theta_0)]$, $\tilde{V} = \mathbb{E}\left[\text{var}\left(\psi(y_i,x_i) - \tilde{\psi}(x_i,u_i;\theta_0)|u_i\right)\right]$.

The results above are given for $S = 1$. For $S > 1$ fixed and finite, the simulated moments $\hat{\psi}_n^s$ are *iid* over $s = 1,\dots,S$. This implies that the CLT and stochastic equicontinuity results can be applied to each $s \in \{1,\dots,S\}$ and also apply to their average $\hat{\psi}_n^S$ by independence with S fixed and finite. The remainder of the proof is identical which concludes the proof. □

B.1.3 Scrambled Indirect Inference

Proof of Proposition A3. Assumption A4 ii. implies a ULLN for $M_n^S(\theta;\psi)$ in ψ for all $\theta \in \Theta$, by Lemma B1. Then Assumption A4 i. implies that Theorem 2.1 in Newey & McFadden (1994) applies for each $\theta \in \Theta$ to $\hat{\psi}_n^S$ so that $\hat{\psi}_n^S(\theta) - \psi_\infty(\theta) = o_p(1)$ pointwise in $\theta \in \Theta$.

Now, to prove that $\hat{\theta}_n^S$ itself is consistent, a ULLN for $\hat{\psi}_n^S$ in θ is needed. Given the pointwise consistency above, it remains to show that $\hat{\psi}_n^S$ is Lispchitz-continuous with stochastically bounded Lipschitz constant. For all $\theta_1, \theta_2 \in \Theta$, the mean-value theorem and the triangular inequality imply:

$$\|\hat{\psi}_n^S(\theta_1) - \hat{\psi}_n^S(\theta_2)\| \leq \|\partial_\theta \hat{\psi}_n^S(\tilde{\theta})\| \times \|\theta_1 - \theta_2\|,$$

where $\tilde{\theta}$ is some intermediate value. The implicit function theorem provides a closed-form for $\partial_\theta \hat{\psi}_n^S$ evaluated at any $\theta \in \Theta$:

$$\partial_\theta \hat{\psi}_n^S(\theta) = -\left[\partial^2_{\psi,\psi'} M_n^S(\theta; \hat{\psi}_n^S(\theta))\right]^{-1} \partial^2_{\psi,\theta'} M_n^S(\theta; \hat{\psi}_n^S(\theta)).$$

Both $\partial^2_{\psi,\psi'} M_n^S(\theta;\psi)$ and $\partial^2_{\psi,\theta'} M_n^S(\theta;\psi)$ satisfy a ULLN in (θ,ψ) by Assumption A4 and Lemma B1. The Continuous Mapping Theorem then implies that $\partial_\theta \hat{\psi}_n^S(\theta) \xrightarrow{p} \mathbb{E}(\partial_\theta \hat{\psi}_n^S(\theta))$ pointwise in $\theta \in \Theta$. Furthermore, Assumption A4 iv. implies that:

$$\left\| \left(\partial^2_{\psi,\psi'} \mathbb{E}[M_n^S(\theta;\psi)]\right)^{-1} \partial^2_{\psi,\theta'} \mathbb{E}[M_n^S(\theta;\psi)] \right\| \leq \bar{M} < \infty,$$

uniformly in (θ,ψ) for some finite bound $\bar{M} \geq 0$. Putting everything together, we have:

$$\begin{aligned}
\|\hat{\psi}_n^S(\theta_1) - \hat{\psi}_n^S(\theta_2)\| &\leq \|\partial_\theta \hat{\psi}_n^S(\tilde{\theta})\| \times \|\theta_1 - \theta_2\| \\
&= \|\mathbb{E}[\partial_\theta \hat{\psi}_n^S(\tilde{\theta})] + o_p(1)\| \times \|\theta_1 - \theta_2\| \\
&\leq [\bar{M} + o_p(1)] \times \|\theta_1 - \theta_2\|
\end{aligned}$$

This implies, as in Jennrich (1969) and Proposition A1, a ULLN for $\hat{\psi}_n^S$ over $\theta \in \Theta$.

To establish the asymptotic normality for $\hat{\theta}_n^S$, first note that the ULLN for $\hat{\psi}_n^S$, $\partial^2_{\psi,\psi'} M_n^S$ and $\partial^2_{\psi,\theta'} M_n^S$ together with the implicit function theorem and the Lipschitz conditions imply a ULLN for $\partial_\theta \hat{\psi}_n^S$ in θ.[18] By the usual mean-value expansion argument, this implies that:

$$\sqrt{n}[\hat{\theta}_n^S - \theta_0] = -\sqrt{n}\left[\partial_\theta \psi_\infty(\theta_0)' W \partial_\theta \psi_\infty(\theta_0) + o_p(1)\right]^{-1} \partial_\theta \psi_\infty(\theta_0)' W [\hat{\psi}_n - \hat{\psi}_n^S(\theta_0)] + o_p(1).$$

To conclude the proof, we need to show that $\sqrt{n}[\hat{\psi}_n^S(\theta_0) - \psi_\infty(\theta_0)] = o_p(1)$. Since $\hat{\psi}_n^S(\theta_0)$ is an M-estimator with the appropriate regularity conditions, the following holds:[19]

$$\sqrt{n}[\hat{\psi}_n^S(\theta_0) - \psi_\infty(\theta_0)] = -\left[\partial^2_{\psi,\psi'} \mathbb{E}[M_n^S(\theta_0; \psi_\infty(\theta_0))] + o_p(1)\right]^{-1} \partial_\psi M_n^S(\theta_0; \psi_\infty(\theta_0)).$$

Since $\psi_\infty(\theta_0)$ is the population minimizer of $\mathbb{E}[M_n^S(\theta_0;\cdot)]$, we have $\partial_\psi \mathbb{E}[M_n^S(\theta_0;\psi_\infty(\theta_0))] = 0$. Applying Theorem 1 with Assumption A4 iii. implies $\partial_\psi M_n^S(\theta_0;\psi_\infty(\theta_0)) = o_p(n^{-1/2})$ which, in turn, implies the desired result and concludes the proof. $\square$

[18] The proof is omitted for brevity but is similar to the previous ULLNs.

[19] The proof is very similar to Proposition A1.

B.2 Dynamic Models

B.2.1 qMC-only Estimator

Proof of Proposition A4. Given the construction in Algorithm 5 and the assumptions the results of Lemma B1 hold and the proof proceeds as in Proposition A1. This concludes the proof. $\square$

B.2.2 Hybrid MC-qMC Estimator

Lemma B3 (Uniform Law of Large Numbers and CLT - MC-qMC)**.** *Suppose that the Assumptions A1, A5 hold then:*

i. $\sup_{\theta\in\Theta} \|\hat{\psi}_T^S(\theta) - \mathbb{E}[\hat{\psi}_T^S(\theta)]\| = o_p(1)$,

ii. $\sup_{\theta\in\Theta} \|\partial_\theta\hat{\psi}_T^S(\theta) - \partial_\theta\mathbb{E}[\hat{\psi}_T^S(\theta)]\| = o_p(1)$,

iii. $\sqrt{TS}\left(\hat{\psi}_T^S(\theta_0) - \mathbb{E}[\hat{\psi}_T^S(\theta_0)]\right) \overset{d}{\to} \mathcal{N}(0, V)$ *where* $V = \lim_{T\to\infty} T\times var[\hat{\psi}_T^S(\theta_0)|u_1, \dots, u_{TS}]$.

Proof of Lemma B3.
Part i. ULLN for $\hat{\psi}_T^S(\theta)$
The main steps as similar to Lemma B1 using pointwise convergence and Lipschitz continuity arguments. The main difficulty is the presence of the Monte-Carlo terms y_t^1, z_t^1 which are dependent and non-stationary. To handle these, as in the proof of Lemma B2, separate $\hat{\psi}_T^S - \mathbb{E}[\hat{\psi}_T^S]$ into two components $(\hat{\psi}_T^S - \mathbb{E}[\hat{\psi}_T^S|u_1, \dots, u_{TS}])$ and $(\mathbb{E}[\hat{\psi}_T^S|u_1, \dots, u_{TS}] - \mathbb{E}[\hat{\psi}_T^S])$ to study the two individually. For the first term, Davydov (1968)'s inequality implies pointwise convergence under mixing and moment conditions. For the second term, the non-stationarity implies that Theorem 1 does not apply directly. The geometric ergodicity conditions will allow to return to a setting where Theorem 1 applies.

As discussed above, for any $\theta \in \Theta$:

$$\hat{\psi}_T^S(\theta) - \mathbb{E}[\hat{\psi}_T^S(\theta)] = \underbrace{\hat{\psi}_T^S(\theta) - \mathbb{E}[\hat{\psi}_T^S(\theta)|u_1, \dots, u_{TS}]}_{\text{heterogeneous dependent vector}} + \underbrace{\mathbb{E}[\hat{\psi}_T^S(\theta)|u_1, \dots, u_{TS}] - \mathbb{E}[\hat{\psi}_T^S(\theta)]}_{\text{non-stationary qMC sequence}}.$$

For the first term, Davydov's inequality implies, up to a universal constant:

$$\begin{aligned}
&\mathbb{E}[\|\hat{\psi}_T^S(\theta) - \mathbb{E}[\hat{\psi}_T^S(\theta)|u_1, \dots, u_{TS}]\|^2|u_1, \dots, u_{TS}] \\
&\leq \frac{1}{[TS]^2} \sum_{t=1}^{TS} \mathbb{E}[\|\tilde{\psi}(y_t^1, z_t^1, u_t; \theta) - \mathbb{E}[\tilde{\psi}(y_t^1, z_t^1, u_t; \theta)|u_t]\|^2|u_t] \\
&+ \frac{1}{[TS]^2} \sum_{t \neq t'} \alpha(|t-t'|)^{1/2} \times \mathbb{E}[\|\tilde{\psi}(y_t^1, z_t^1, u_t; \theta) - \mathbb{E}[\tilde{\psi}(y_t^1, z_t^1, u_t; \theta)|u_t]\|^4|u_t]^{1/4} \\
&\qquad\qquad\qquad\qquad \times \mathbb{E}[\|\tilde{\psi}(\tilde{y}_{t'}, \tilde{z}_{t'}, u_{t'}; \theta) - \mathbb{E}[\tilde{\psi}(y_{t'}^1, z_{t'}^1, u_{t'}; \theta)|u_{t'}]\|^4|u_{t'}]^{1/4}.
\end{aligned}$$

Note that $\mathbb{E}[\|\tilde{\psi}(y_t^1, z_t^1, u_t; \theta) - \mathbb{E}[\tilde{\psi}(y_t^1, z_t^1, u_t; \theta)|u_t]\|^2|u_t]$ is not stationary, so that Theorem 1 does not apply directly. However, by geometric ergodicity we have for any function g with bounded fourth moment:

$$\begin{aligned}
&\|\mathbb{E}[g(y_t^1, z_t^1, u_t; \theta) - g(y_t^\infty, z_t^\infty, u_t; \theta)|u_t]\| \\
&= \| \int g(y^1, z^1, u_t; \theta)[f_t(y^1, z^1) - f_\infty(y^1, z^1)]dy^1 dz^1 \| \\
&\leq \int \|g(y^1, z^1, u_t; \theta)\| \times |f_t(y^1, z^1) - f_\infty(y^1, z^1)|dy^1 dz^1 \\
&\leq \left(\int \|g(y^1, z^1, u_t; \theta)\|^2 \times |f_t(y^1, z^1) - f_\infty(y^1, z^1)|dy^1 dz^1 \right)^{1/2} \left(\int |f_t(y^1, z^1) - f_\infty(y^1, z^1)|dy^1 dz^1 \right)^{1/2} \\
&\leq \sqrt{2} \times \bar{K}_g \times \|f_t - f_\infty\|_{TV}^{1/2} \\
&\leq \sqrt{2C_1} \times \bar{K}_g \times \rho^{t/2},
\end{aligned}$$

where $\bar{K}_g \geq 0$ is a bound for the moment conditional on u_t fixed. This bound is finite by Assumption A5 iii. and v. for $\tilde{\psi}$ and $\partial_\theta \tilde{\psi}$, respectively. Under the geometric ergodicity assumption, $\rho \in [0, 1)$ so that $\sum_{t \geq 0} \rho^{t/2} < +\infty$ and:

$$\begin{aligned}
&\frac{1}{[TS]^2} \sum_{t=1}^{TS} \mathbb{E}[\|\tilde{\psi}(y_t^1, z_t^1, u_t; \theta) - \mathbb{E}[\tilde{\psi}(y_t^1, z_t^1, u_t; \theta)|u_t]\|^2|u_t] \\
&= \frac{1}{[TS]^2} \sum_{t=1}^{TS} \left(\mathbb{E}[\|\tilde{\psi}(y_t^\infty, z_t^\infty, u_t; \theta) - \mathbb{E}[\tilde{\psi}(y_t^\infty, z_t^\infty, u_t; \theta)|u_t]\|^2|u_t] \right) + O(1/[TS]^2) \\
&= \mathbb{E} \left(\mathbb{E}[\|\tilde{\psi}(y_t^\infty, z_t^\infty, u_t; \theta) - \mathbb{E}[\tilde{\psi}(y_t^\infty, z_t^\infty, u_t; \theta)|u_t]\|^2] \right) /[TS] + o_p(1/[TS]) + O(1/[TS]^2),
\end{aligned}$$

where the last equality is due to Theorem 1 using the bounded fourth moment assumption to find the finite variance condition needed in the Theorem.

The second term, which is a non-stationary qMC sequence, can be handled using the geometric ergodicity condition and the bounded fourth moment asusmption to get:

$$\begin{aligned}
&\frac{1}{TS}\sum_{t=1}^{TS}\Big(\mathbb{E}[\tilde{\psi}(y_t^1, z_t^1, u_t; \theta)|u_t] - \mathbb{E}[\tilde{\psi}(y_t^1, z_t^1, u_t; \theta)]\Big) \\
&= \frac{1}{TS}\sum_{t=1}^{TS}\Big(\mathbb{E}[\tilde{\psi}(y_t^\infty, z_t^\infty, u_t; \theta)|u_t] - \mathbb{E}[\tilde{\psi}(y_t^\infty, z_t^\infty, u_t; \theta)]\Big) + O(1/[TS]) \\
&= o_p(1/\sqrt{TS}) + O(1/[TS]).
\end{aligned}$$

Finally, the geometric ergodicity imply that $(y_t^1, z_t^1)_{t\geq 1}$ is α-mixing with exponential decay. This implies that $\frac{1}{[TS]^2}\sum_{t\neq t'}\alpha(|t-t'|) = O(1/[TS])$ where α are the α-mixing coefficients. Furthermore, by assumption $\mathbb{E}[\|\tilde{\psi}(y_t^1, z_t^1, u_t; \theta) - \mathbb{E}[\tilde{\psi}(y_t^1, z_t^1, u_t; \theta)|u_t]\|^4 |u_t]^{1/4}$ is bounded for all $t \geq 1$. Altogether, these imply that:

$$\begin{aligned}
\hat{\psi}_T^S(\theta) - \mathbb{E}[\hat{\psi}_T^S(\theta)] &= \underbrace{\hat{\psi}_T^S(\theta) - \mathbb{E}[\hat{\psi}_T^S(\theta)|u_1, \dots, u_{TS}]}_{=O_p(1/\sqrt{TS})} + \underbrace{\mathbb{E}[\hat{\psi}_T^S(\theta)|u_1, \dots, u_{TS}] - \mathbb{E}[\hat{\psi}_T^S(\theta)]}_{=o_p(1/\sqrt{TS})} \\
&= O_p(1/\sqrt{TS}),
\end{aligned}$$

which implies pointwise convergence.

As in Proposition A1, take a cover $\{\theta_1, \dots, \theta_J\}$ of Θ and:

$$\begin{aligned}
&\sup_{\theta\in\Theta}\|\hat{\psi}_T^S(\theta) - \mathbb{E}[\hat{\psi}_T^S(\theta)]\| \\
&\quad\leq \max_{j\in\{1,\dots,J\}}\|\hat{\psi}_T^S(\theta_j) - \mathbb{E}[\hat{\psi}_T^S(\theta_j)]\| + \sup_{\theta\in\Theta}\min_{j\in\{1,\dots,J\}}\Big\|[\hat{\psi}_T^S(\theta) - \hat{\psi}_T^S(\theta_j)] - \mathbb{E}[\hat{\psi}_T^S(\theta) - \hat{\psi}_T^S(\theta_j)]\Big\|.
\end{aligned}$$

The first term can be handled with the pointwise convergence result above. For the second term, note that:

$$\|\hat{\psi}_T^S(\theta) - \hat{\psi}_T^S(\theta_j)\| \leq \frac{1}{TS}\sum_{t=1}^{TS} C_2(y_t^1, z_t^1, u_t; \theta_j) \times \|\theta - \theta_j\|.$$

It is sufficient to show that $\sum_{t=1}^{TS} C_2(y_t^1, z_t^1, u_t; \theta_j)/[TS]$ is a $O_p(1)$ for each $j \in \{1, \dots, J\}$. Since C_2 satisfies the conditions for the pointwise convergence derived above, using the same arguments as for $\hat{\psi}_T^S$ we have:

$$\frac{1}{TS}\sum_{t=1}^{TS} C_2(y_t^1, z_t^1, u_t; \theta_j) \xrightarrow{p} \mathbb{E}[C_2(y_t^\infty, z_t^\infty, u_t; \theta_j)].$$

As in the proof of Lemma B1, for J and T large enough we have:

$$\sup_{\theta \in \Theta} \|\hat{\psi}_T^S(\theta) - \mathbb{E}[\hat{\psi}_T^S(\theta)]\| \leq \varepsilon,$$

with probability going to 1, which implies the desired result.

Part ii. ULLN for $\partial_\theta \hat{\psi}_T^S$
Given the stated assumptions, the same results as above apply to $\partial_\theta \hat{\psi}_T^S$ uniformly in $\theta \in \Theta$.

Part iii. CLT for $\sqrt{TS}\left(\hat{\psi}_T^S(\theta_0) - \mathbb{E}[\hat{\psi}_T^S(\theta_0)]\right)$
In part i., it was shown that $\mathbb{E}[\hat{\psi}_T^S(\theta_0)|u_1, \dots, u_{TS}] - \mathbb{E}[\hat{\psi}_T^S(\theta_0)] = o_p(1/\sqrt{TS})$. Then, the bounded fourth moment in Assumption A5 ii., the mixing condition i. and the variance condition vi. imply that the CLT for heterogeneous dependent arrays (White, 1984, Theorem 5.20) can be applied and:

$$\sqrt{TS}\left(\hat{\psi}_n^S(\theta_0) - \mathbb{E}[\hat{\psi}_T^S(\theta_0)|u_1, \dots, u_{TS}]\right) \xrightarrow{d} \mathcal{N}(0, V),$$

where $V = \lim_{T\to\infty} T \times \text{var}[\hat{\psi}_T^S(\theta_0)|u_1, \dots, u_{TS}]$. This concludes the proof. □

Proof of Proposition A5. Given the assumptions, Lemma B3 applies and the proof proceed as in Proposition A1. This concludes the proof. □