Parametric Risk-Neutral Density Estimation via Finite Lognormal-Weibull Mixtures*

Yifan Li† Ingmar Nolte‡ Manh Cuong Pham§

University of Manchester Lancaster University Lancaster University

Abstract

This paper proposes a new parametric risk-neutral density (RND) estimator based on a finite lognormal-Weibull mixture (LWM) density. We establish the consistency and asymptotic normality of the LWM method in a general misspecified parametric framework. Based on the theoretical results, we propose a sequential test procedure to evaluate the goodness-of-fit of the LWM model, which leads to an adaptive choice for the number and type of mixture components. Our simulation results show that, in finite samples with various observation error specifications, the LWM method can approximate complex RNDs generated by state-of-the-art multi-factor stochastic volatility models with a few (typically less than 4) mixtures. Application of the LWM model on index options confirms its reliability in recovering empirical RNDs with a heavy left tail or bimodality, which can be incorrectly identified as bimodality or a heavy left tail by existing (semi)-nonparametric methods if the goodness-of-fit to the observed data is ignored.

JEL classification: C13, C52, C58, G13

Keywords: risk-neutral density, parametric modelling, mixture-of-distribution method.

*We are grateful to the co-editor (Torben Andersen), the associate editor, and two anonymous referees for their insightful comments and suggestions, which greatly improved the quality of this paper. We thank Olivier Scaillet, Eric Renault, Yang Zhang, and all participants in the Vienna Workshop on Econometrics of Option Markets and the Liverpool Econometrics Workshop 2022 for their helpful comments. We would like to acknowledge the financial support from the ESRC-FWF bilateral grant titled "Bilateral Austria: Order Book Foundations of Price Risks and Liquidity: An Integrated Equity and Derivatives Markets Perspective", Grant Ref: ES/N014588/1 and the Austrian Science Fund (FWF): Research project: I-2762-G27. The usual disclaimer applies.

†Corresponding author: Alliance Manchester Business School, Booth Street W, Manchester, M13 9SS, UK. Phone +44 16130 66402, e-mail: yifan.li@manchester.ac.uk.

‡Lancaster University Management School Bailrigg, Lancaster LA1 4YX, UK. Phone +44 15245 92644, email: I.Nolte@lancaster.ac.uk

§Lancaster University Management School, Bailrigg, Lancaster, LA1 4YX, UK. Phone +44 15245 94816, e-mail: m.c.pham@lancaster.ac.uk.

1 Introduction

The risk-neutral density (RND), also termed the state price density, summarizes the distribution of asset prices at a fixed point of time in the future under the risk-neutral probability measure. The RND is believed to contain rich forward-looking information about the risk preference and market expectations of the underlying asset. Applications of the RND can be found in Aït-Sahalia and Lo (2000), Bliss and Panigirtzoglou (2004), Liu et al. (2007), Shackleton et al. (2010), Birru and Figlewski (2012), Christoffersen et al. (2013), Ghysels and Wang (2014), Cuesdeanu and Jackwerth (2018) among many others.

Since the seminal works of Breeden and Litzenberger (1978) and Banz and Miller (1978) which show that the RND can be expressed as the second-order derivative of option prices as a function of the strike, numerous methods[1] have been proposed to extract RNDs from the observed option prices. Popular RND extraction methods can be classified into three classes: (1) parametric specification of the RND, such as the mixture-of-lognormal (MLN) approach of Ritchey (1990), Melick and Thomas (1997), the normal inverse Gaussian of Eriksson et al. (2009), the generalized extreme value distribution of Markose and Alentorn (2011); (2) semi-nonparametric (SNP) approximations of the RND, including the Edgeworth expansion or Gram-Charlier series of Jarrow and Rudd (1982), Longstaff (1995), Jondeau and Rockinger (2001), the confluent hypergeometric functions of Abadir and Rockinger (2003), Bu and Hadri (2007), the mixture-based density approximation of Bondarenko (2003), Yuan (2009) and the recent sieve estimator of Lu and Qu (2021) based on the Gauss-Hermite expansion; (3) nonparametric (NP) methods, for example, the kernel or curve-fitting approach of Aït-Sahalia and Duarte (2003), Bliss and Panigirtzoglou (2004), Yatchew and Härdle (2006), Härdle and Hlávka (2009), Figlewski (2010), Birru and Figlewski (2012) and the implied binomial tree of Rubinstein (1994).

The dense literature on RND estimation reflects the difficulty of the estimation problem. In detail, one needs to extract a complex empirical RND from a relatively small cross-section of option prices. For some recent examples, Song and Xiu (2016), Andersen et al. (2017, 2021a) suggest that an empirical RND is typically unimodal with a heavy left tail, while Clark and Amen (2018), Ferreira et al. (2022), Kostakis et al. (2023) show that the RND can become bimodal in the presence of major binary event risk, such as the US presidential election or the Brexit referendum. This explains the wide implementation of (S)NP methods in recent empirical studies which are designed to approximate an arbitrary unknown density function. Meanwhile, the parametric approach receives much less attention, as it is believed that existing parametric RND models may not be flexible enough to fit the diverse

[1] Comprehensive surveys of various methods can be found in Jackwerth (1999, 2004), Taylor (2005), Figlewski (2018). See also Fengler and Hin (2015), Ludwig (2015), Orosi (2015), Feng and Dang (2016), Kundu et al. (2019), Dalderop (2020), Monteiro and Santos (2020) for some recent developments.

empirical RND shapes.

Motivated by the relatively less developed literature on the parametric RND extraction approaches, this paper proposes a novel parametric method by modelling RNDs with a finite lognormal-Weibull mixture (LWM) density. The proposed method has three main features: (a) it extends the parametric MLN method by adding Weibull mixture components, which is designed to approximate the heavy left tail and the potential multi-modality of the empirical RND; (b) the LWM-implied option prices are available in closed-form, and non-arbitrage conditions for the RND estimates are guaranteed to hold by simple parameter constraints; (c) the number and type of mixtures are chosen by an adaptive algorithm based on residual diagnostics, which produce a parsimonious model that balances misspecification and overfitting.

We make three core contributions to the RND estimation literature. First, we develop asymptotic theory for the parametric RND estimator under an infill asymptotic setting with a fixed strike range. To the best of our knowledge, we are among the first to formalize a parametric RND estimation framework which simultaneously accounts for potential model misspecification and a general observation error. This allows us to construct asymptotic confidence bounds for the RND estimates and conduct post-estimation diagnostic tests, which further provides rigorous econometric tools to examine the goodness-of-fit of the LWM method.

Second, our comprehensive simulation results suggest that, with only a small number of mixtures, the LWM density can reliably approximate complex RNDs generated by affine jump-diffusive models featuring heavy tails or bimodality. Under various settings of samples sizes and observation errors, the LMW approach is comparable to popular optimally tuned (S)NP methods in Aït-Sahalia and Duarte (2003), Bondarenko (2003), Bliss and Panigirtzoglou (2004), Lu and Qu (2021), and sizeable gains in precision are expected if one only adopts 'recommended' tuning parameters in the existing literature. This ensures that the LWM approach can generate reliable RND estimates under rapidly changing market conditions that are more robust to tuning than popular (S)NP methods.

Third, we highlight the practical importance of post-estimation residual diagnostic tests. In our empirical study, we re-examine a bimodal RND estimate from S&P index options in Lu and Qu (2021) and a unimodal RND estimate from FTSE100 index options in Kostakis et al. (2023). Our findings convincingly show that the (S)NP-based empirical RND estimates with recommended tuning parameters can produce spurious bimodality or fail to detect genuine bimodality. This is due to insufficient flexibility of the (S)NP-based RND model caused by poor choices of tuning parameters. By contrast, the LWM method provides more reliable RND estimates in both cases with a significantly better fit to the observed data.

The remainder of the paper is structured as follows. Section 2 lays down the econometric setting

for the parametric LWM approach. Section 3 presents the main theoretical results of the paper. The simulation study and empirical illustrations can be found in Sections 4 and 5, respectively. Section 6 concludes.

2 Econometric Settings

2.1 Risk-Neutral Density Estimation

Let S_t denote the price of a stock at time t. We are interested in estimating the risk-neutral density[2] (RND) of S_T for some $T > t$, denoted as $f_t^*(x)$, based on observations of option prices. Throughout the paper, let F_t denote the price of a forward contract of the stock expiring at time T, and we have $S_T = F_T$ by design. As S_t is observed and F_t can be computed from S_t using observed quantities (e.g., risk-free rates and dividend yields), we shall assume that F_t is also observed.

Breeden and Litzenberger (1978) show that $f_t^*(x)$ is closely related to the price function of European options. Let $O^*(K)$ denote the theoretical (put or call) price of an European option at time t with time-to-maturity $\tau := T - t$ and constant risk-free rate r_f, then $f_t^*(x)$ can be derived from $O^*(K)$ as:

$$f_t^*(x)|_{x=K} = e^{-r_f \tau} \frac{\partial^2 O^*(K)}{\partial K^2}. \tag{1}$$

Although the option price depends on additional state variables such as spot volatility factors or jumps at time t, we shall suppress these inputs and assume them to be identical for all options in a cross-section observed at time t with the same time-to-maturity. Consequently, we consider option prices as a univariate function of the strikes in this paper.

To ensure that $f_t^*(x)$ exists so that the estimation problem is well-defined, we firstly state an assumption about the theoretical option pricing function. For illustrative purpose, we formulate our assumption in terms of a call option. Similar conditions can easily be derived for a put option using the put-call parity, and are thus omitted in the paper.

Assumption 1. *We assume that there exists a true latent European call option pricing function $C^*(K)$ which is at least twice differentiable on $\mathbb{R}^+$ and satisfies the following conditions: (1) $\frac{\partial C^*(K)}{\partial K} \in [-e^{r_f \tau}, 0]$; (2) $\frac{\partial^2 C^*(K)}{\partial^2 K} = e^{r_f \tau} f_t^*(K) \geq 0$; (3) $\int_0^\infty f_t^*(x)dx = 1$; (4) $\int_0^\infty x f_t^*(x)dx = F_t$. The function $f_t^*(x)$ is called the true latent RND of the underlying price at expiry F_T.*

Remark 1. *Conditions (1)-(4) can be found in Aït-Sahalia and Duarte (2003), which are non-arbitrage conditions for the cross-section of option prices. Specifically, conditions (2) and (3) ensure that $f_t^*(x)$*

[2] In detail, the risk-neutral density of the stock at time t with a horizon $T > t$ is the probability density function of S_T conditional on the information set at time t, $\mathcal{F}_t$, under the risk-neutral probability measure.

is a well-defined density function on $\mathbb{R}^+$. *Condition (4) implies that* $\mathrm{E}^{\mathbb{Q}}[F_T|\mathcal{F}_t] = F_t$ *to eliminate arbitrage opportunities using a forward contract, where* $\mathrm{E}^{\mathbb{Q}}[\cdot]$ *denotes the expectation under the risk-neutral probability measure.*

We will also denote $O^*(K)$ as the general true option pricing function (put or call) implied by Assumption 1. In practice, we do not observe $O^*(K)$ for a continuum of K, but rather on a discrete strike grid with observation errors. We impose the following assumption on the strike grid and the asymptotic setting of this paper:

Assumption 2. *The observed strike prices are contained in a fixed and compact interval* $\mathcal{K} = [\underline{K}, \overline{K}]$ *for some finite* $0 < \underline{K} < \overline{K}$, *and denote* $m(\mathcal{K}) := \overline{K} - \underline{K}$ *as the length of the strike range. With a sample size of* N, *the strike grid is:*

$$\underline{K} = K_{1,N} < K_{2,N} < \ldots < K_{N,N} = \overline{K},$$

which is a (sequence of) non-random and strictly increasing partition of $\mathcal{K}$. *We assume that there exists a deterministic, Lipschitz and strictly increasing function* $\kappa : \mathcal{K} \mapsto [0, 1]$ *such that as* $N \to \infty$*:*

$$\sup_{1 \le n \le N} \left| \kappa(K_{n,N}) - \frac{n}{N} \right| = o(N^{-1}). \tag{2}$$

Our asymptotic setting is of the infill type similar to Andersen et al. (2021b), in which we observe an increasing amount of strike prices on a fixed range. This is a reasonable assumption in practice as equity options are only actively traded on an economically meaningful range of moneyness[3]. Eq. (2) assumes an eventually equidistant strike grid up to an invertible transformation $\kappa(x)$, which is designed to capture the empirical observation that the traded strike grid gets sparser as the moneyness becomes more extreme. For the ease of notation, we will omit the subscript N whenever no confusion is caused. The following assumption characterizes the option pricing errors:

Assumption 3. *On a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$, *let* $Z = (Z_t)_{t \in \mathbb{Z}}$ *define a strictly stationary and exponentially* α*-mixing time series satisfying: (1)* $\mathrm{E}[Z_t] = 0$*; (2)* $\mathrm{Var}[Z_t] = 1$*; (3)* $\mathrm{E}[Z_t^4] < \infty$*; (4) Denote* $\sigma_Z^2 := \sum_{j=-\infty}^{\infty} \mathrm{E}[Z_t Z_{t-j}]$ *as the long run variance of* Z_t, *which satisfies* $\sigma_Z^2 < \infty$.

For each $N \in \mathbb{N}$, *the observed option prices* $\{O_n\}_{n=1:N}$ *are generated by:*

$$O_n = O^*(K_n) + u_n, \tag{3}$$

where $u_n = \gamma(K_n) Z_{n,N}$ *is the pricing error at the* n*th strike grid. The triangular array* $\{Z_{n,N}\}_{n=1:N}$

[3] Equity options with an absolute moneyness of >10 are seldomly traded. See e.g., Figure 7 of Andersen et al. (2021a) for some recent empirical evidence.

is a realization of N observations of Z. The function $\gamma(x)$ is a bounded and strictly positive Hölder continuous function on $\mathcal{K}$.

Remark 2. *It is worth pointing out that Assumption 3 implies a conditional exogeneity-type assumption for the pricing errors $\mathrm{E}[u_n|K_n] = 0, \forall n$, so that the pricing errors are centred and independent from the strike prices. This serves as our identification assumption of both the parametric model estimation and the post-estimation diagnostics. Importantly, this assumption precludes strike-dependent deviations from the efficient option prices, such as a 'clientele effect' when large investors trade at certain strike ranges, which jointly affects all observed prices in this range and can induce a strong dependence between K_n and u_n. In this paper, we shall exclude this possibility to avoid ambiguities in the definition of efficient option prices.*

Assumption 3 is similar to those used in Andersen et al. (2021b), which provides a very flexible structure to account for heteroscedastic and spatial correlated observation errors in option prices. The function $\gamma(x)$ captures possible heteroscedasticity of pricing errors related to the strike grid. The Hölder continuity of $\gamma(x)$ is a technical assumption used to prove the convergence of the variance-covariance matrix, which can be further weakened to allow for finitely many discontinuities on $\mathcal{K}$. The $Z_{n,N}$ process allows spatial correlation across the strike grid. Although the assumptions on Z_t can be further weakened (e.g. α-mixing with a polynomial mixing rate), from the empirical findings in Andersen et al. (2021b) we believe that these assumptions are sufficient to capture the dynamics in the empirical option pricing errors.

Note that Assumption 3 does not cover the case where more than one observation (e.g., both a put and a call, or multiple cross-sections of option prices) is observed at the same strike price. This requires a panel-type assumption for the observation errors, which is beyond the scope of this paper. To simplify the discussion, we shall stick with Assumption 3 and focus on a single cross-section of out-of-the-money (OTM) puts and calls due to their relatively higher liquidity than the in-the-money (ITM) options.

2.2 The Lognormal-Weibull Mixture Model

The true RND $f_t^*(x)$ can take arbitrary forms as long as Assumption 1 is satisfied. In this paper, we consider a parametric approach by approximating $f_t^*(x)$ with a Lognormal-Weibull-Mixture (LWM) density. The density takes the following form:

Definition 1. *For non-negative integers M_1 and M such that $(M_1 \vee 1) \leq M$, an M-LWM density with M_1 lognormal components (and $M_2 = M - M_1$ Weibull components) is defined as the following discrete*

mixture density:

$$
\begin{aligned}
f_M(x;\theta) &= \sum_{i=1}^{M} w^{(i)} \Big(f_{LN}(x; F^{(i)}, \sigma^{(i)2}\tau) \mathbb{1}_{\{i \le M_1\}} + f_{WB}(x; F^{(i)}, k^{(i)}) \mathbb{1}_{\{i > M_1\}} \Big), \\
s.t. \quad & w^{(i)}, F^{(i)}, \sigma^{(i)}, k^{(i)} > 0, \quad \sum_{i=1}^{M} w^{(i)} = 1, \quad \sum_{i=1}^{M} w^{(i)} F^{(i)} = F_t.
\end{aligned} \tag{4}
$$

where $f_{LN}(x;\mu,\sigma^2)$ and $f_{WB}(x;\mu,k)$ are the density function of a lognormal and a Weibull distribution with the following parametrization, respectively:

$$
f_{LN}(x;\mu,\sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\Big(-\frac{1}{2\sigma^2}(\ln x - \ln\mu + 0.5\sigma^2)^2 \Big), \quad \mu, \sigma^2 > 0 \tag{5}
$$

$$
f_{WB}(x;\mu,k) = \frac{g(x;\mu,k)e^{-g(x;\mu,k)}}{kx}, \quad g(x;\mu,k) = \left(\frac{x\Gamma(1+k)}{\mu} \right)^{1/k}, \quad \mu, k > 0, \tag{6}
$$

in which $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$ is the Gamma function. We collect all the parameters in the $3M$-by-1 parameter vector $\theta = ((w^{(i)})_{i=1:M}, (F^{(i)})_{i=1:M}, (\sigma^{(i)})_{i=1:M_1}, (k^{(i)})_{i=1:M_2})^\intercal$ such that for each type of parameters, the lognormal parameters always precede the Weibull ones.

Remark 3. *Definition 1 suggests that there are $3M-2$ free parameters in an M-LWM density due to the two equality constraints. For theoretical analysis purposes, it is more convenient to drop $w^{(M)}$ and $F^{(M)}$ and consider $\vartheta = ((w^{(i)})_{i=1:M-1}, (F^{(i)})_{i=1:M-1}, (\sigma^{(i)})_{i=1:M_1}, (k^{(i)})_{i=1:M_2})^\intercal$. Clearly, each θ uniquely defines a ϑ and vice versa. We will refer to θ (resp. ϑ) as the full (resp. free) parameter vector in the remainder of the paper.*

We summarize some important properties of the M-LWM density. First, for X_{LN} and X_{WB} having the densities as in Eq. (5) and Eq. (6) respectively, we have:

$$
\begin{aligned}
&\mathrm{E}[X_{LN}] = \mathrm{E}[X_{WB}] = \mu, \mathrm{Var}[X_{LN}] = e^{\sigma^2-1}\mu^2, \mathrm{Var}[X_{WB}] = v(k)\mu^2, \\
&\mathrm{E}[\ln X_{LN}] = \ln\mu - \frac{\sigma^2}{2}, \mathrm{E}[\ln X_{WB}] = \ln\mu - \ln\Gamma(1+k) - k\gamma, \\
&\mathrm{Var}[\ln X_{LN}] = \sigma^2, \mathrm{Var}[\ln X_{WB}] = k^2\pi^2/6.
\end{aligned} \tag{7}
$$

with $v(k) = \frac{\Gamma(1+2k)}{\Gamma(1+k)^2} - 1$. Note that $v(k)$ is monotonically increasing on $[0,1]$ with $v(0) = 0$ and $v(1) = 1$. The parametrization of the Weibull distribution ensures that (μ, k) has the same interpretation as (μ, σ) for the Lognormal distribution. Specifically, the condition $\sum_{i=1}^{M} w^{(i)}F^{(i)} = F_t$ ensures that Assumption 1.(3)-(4) hold by construction. Also, σ and k can be interpreted as the scale parameters of the corresponding log-transformed distributions, as the variance of the normal (resp. Gumbel) distribution is proportional to σ^2 (resp. k^2). By setting $M = M_1$, the M-LWM density reduces to a mixture-of-lognormal (MLN) density as in Ritchey (1990), Melick and Thomas (1997). Similarly,

setting $M_1 = 0$ gives a pure mixture-of-Weibull density.

The M-LWM density implies the following option pricing formula:

$$O(K;\theta) = \sum_{i=1}^{M} w^{(i)} \left(O_{LN}(K; F^{(i)}, \sigma^{(i)}) \mathbb{1}_{\{i \leq M_1\}} + O_{WB}(K; F^{(i)}, k^{(i)}) \mathbb{1}_{\{i > M_1\}} \right), \tag{8}$$

where $O_{LN}(\cdot)$ is the Black's (1976) option pricing formula for options on forward contracts, and additional inputs such as τ and r_f are omitted for brevity. For example, the Black (1976) model for call options is:

$$C_{LN}(K; F, \sigma) = e^{-r_f \tau}[F\Phi(d_1) - K\Phi(d_2)], \tag{9}$$

where $d_1 = \frac{\ln(F/K) + \sigma^2 \tau/2}{\sigma\sqrt{\tau}}$, $d_2 = d_1 - \sigma\sqrt{\tau}$, and $\Phi(\cdot)$ is the CDF of a standard normal distribution. The Weibull-implied option price, $O_{WB}(\cdot)$, can be computed by direct integration (e.g., Eqs. (1) and (2) in Lu and Qu (2021)). For instance, the Weibull-implied call option price takes the following closed-form:

$$C_{WB}(K; F, k) = e^{-r_f \tau}[F\Upsilon_{1+k}(g(K; \mu, k)) - Ke^{-g(K;\mu,k)}], \tag{10}$$

where $\Upsilon_k(z) = \Gamma(k)^{-1} \int_z^\infty t^{k-1} e^{-t} dt$ is the (normalized) upper incomplete Gamma function.

Given the observed option prices $(O_n)_{n=1:N}$, the number of mixtures M and the number of lognormal components M_1, one can estimate the parameter vector ϑ by minimizing the weighted squared pricing error:

$$\begin{aligned} &\widehat{\theta}_N = \underset{\theta}{\operatorname{argmin}}\, Q_N(\theta) := \frac{1}{2N} \sum_{n=1}^{N} \omega_n \varepsilon_n(\theta)^2 \\ \text{s.t.} \quad &w^{(i)}, F^{(i)}, \sigma^{(i)}, k^{(i)} > 0, \quad \sum_{i=1}^{M} w^{(i)} = 1, \quad \sum_{i=1}^{M} w^{(i)} F^{(i)} = F_t. \end{aligned} \tag{11}$$

where $\varepsilon_n(\theta) := O_n - O(K_n; \theta)$ is the pricing error of the M-LWM model at the nth strike grid, and $\{\omega_n\}_{n=1:N}$ is a set of positive regression weights. The optimization problem in Eq. (11) is a constrained weighted nonlinear least square (WNLS) problem, which is typically solved numerically by standard optimization algorithms. The estimated parameter vector $\widehat{\theta}_N$ directly provides an estimate of the RND, i.e., $f_M(x; \widehat{\theta}_N)$.

As to the regression weights in Eq. (11), different weighting schemes are proposed in the literature. The simplest choice is $\omega_n = 1, \forall n$, which corresponds to an equal weighting scheme. Aït-Sahalia and Duarte (2003) assign higher weights for options traded more liquidly and propose to choose ω_n based on open interests or the bid-ask spreads. Christoffersen et al. (2018) set ω_n based on the implied Vega, so that $Q_n(\theta)$ approximates the mean squared error in the implied volatility domain.[4]

[4] Christoffersen et al. (2018) suggest to set $\omega_n = 1/v(K_n)^2$, where $v(K_n)$ is the option Vega computed from the implied

Before deriving asymptotic properties for $\widehat{\theta}_N$, we provide some motivations for adding Weibull mixtures to the classic MLN density, which is a key innovation of this paper. The use of the Weibull distribution to model option prices can be traced back to Savickas (2002), who noticed that the empirical RND can be better captured by a negatively skewed[5] Weibull distribution, while a lognormal density is always positively skewed. More recent evidence in Andersen et al. (2017, 2021a), Lu and Qu (2021) also suggest that the empirical RND typically has a heavy negative left tail, generating a negative skewness. As a result, the Weibull components in the LWM density can fit the left tail of the empirical RND much more efficiently than a pure MLN density. In fact, one can easily show[6] that $f_{WB}(x)$ has the same left tail behaviour as a type III extreme value distribution, which is used in Figlewski (2010) to fit the left tail of the empirical RND.

A graphical illustration of Weibull RNDs, the associated call prices and implied volatilities are presented in Fig. 2.1. The choices of the scaling parameter $k \in \{0.15, 0.2776, 0.4\}$ represent a negative, neutral, or positive skewness, respectively. Interestingly, the implied volatility (IV) curves slope downwards for all choices of k. This provides a better fit to the empirical volatility smile or smirk pattern than a flat IV curve implied by the lognormal RND, which is consistent with Savickas (2002).

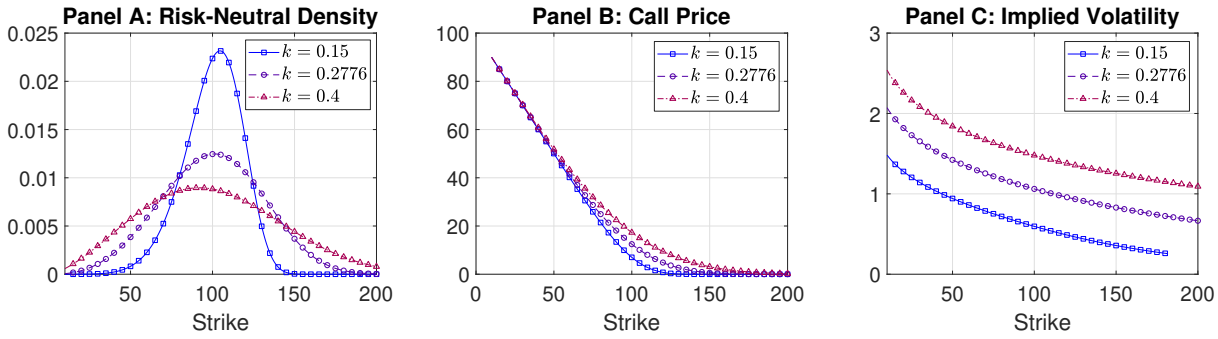

Figure 2.1: Examples of Weibull densities $f_{WB}(x)$, the associated call prices $C_{WB}(K; F, k)$ and the corresponding Black model implied volatility. The parameters used to generate the call prices and the implied volatilities are: $F = 100$, $r_f = 0$, $\tau = 22/252$.

To demonstrate the importance of the Weibull components in the LWM mixture for empirical RND extraction, we simulate two RNDs with opposite skewness coefficients from a 3-factor stochastic volatility (3FSV) model and their corresponding true option prices (see DGP I and DGP III in Section 4 and Online Appendix C for details). We fit a 3-LWM density by solving Eq. (11) with all possible choices of M_1 and present the fitted densities in Fig. 2.2. The left panel of the figure reveals that the 3-MLN density has a poor fit to the heavy left tail of the RND, which is generated by a slow-decaying

volatility at strike grid K_n. We thank a referee for pointing this out.

[5]The skewness coefficient of the Weibull distribution in Eq. (6) takes the explicit form $\frac{2\Gamma(1+k)^3-3\Gamma(1+k)\Gamma(2+k)+\Gamma(1+3k)}{(\Gamma(1+2k)-\Gamma(1+k)^2)^{3/2}}$, which is independent from F and is negative whenever $k < 0.2775975$.

[6]Note that with the linear change of variable $-ky = x\Gamma(1+k)$, $f_{WB}(x)$ is proportional to a generalized extreme value distribution $f_{GEV}(y)$ with scale parameter μ, shape parameter $-k$, and location parameter $-\mu/k$.

jump process. Replacing more lognormal components by Weibull components progressively improves the RND fit, and in this case the pure mixture-of-Weibull density has the best overall fit. The right panel of the figure tells the opposite story: a pure mixture-of-Weibull density can hardly fit the heavy right-tail of the RND, while a 3-LWM density with $M_1 = 2$ provides the overall smallest squared error.

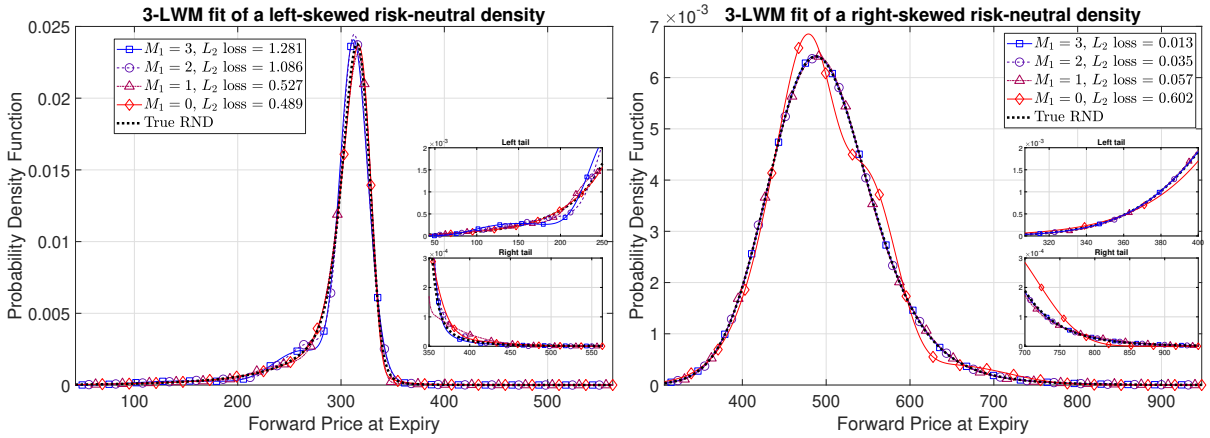

Figure 2.2: Examples of RND estimates based on the 3-LWM density. The left-skewed (resp. right-skewed) true RND is generated from DGP I (resp. DGP III) of a 3FSV model in Section 4 and Online Appendix C. For each choice of $M_1 \in \{0, 1, 2, 3\}$, we estimate the 3-LWM parameters by solving Eq. (11) based on the true option prices with an equal weighting scheme. The option prices are computed on an equidistant strike grid with 100 observations. The L_2 loss is defined as $\sqrt{\int_{\mathcal{K}} |\widehat{f}(x) - f^*(x)|^2 dx} \times 10^2$, where $\widehat{f}(x)$ is the fitted 3-LWM density and $f^*(x)$ is the true RND, respectively.

Fig. 2.2 clearly shows the superior flexibility of the 3-LWM density over a pure 3-MLN density, as vast mean square improvements can be achieved by appropriate mixing of Weibull and lognormal components. Changing the number of lognormal and Weibull mixtures of the 3-LWM density adjusts adaptively to the tail shapes of the RND, but importantly, without introducing additional parameters. By increasing the number of mixtures, a general M-LWM density is expected to provide more accurate approximations to highly complex RNDs. It is worth mentioning that the number of parameters involved in the finite LWM density is usually much less than a (semi)-nonparametric RND estimator for a comparable approximation quality. This leads to a significant gain in precision for the parametric LWM method, on which we elaborate in our simulation and empirical analyses.

3 Main Results

Our main results can be divided into three parts: First, we establish consistency and asymptotic normality of the estimator $\widehat{\theta}_N$ in Eq. (11) as $N \to \infty$ under the assumption of a misspecified model. Second, we propose a method to choose the number of lognormal mixtures M_1 and the number of mixtures M empirically. Last, we explain computational details and conclude with some further discussions of the LWM method.

3.1 Asymptotic Properties of the LWM-Based RND Estimates

The econometric problem in Eq. (11) belongs to the general class of constrained WNLS regressions with dependent and heterogeneous innovations, which is a very well-studied problem in the literature. Asymptotic results have been established in much more general settings than ours, e.g. Jennrich (1969), White (1980), Wu (1981), Domowitz and White (1982), White and Domowitz (1984), Andrews (1987), Pötscher and Prucha (1991a,b) for the unconstrained case, and Geyer (1994), Wang (1996), Andrews (1999), Wang (2000, 2004) for the constrained case. By applying these asymptotic results to our specific setting, we derive asymptotic properties for the parameter estimates $\widehat{\theta}_N$ in Eq. (11). However, such application is not entirely trivial as the convergence of $\widehat{\theta}_N$ cannot be established in the usual sense.

We begin with a series of assumptions and definitions that clarify the asymptotic setting of our results.

Assumption 4. *For any fixed integers $1 \leq M$, we assume that $\theta \in \Theta_M$, where Θ_M is a compact subset of $\mathbb{R}^{3M}$ which satisfies $\forall i$, $w^{(i)} \in [0,1]$, $\sum_{i=1}^{M} w^{(i)} = 1$, $\sum_{i=1}^{M} w^{(i)} F^{(i)} = F_t$, and there exist universal constants $0 < \underline{S} < \overline{S}$ such that $\sigma^{(i)}, k^{(i)}, F^{(i)} \in [\underline{S}, \overline{S}]$, $\forall i$.*

Remark 4. *The compactness of Θ_M is a common requirement to ensure the consistency of parameter estimates. In our case, since $\underline{S}$ (resp. $\overline{S}$) can be set to be arbitrarily small (resp. large), in practice we only need to set it to cover reasonable parameter choices and the compactness of Θ_M is guaranteed. This also insures that $\mathscr{F}_M$ only contains smooth density functions on $\mathcal{K} \times \Theta_M$, as all $\sigma^{(i)}$s and $k^{(i)}$s are bounded away from 0. We do not specify M_1 here, as the parameter bounds are common for both the lognormal and the Weibull mixtures.*

Remark 5. *In view of Remark 3, we have $\vartheta \in \Theta_M$, where Θ_M is a compact subset of Θ_M, understood as the space of the free parameter vector. It is also clear that there exists a bijection $T : \Theta_M \mapsto \Theta_M$ such that $T(\vartheta) = \theta$ with the inverse function $T^{-1}(\theta) = \vartheta$ that generates a unique free parameter vector from a full parameter vector. As a result, it suffices to prove the consistency and asymptotic normality of the WNLS estimator $\widehat{\vartheta}_N = T^{-1}(\widehat{\theta}_N)$ in the free parameter space, which is the objective of this section.*

The main challenge of establishing the desired convergence is that $O(K;\theta)$ is only uniquely define on Θ_M up to a permutation of the parameters that belong to mixtures of the same type. In particular, one can easily construct a sequence of $\widehat{\theta}_N$ that does not converge to any element in Θ_M but still minimizes $Q_N(\theta)$ as $N \to \infty$, since the minimizer is not unique in general. To resolve this issue, we introduce the notion of a quotient space of the equivalence class of parameters:

Definition 2. *For some natural number N, let P_N denote a N-by-N permutation matrix such that for any $v \in \mathbb{R}^{N\times 1}$, $P_N v$ is a permuted version of v. Let $\mathcal{P}_N$ collects all $N!$ unique permutation matrices.*

With $M_1 + M_2 = M$, pick $P_{M_1} \in \mathcal{P}_{M_1}$ and $P_{M_2} \in \mathcal{P}_{M_2}$ and construct the $3M$-by-$3M$ block-diagonal matrix $\Pi_M = diag(P_{M_1}, P_{M_2}, P_{M_1}, P_{M_2}, P_{M_1}, P_{M_2})$. Let $\mathbf{\Pi}_M$ collects all $M_1! \cdot M_2!$ unique Π_M formed by exhausting different choices of P_{M_1} and P_{M_2}. For all $\theta \in \Theta_M$, define an equivalence relation $\sim_p$ on Θ_M such that $\theta \sim_p \theta_p$ iff:

$$\theta_p = \Pi_M \theta \in \Theta_M, \exists \Pi_M \in \mathbf{\Pi}_M. \tag{12}$$

The equivalence class $[\theta]$ is defined as: $[\theta] = \{\theta_p \in \Theta_M : \theta \sim_p \theta_p\}$.

Denote the set $\Theta_M \backslash \sim_p$ as the collection of all equivalence classes on Θ_M, and let $\pi : \Theta_M \mapsto \Theta_M \backslash \sim_p$ denote the quotient map which maps θ into $[\theta]$. The quotient space $\widetilde{\Theta}_M$ is defined by the set $\Theta_M \backslash \sim_p$ with the topology $\{U \subset \widetilde{\Theta}_M : \pi^{-1}(U)$ is open in $\Theta_M\}$.

A function $g : \Theta_M \mapsto Y$ is called permutation invariant if for every $\theta \in \Theta_M$, it holds that $g(\theta) = g(\theta_p)$ for all $\theta_p \in [\theta]$. We denote $\mathcal{G}$ as the set of all permutation invariant functions. Every $g \in \mathcal{G}$ has a well-defined dual function $\widetilde{g} : \widetilde{\Theta}_M \mapsto Y$ on the quotient space, defined by $\widetilde{g}([\theta]) = g(\theta)$.

Remark 6. *In view of Remark 3, for the free parameter vector $\vartheta = T^{-1}(\theta)$, we shall define the equivalent class $[\vartheta] := \{\vartheta \in \Theta_M : T^{-1}(\theta) \sim_p \theta\}$, and the corresponding quotient parameter space $\widetilde{\Theta}_M$ is defined analogously. Note that $T^{-1} \notin \mathcal{G}$, as different elements of θ are removed after a permutation. Therefore, functions g and $\widetilde{g}$ defined on Θ_M and $\widetilde{\Theta}_M$ can also be generalized to take value in Θ_M and $\widetilde{\Theta}_M$ by a composition of g and T without loss of generality. We shall also abuse the notation and write $g(\vartheta) := g(T(\vartheta))$ and $\widetilde{g}([\vartheta]) := \widetilde{g}([T(\vartheta)])$ to avoid notational clutter whenever no confusion is caused.*

Remark 7. *Intuitively, $[\theta]$ contains all parameter vectors formed by permuting the orders of the parameters that belong to the same distribution type in θ. Clearly, all elements in $[\theta]$ define the same M-LWM density $f_M(K;\cdot)$, which also implies the same pricing function $O(K;\cdot)$ and the weighted squared loss $Q_N(\cdot)$, i.e., they belong to $\mathcal{G}$ by construction. The case with $M_1 M_2 = 0$ can be constructed analogously by removing the corresponding permutation matrix from Π_M. Importantly, the minimizer of $Q_N(\vartheta)$ on Θ_M is not unique due to permutation, but one can construct a unique minimizer $[\vartheta]$ of $\widetilde{Q}_N([\vartheta])$ on $\widetilde{\Theta}_M$, which allows for the identification of the true parameter vector. Therefore, we establish consistency and asymptotic normality of parameter estimates in the sense of equivalence classes, such that $[\widehat{\vartheta}_N]$ converges to $[\vartheta^*]$ on $\widetilde{\Theta}_M$ for some $\vartheta^* \in \Theta_M$. Note that this sense of convergence is also applicable to models possessing a label switching problems, e.g. mixture models or regime-switching models.*

We impose the following regularity assumption on the regression weights:

Assumption 5. *The weights $\{\omega_n\}_{n=1:N}$ are generated by a strictly positive, bounded and Hölder continuous function $\omega(x)$ such that $\omega_n = \omega(K_n)$.*

The above assumption allows arbitrary positive bounded and fixed weights in finite sample, which covers the aforementioned choices of weights considered in the literature by a conditioning argument.

The Hölder continuity of $\omega(x)$ plays the same role as that of $\gamma(x)$. Stochastic and exogenous weights can be accommodated by generalizing Assumption 5 in view of White (1980, 1981), Domowitz and White (1982), which does not fundamentally change the asymptotic analysis of this paper and is thus not considered here.

We impose an identification assumption about the (pseudo) true parameter vector ϑ^* in the setting that the M-LWM model is in general misspecified:

Assumption 6. *Suppose Assumptions 1-5 hold true. Consider an M-LWM density with M_1 lognormal components for some fixed M and M_1. Choose a weighting function ω and define $\widetilde{Q}([\vartheta])$ as the following definite integral, understood as the limiting integrated squared weighted pricing error of the M-LWM model on $\mathcal{K}$:*

$$\widetilde{Q}([\vartheta]) := \int_{\mathcal{K}} (O^*(K) - \widetilde{O}(K;[\vartheta]))^2 \omega(K) d\kappa(K). \tag{13}$$

We assume that $\widetilde{Q}([\vartheta])$ has a unique minimizer $[\vartheta^*]$ in the interior of $\widetilde{\Theta}_M$.

Remark 8. *Assumption 6 rules out the ill-behaved case where two (or more) mixtures of the same type have identical scale and shape parameters, in which case the weight parameters are not uniquely identified. Also, the interior requirement of $[\vartheta^*]$ and Eq. (11) eliminates the boundary solutions to the optimization problem, e.g., $w^{(i)} = \{0, 1\}$, or $F^{(i)}, \sigma^{(i)}, k^{(i)} \in \{\underline{S}, \overline{S}\}$. This ensures that we can transform the constrained WNLS problem in Eq. (11) into an unconstrained problem as the inequalities for the true free parameter vector are not binding. It is worth noting that such an identification assumption is not required for the sieve method of Lu and Qu (2021), which is replaced a diverging (instead of fixed) strike range $\mathcal{K}$ to allow for the identification of the entire RND function. In our fixed strike range setting, Assumption 6 only allows for parameter identification of RND on $\mathcal{K}$, and the RND tails outside $\mathcal{K}$ are unidentified. Please see Remark 12 for further details.*

Remark 9. *Eq. (13) states that $\widetilde{O}(K;[\vartheta^*])$ can be interpreted as the best WNLS approximation of the observed option prices based on the M-LWM model with M_1 lognormal components and M_2 Weibull components. In the general misspecified case, the value of $[\vartheta^*]$ depends on the choice of $\omega(x)$. However, in the special case where the M-LWM model correctly specifies the DGP with $O^*(K) \equiv \widetilde{O}(K;[\vartheta^*]), \forall K \in \mathcal{K}$, then it is clear that different choices of $\omega(x)$ do not alter $[\vartheta^*]$.*

We present the consistency and asymptotic normality of $\widehat{\vartheta}_N$ in the results below:

Theorem 1. *Under Assumptions 1-6, it holds as $N \to \infty$ that $[\widehat{\vartheta}_N] \xrightarrow{a.s.} [\vartheta^*]$.*

Theorem 2. *Under the same condition as in Theorem 1, as $N \to \infty$ we have:*

$$[\widehat{\vartheta}_N] \xrightarrow{d} [\vartheta^* + N^{-1/2}(\boldsymbol{A}^*)^{-1}\xi], \tag{14}$$

where $\xi \sim \mathcal{N}(0, \boldsymbol{B}^*)$ *is a* $(3M-2)$*-by-1 dimensional multivariate normal vector, and the expressions for the* $(3M-2)$*-dimensional positive definite matrices* $\boldsymbol{A}^*$ *and* $\boldsymbol{B}^*$ *can be found in the proof.*

Intuitively, the results in Theorems 1 and 2 suggest that, as $N \to \infty$, $\widehat{\vartheta}_N$ may fail to converge towards a particular ϑ^*, since the optimizer of Eq. (11) is not unique in Θ_M. However, $\widehat{\vartheta}_N$ must belong to a sequence which converges to one of the elements in $[\vartheta^*]$, say ϑ_0^* (see Lemma A.1(d)). Theorem 2 then ensures that asymptotically we must have:

$$\sqrt{N}(\widehat{\vartheta}_N - \vartheta_0^*) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{C}^*), \quad \boldsymbol{C}^* := (\boldsymbol{A}^*)^{-1}\boldsymbol{B}^*(\boldsymbol{A}^*)^{-1}. \tag{15}$$

The sandwich form of $\boldsymbol{C}^*$ is standard in the literature of unconstrained WNLS problems (see, e.g., White (1980), White and Domowitz (1984)). The above result is not feasible as both $\boldsymbol{A}^*$ and $\boldsymbol{B}^*$ are limiting quantities that need to be estimated. Following Domowitz and White (1982), the following estimator of $\boldsymbol{A}^*$ is consistent:

$$\widehat{\boldsymbol{A}} = \frac{1}{N}\sum_{n=1}^{N} \omega_n \Big(\nabla O(K_n; \widehat{\vartheta}_N) \nabla O(K_n; \widehat{\vartheta}_N)^\intercal - \nabla^2 O(K_n; \widehat{\vartheta}_N) \widehat{\varepsilon}_n \Big), \tag{16}$$

where $\nabla O(K_n; \widehat{\vartheta}_N)$ and $\nabla^2 O(K_n; \widehat{\vartheta}_N)$ are the $(3M-2)$-by-1 gradient vector and the Hessian matrix of $O(K_n; \vartheta)$ w.r.t. ϑ evaluated at $\widehat{\vartheta}_N$, and $\widehat{\varepsilon}_n := \varepsilon_n(\widehat{\vartheta}_N)$ is the nth fitted residual of Eq. (11). A positive semi-definite estimator of $\boldsymbol{B}^*$ can be constructed can be constructed using the simple Newey and West (1987) estimator with the Bartlett kernel:

$$\widehat{\boldsymbol{B}} = \widehat{\boldsymbol{\Omega}}_0 + \sum_{j=1}^{h} \Big(1 - \frac{j}{h}\Big)(\widehat{\boldsymbol{\Omega}}_j + \widehat{\boldsymbol{\Omega}}_j^\intercal), \tag{17}$$

in which $\widehat{\boldsymbol{\Omega}}_j := \sum_{n=j+1}^{N} \omega_n \omega_{n-j} \widehat{\varepsilon}_n \widehat{\varepsilon}_{n-j} \nabla O(K_n; \widehat{\vartheta}_N) \nabla O(K_{n-j}; \widehat{\vartheta}_N)^\intercal / N$, and h is a suitably chosen[7] truncation lag. As discussed in Domowitz and White (1982), $\widehat{\boldsymbol{B}}$ consistently estimates $\boldsymbol{B}^*$ when the M-LWM model correctly specifies the DGP on the strike range, i.e., $O^*(K) \equiv O(K; \vartheta^*), \forall K \in \mathcal{K}$, but is in general inconsistent in the misspecified case. However, this does not prevent us from conducting valid statistical inference, as the correctness of the model is usually assumed in the null hypothesis of statistical tests.

The asymptotic normality established in Theorem 2 allows us to construct confidence bounds for any continuous and permutation invariant function of ϑ_N via the Delta method:

Corollary 1. *Let* $g : \Theta_M \mapsto \mathbb{R}^k$ *be an element in* $\mathcal{G}$ *with continuous and non-zero first-order derivative.*

[7] In simulation and empirical applications, we use $h = \lfloor 0.75 N^{1/3} \rfloor$ following Andrews (1991).

Under the same condition as in Theorem 1, it holds that:

$$\begin{aligned} g(\widehat{\vartheta}_N) &\xrightarrow{a.s.} g(\vartheta^*), \\ \sqrt{N}(g(\widehat{\vartheta}_N) - g(\vartheta^*)) &\xrightarrow{d} \mathcal{N}(0, \boldsymbol{J}_g(\vartheta^*)\boldsymbol{C}^*\boldsymbol{J}_g(\vartheta^*)^\intercal), \end{aligned} \tag{18}$$

where $\boldsymbol{J}_g(\vartheta^*)$ is the k-by-$(3M-2)$ Jacobian matrix of g w.r.t. ϑ evaluated at ϑ^*.

Remark 10. *The requirement that $g \in \mathcal{G}$ eliminates the permutation dependence of the limiting distribution (as in Theorem 2) by construction. Similar to the asymptotic variance of $\widehat{\vartheta}_N$, one can estimate the asymptotic variance of $g(\widehat{\vartheta}_N)$ consistently via the plug-in method when the M-LWM model is correctly specified. This allows us to estimate the confidence bounds of the LWM-based RND estimate pointwise for each $x \in \mathbb{R}^+$ by setting $g(\vartheta) = f_M(x;\vartheta)$. However, as $\boldsymbol{C}^*$ can only be consistently estimated under the correct model specification which is unlikely to hold in practice, one should interpret the estimated confidence bounds with caution (further discussions are provided in the simulation section). More importantly, it allows us to construct specification tests to evaluate the goodness-of-fit of the M-LWM model which implicitly assumes a correct model specification. We shall elaborate on this point in the following section.*

3.2 Choosing the Number of Mixtures

To implement the LWM approach in practice, one needs to choose both the number of mixtures M and the number of lognormal components M_1. The two choices are fundamentally quite different problems, which we shall elaborate on in this section.

Given a fixed M, choosing M_1 is relatively straightforward. As depicted in Figure 2.2, different choices of M_1 alter the tail shapes of the M-LWM density, but do not increase its overall flexibility since the number of parameters does not change. As a result, one can simply perform an exhaustive search and minimize Eq. (11) across all different choices of $M_1 \in \{0, \dots, M\}$. This method is computationally costly as it requires $M+1$ numerical solutions of the WNLS problem. A computation-wise more efficient approach is presented at the end of this section, which chooses both M and M_1 simultaneously based on a sequential testing procedure. Note that in finite samples, the choice of M_1 computed from noisy option prices is not necessarily the same as the limiting optimal choice, i.e., M_1^* that minimizes the weighted integrated squared pricing error in Eq. (13) over all different M-LWM densities. This scenario can be considered as a specific form of model misspecification, which is covered by our asymptotic results in Theorems 1 and 2.

The choice of M, on the other hand, requires more elaboration. Heuristically, a larger M always improves the flexibility of the candidate RND, however, with the risk of overfitting the empirical

data. In this paper, we take a conventional parametric view and choose M as the most parsimonious M-LWM model that provides a good fit to the observed data without any obvious patterns in the residuals. To this end, we propose a sequential testing procedure to check for misspecification of an existing M-LWM model. We add mixtures to the model sequentially until the specification test is no longer rejected, which gives the desired choice of M.

Given some choices of M and M_1, we propose a simple lack-of-fit test in the spirit of Gallant (1977) to determine whether this M-LWM model is the correct DGP that generates the efficient option prices on the strike range, i.e., $O^*(K) \equiv O(K;\vartheta^*), \forall K \in \mathcal{K}$, which is our null hypothesis. In detail, we test whether there are any unexplained structures in the WNLS regression residuals. To this end, suppose that we have estimated this M-LWM model using some weights $\{\omega_n\}_{n=1:N}$. We obtained the estimated parameter vector $\widehat{\vartheta}_N$ and the weighted WNLS residuals $\widehat{\boldsymbol{\epsilon}} = \{\widehat{\epsilon}_n\}_{n=1:N}$ with $\widehat{\epsilon}_n := \sqrt{\omega_n}\widehat{\varepsilon}_n$. Consider the following auxiliary fast Fourier regression with some fixed $p, q \in \mathbb{N}$:

$$\widehat{\epsilon}_n = \sum_{i=0}^{p} \beta_i \widetilde{K}_n^i + \sum_{j=1}^{q} \big(\gamma_{s,j} \sin(2\pi \widetilde{K}_n) + \gamma_{c,j} \cos(2\pi \widetilde{K}_n)\big) + e_n, \tag{19}$$

where $\widetilde{K}_n := (K_n - F_t)/m(\mathcal{K})$ is the moneyness normalized to a unit subinterval of $[-1, 1]$, and p, q are the corresponding degrees for the polynomial and the Fourier series. A HAC-robust Wald test can be constructed by testing the explanatory power of the auxiliary regression based on the following result:

Proposition 1. *Write the auxiliary regression in Eq. (19) in matrix form:*

$$\widehat{\boldsymbol{\epsilon}} = \underbrace{\boldsymbol{X}}_{N\times(p+2q+1)} \underbrace{\boldsymbol{\beta}}_{(p+2q+1)\times 1} + \boldsymbol{e},$$

and let $\widehat{\boldsymbol{\beta}}$ denote the least square estimator of $\boldsymbol{\beta}$. Under the null hypothesis that the M-MLN model is correct on $\mathcal{K}$, it holds as $N \to \infty$ that:

$$\text{Wald}_{p,q} := N \widehat{\boldsymbol{\beta}}^{\intercal} \widehat{\boldsymbol{V}}^{-1} \widehat{\boldsymbol{\beta}} \sim \chi^2_{p+2q+1}, \tag{20}$$

where $\widehat{\boldsymbol{V}}$ is the following consistent and positive semi-definite estimator of $\text{avar}(\sqrt{N}\widehat{\boldsymbol{\beta}})$*:*

$$\widehat{\boldsymbol{V}} = \widehat{\boldsymbol{V}}_0 + \sum_{j=1}^{h} \Big(1 - \frac{j}{h}\Big)(\widehat{\boldsymbol{V}}_j + \widehat{\boldsymbol{V}}_j^{\intercal}), \tag{21}$$

in which $\widehat{\boldsymbol{V}}_j := \sum_{n=j+1}^{N} \widehat{\epsilon}_n \widehat{\epsilon}_{n-j} \widetilde{\boldsymbol{X}}_n(\widehat{\vartheta}_N) \widetilde{\boldsymbol{X}}_{n-j}(\widehat{\vartheta}_N)^{\intercal}/N$ with $h = \lfloor 0.75 N^{1/3} \rfloor$, and $\widetilde{\boldsymbol{X}}_n$ is the nth column of the $(p+2q+1)$-by-N matrix:

$$(\boldsymbol{X}^{\intercal}\boldsymbol{X})^{-1}\boldsymbol{X}^{\intercal}(N \cdot \boldsymbol{I}_N - \boldsymbol{W}_N^{1/2} \boldsymbol{J}_O(\widehat{\vartheta}_N) \widehat{\boldsymbol{A}}^{-1} \boldsymbol{J}_O(\widehat{\vartheta}_N)^{\intercal} \boldsymbol{W}_N^{1/2}),$$

where $\boldsymbol{J}_O(\widehat{\vartheta}_N) = [\nabla O(K_1;\widehat{\vartheta}_N),\ldots,\nabla O(K_N;\widehat{\vartheta}_N)]^\intercal$ *is the* N*-by-*$(3M-2)$ *Jacobian matrix (see Proposition 2), and* $\boldsymbol{W}_N$ *is the* N*-by-*N *diagonal matrix formed by the weights* $\{\omega_n\}_{n=1:N}$.

Remark 11. *Under the null hypothesis and from Assumption 3,* $\widehat{\epsilon}_n$ *is asymptotically a linear transformation of* u_n*, thus asymptotically* $\mathrm{E}[\widehat{\epsilon}_n|K_n] = 0, \forall n$ *and the fast Fourier regression in Eq. (19) should have no explanatory power. Therefore, a rejection of the Wald test suggests that there are still certain patterns in the residuals left unexplained by the* M*-LWM implied option pricing function, which indicates that we should increase* M *and consider a more flexible family of RNDs. To perform the test in practice, one needs to decide the orders of* p *and* q *for the auxiliary regression. As the Fourier regression can approximate any function with a sufficiently large order, we need to choose small* p *and* q *relative to* N *to avoid over-fitting (and hence spurious rejection of the test). We provide guidance on the choices of* p *and* q *via simulation in Section 4.*

Remark 12. *It is worth pointing out that the null hypothesis of the test does not rule out any misspecification of the* M*-LWM model outside the strike range. Because, under Assumption 2 with a fixed strike range, the shape of the RND tails outside* $\mathcal{K}$ *cannot be uniquely identified based on the observed option prices in* $\mathcal{K}$*. To see this, we note that two arbitrary well-defined RNDs* $f_1(x)$ *and* $f_2(x)$ *with* $f_1(x) \equiv f_2(x), \forall x \in \mathcal{K}$ *imply identical option prices on* $\mathcal{K}$ *iff the following equalities hold for* $k \in \{0,1\}$*:*

$$\int_0^K x^k f_1(x)dx = \int_0^K x^k f_2(x)dx, \quad \int_K^\infty x^k f_1(x)dx = \int_K^\infty x^k f_2(x)dx, \tag{22}$$

which can be derived from the option pricing functions given the RND (see for example, Eqs.(1) and (2) in Lu and Qu (2021)). Eq. (22) allows us to replace the fitted M*-LWM tails with arbitrary tails from other distributions (such as the GEV tails proposed by Figlewski (2018)) outside the strike range without changing the fitted option prices.*

With the proposed specification test, we suggest the following scheme to iteratively choose an appropriate M and the associate choice of M_1 (denote by $M_1^{(M)}$) simultaneously. To this end, we shall fix the regression weights $\{\omega_n\}_{n=1:N}$ and denote $Q_N^{(M)}(\vartheta_N; M_1)$ as the weighted squared regression error evaluated at ϑ with M mixtures and M_1 lognormal components.

The upper bound $\overline{M}$ sets the maximum number of mixtures for the M-LWM model. In our simulation and empirical analysis, we set $\overline{M} = 5$, although we find that $\overline{M} = 4$ is usually sufficiently flexible for unimodal RNDs. The sequential test is only intended to eliminate severe form of misspecification caused by a small M, but the overall flexibility of the model is still capped by $\overline{M}$. Importantly, we do not allow M to diverge to infinity and only consider a finite number of mixtures.

Note that the initialization step is simply the exhaustive search method discussed in the beginning of the section for $M = 1$. One also can replace the branching step by an exhaustive search, i.e., take

A sequential testing procedure to choose M and M_1

Step 1: (initialization) Set $M = 1$. Choose $M_1^{(1)} = \text{argmin}_{M_1 \in \{0,1\}} Q_N^{(1)}(\widehat{\vartheta}_N; M_1)$.
Step 2: (branching) Set $M = M + 1$ and choose $(\widehat{\vartheta}_N, M_1^{(M)})$ as:

$$(\widehat{\vartheta}_N, M_1^{(M)}) = \underset{M_1 \in \{M_1^{(M-1)}, M_1^{(M-1)}+1\}, \vartheta \in \Theta_M}{\text{argmin}} Q_N^{(M)}(\vartheta; M_1). \tag{23}$$

Step 3: (testing) For the estimated M-LWM model with $M_1^{(M)}$ lognormal mixtures, choose some p, q and a significance level $\alpha\%$. Perform the auxiliary regression test in Proposition 1.
Step 4: (iteration) If the test in Step 3 is rejected at $\alpha\%$ and $M < \overline{M}$, go to Step 2. Otherwise, terminate the algorithm with the current M and $M_1^{(M)}$.

the argmin in Eq. (23) over the set $\{0, \ldots, M\}$. The current branching step is computationally more appealing as only two WNLS regressions are solved for each M. Intuitively, the branching step adds a lognormal or a Weibull component to the existing M-LWM density depending on the value of the objective function. Although the chosen $M_1^{(M)}$ here may not be the optimal over all $M + 1$ possible values, this procedure is guaranteed to produce a nesting sequence of M-LWM models. This allows us to exploit the estimation results in previous steps and solve the WNLS regressions in an evolutionary manner, which is detailed in the following section.

3.3 Computational Concerns

In practice, one cannot implement the LWM method if Eq. (11) cannot be optimized reliably. Indeed, Bondarenko (2003) points out that, even for the pure MLN model, the optimization problem in Eq. (11) is non-trivial to solve. Standard gradient-based local optimizers may converge to different local minima, and the resulting RND estimates may have large spurious spikes due to overfitting. As a result, the existing literature only considers the MLN model with $M \leq 3$ due to the difficulty in estimating the parameter vector.

To solve the computation issues of implementing the LWM method, we develop several important features in optimising Eq. (11), allowing us to estimate the LWM model with a relatively larger M in a computationally efficient manner. First, to alleviate the problem of starting point dependency, in Online Appendix B.1 we develop a global search algorithm which escapes from a local minimum by re-solving Eq. (11) using multiple random draws of starting points. It incorporates an 'evolutional' feature similar to the implementation of the 3-MLN approach in Bondarenko (2003), which fits the M-LWM density using the optimal $(M-1)$-LWM density parameters as initial values. As the M-LWM density nests the $(M-1)$-LWM density, we use the optimized objective function value $Q_N^{(M-1)}(\widehat{\vartheta}_N; M_1^{(M-1)})$ as an upper bound for the objective function value in the branching step. This ensures that adding one mixture to the $(M-1)$-LWM density always improves its fit due to the nesting feature, which

further stabilizes the estimation procedure.

Inevitably, the global search algorithm requires multiple local searches which is computationally expensive. To improve the efficiency of the local search algorithms, we derive analytical gradient and Hessian for Eq. (11) in Online Appendix B.2, which can be expressed as a weighted average of option Greeks for each mixture component. To the best of our knowledge, we are among the first to document such analytical expressions.[8] We also acknowledge that the LWM approach is in general much slower than (semi)-nonparametric approaches that solve convex optimization problems (e.g., Aït-Sahalia and Duarte (2003), Bondarenko (2003), Bliss and Panigirtzoglou (2004), Lu and Qu (2021)) in exchange for a precise parsimonious parametric RND estimator.

Finally, the estimated M-LWM density can contain spurious spikes when M is relatively large (e.g., $M \geq 4$) due to very small scale parameter estimates for some mixtures, which can be caused by a small N and large observation errors. A simple way to eliminate the spurious spikes is to reduce M until the spurious spike disappears. However, the resulting LWM model may not be flexible enough due to a reduction in M, which is undesirable in practice. To resolve this issue, we introduce a variance constraint in the spirit of Hathaway (1985) to effectively remove spurious spikes while preserving most of the flexibility of a M-LWM model. This is achieved by solving Eq. (11) with the following additional linear parameter constraint:

$$\min_{i=1:M} \widetilde{k}^{(i)} \geq c \max_{i=1:M} \widetilde{k}^{(i)}, \quad \widetilde{k}^{(i)} = \begin{cases} \sigma^{(i)}\sqrt{6\tau}/\pi, & 1 \leq i \leq M_1 \\ k^{(i)}, & M_1 + 1 \leq i \leq M \end{cases}. \tag{24}$$

From Eq. (7) it is not hard to see that $\widetilde{k}^{(i)}$ translates $\sigma^{(i)}$ to the corresponding $k^{(i)}$ parameter for a Weibull distribution with the same variance of the log-transformed density. The constant c takes a value in the interval $[0, 1]$ and is the ratio of the smallest component standard deviation to the largest one. Larger (resp. smaller) c imply a more (resp. less) stringent constraint, as Eq. (24) is inactive when $c = 0$, while $c = 1$ forces all the components to have equal variance. In essence, the constraint in Eq. (24) restricts the parameter space of the M-LWM model, which rules out local maximizers whose component variances (captured by $\widetilde{k}^{(i)}$) are vastly different, hence removing the spurious spikes due to a very small $\widetilde{k}^{(i)}$.

As a general principle for the choice of c in practice, an ideal c should be large enough so that the spurious spikes are removed, while small enough so that Eq. (7) does substantially restrict the flexibility of the M-LWM density. In our simulation, we consider up to 5 mixtures which may overfit the noisy prices, and some c in the range of $[0.05, 0.15]$ can very effectively remove the spurious spikes.

[8] Melick and Thomas (1997) claim to have used the analytical gradient for the lognormal-implied option prices in their optimization procedures without revealing the detailed expressions. The Weibull-implied option Greeks are not documented in the literature.

Our simulation results show that the performance of the RND estimates is fairly robust to different choices of c in this range, as it only removes the very small $\widetilde{k}^{(i)}$ that causes a spurious spike, but otherwise has very little impact to the overall flexibility of the LWM density. We recommend the choice of $c = 0.1$ in empirical applications, but we find that Eq. (24) is often not binding in practice.

3.4 Further Discussions

We conclude this section by making several remarks about the parametric LWM approach and discussing the various similarities and differences compared to the popular (S)NP approaches.

First, we point out that our theoretical results can be generalized to cover a mixture of arbitrary parametric densities defined on the positive real line[9] (e.g., Gamma distribution, F-distribution, log-t distribution) to further enrich the shape of the RND. However, for each mixture component, one needs to ensure that: (1) the implied option price is analytically tractable and numerically fast to compute, so that the parameters can be reliably estimated; (2) Assumption 1 holds for the mixture density, which implies the existence of moments for each component; (3) the flexibility to accommodate heavy tails of the RND. The Weibull and the lognormal densities appear to be the simplest ones which satisfy all three criteria. A more sophisticated mixture density may potentially further improve the results of this paper, which is left for future research.

Second, comparing with the (S)NP framework, our parametric mixture approach takes a drastically different strategy to balance under- and over-fitting. In detail, the (S)NP methods typically assume or imply an arbitrarily flexible candidate density that can asymptotically converge to any RND in practice, i.e., the model is asymptotically correctly specified. Therefore, the (S)NP approach should in theory be preferred in the limit due to this universal approximation property. However, in practice the sample size is typically small with noisy observations, and an infinitely flexible density can easily overfit the data with unrealistic RND estimates. Therefore, a central task for the (S)NP method is to guard against overfitting by carefully restricting the flexibility of the RND candidates. This is typically done by choosing some tuning parameters using 'rule of thumbs' based on previous works (Bondarenko, 2003, Bliss and Panigirtzoglou, 2004) or by asymptotic and cross-validation arguments (Yuan, 2009, Lu and Qu, 2021). Nevertheless, as we shall show in the simulation section, the finite sample optimal choices of the tuning parameters can depend heavily on the unknown true RND. As a result, asymptotically valid tuning parameter choices, which are the recommended options in the literature, may be very different from the finite sample optimal ones. Therefore, although a properly tuned (S)NP method can have good finite sample performance, blindly adopting recommended tuning

[9] Alternatively, one can model the RND for the log-return and choose parametric densities defined on the entire real line. Nevertheless, as the option prices implied by the log-returns (e.g., a normal distribution) and its corresponding simple returns (e.g., a log-normal distribution) coincide, the estimation problem remains unchanged.

parameter choices can lead to undesirable performance deterioration, as it is not informative enough about the goodness-of-fit of the (S)NP model.

Distinct from the (S)NP framework which builds a finite-sample RND from restricting an asymptotically correct model, we acknowledge the misspecification of our parametric model and prioritizing on eliminating severe form of misspecification with a parsimonious model. The parsimony originates from two sources: (1) the M-LWM density is highly flexible and adaptive to different tail shapes in the empirical RNDs; (2) the sequential test procedure gradually increases the simplest misspecified model until there are no clear patterns remaining in the residuals. As a result, our approach usually provides a mildly misspecified fit to the observed data with typically less parameters than a correctly tuned (S)NP method, which translates into a better finite sample performance.

4 Simulation Study

In this section, we conduct a comprehensive simulation study to (1) evaluate the performance of the LWM-based RND estimates under different RND and noise specifications and (2) examine the effectiveness of the sequential test on choosing the number of mixtures.

We consider four data generating process (DGP) specifications for the true RND $f_t^*(x)$ generated by the start-of-the-art three-factor stochastic volatility (3FSV) model of Andersen et al. (2015a,b), which has been shown to explain the risk-neutral price and volatility dynamics of empirical assets very well. The four densities are plotted in Figure 4.1. In detail, the density corresponding to the "Mid Volatility" DGP II in Figure 4.1 shows a negatively skewed RND function that is adapted from an empirically estimated RND by Andersen et al. (2015a) using S&P 500 equity-index options sampled every Wednesday between January 1996 and July 2010. The "Low Volatility" DGP I and "High Volatility" DGP III are modified versions of the "Mid Volatility" case, offering alternative RNDs with different symmetrical and dispersion properties, as depicted in Figure 4.1. Inspired by the observation in Lu and Qu (2021), Kostakis et al. (2023), we also consider a bimodal RND, which is formed by linearly combining the RNDs in DGP II and III, with the respective weights of 0.2 and 0.8. Choices of parameters and the technical details of extracting the above four RNDs and the associated option prices are presented in Online Appendix C.

For each of the four DGPs, we compute the theoretical prices of call options with time-to-maturity (τ) of one month on a DGP-specific strike range $\mathcal{K}$, shown in Table 4.1, such that theoretical call and put prices on this range are no less than 0.01. This is motivated by an empirical fact that options are traded at a minimum of 1 cent in practice. For each DGP, we consider different sample sizes $N \in \{25, 100\}$ equidistantly spaced on the strike range to examine the behaviour of our method

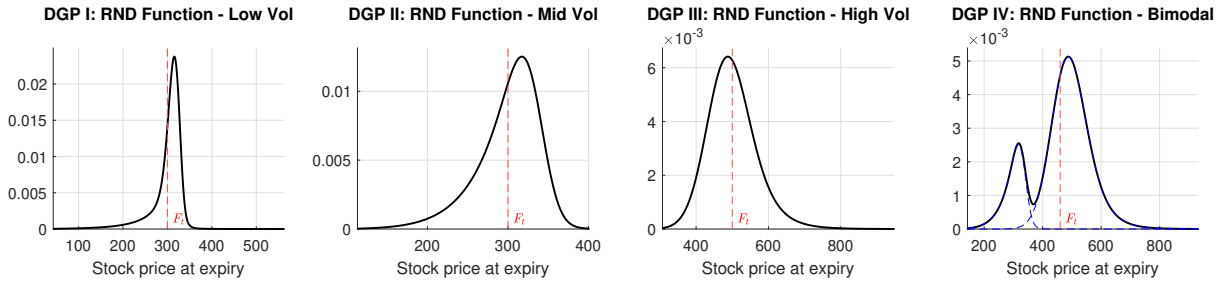

Figure 4.1: Plot of the true RND functions for the four three-factor double exponential stochastic volatility DGPs in our simulation. Details of these DGPs and the choice of parameters can be found in Online Appendix C. F_t denotes the true forward price.

under small or large sample sizes, which are typical for empirical cross-sections of option prices. For simplicity, the risk-free rate is assumed to be zero.

	DGP I	DGP II	DGP III	DGP IV
	Low Volatility	Mid Volatility	High Volatility	Bimodal
F_t	300	300	500	460
$\mathcal{K}$	[43, 563]	[113, 401]	[307, 948]	[143, 933]

Table 4.1: Forward price (F_t) and strike range ($\mathcal{K}$) for the four DGPs in our simulation.

4.1 Baseline Option Error Specification

We begin with a simple setting for the observed option prices with homoscedastic[10] additive observation errors defined on the price domain. On the observed strike grid $(K_n)_{n=1:N}$, we generate observed option prices based on the formula $C_n = \max(C_n^* + u_n, 0)$, where C_n^* and C_n are the theoretical and observed call price implied by the 3FSV model at K_n, respectively. The observation errors $u_n = \gamma_u \times Z_n$, where $\{Z_n\}_{n=1:N}$ is a sequence of i.i.d uniform $U(-1, 1)$ random variables. Alternative option error designs allowing for heteroskedastic and/or autocorrelated errors are also considered and will be summarized in the next subsection, with details presented in Online Appendix F.3. We consider three different values for the proportionality factor γ_u, namely 0.1212 (the small error variance case), 0.1715 (the medium error variance case) and 0.2425 (the large error variance case), so that the variance of u_n is respectively about the same as, twice, and four times the estimated SPX options' error variance in our empirical application.

For each DGP, sample size N and error variance size, we draw 1000 random cross-sections of observed call option prices. For each path, we follow the standard practice in the empirical option literature to remove options with zero prices and then estimate the RND using the LWM-based method with M ranging from 2 to 5, using an equal weighting (EW) scheme for our NLS estimation, which is appropriate given the homoskedasticity of the option errors. The $M = 1$ case produces heavily

[10] This is motivated by an observation that the option errors estimated for the SPX options traded on 20 June 2013 considered in our empirical application appear roughly homoskedastic (see Fig. 5.3).

biased RND estimates under all DGPs considered here, and is thus not reported for brevity. The LWM parameters are estimated based on the computational details in Online Appendix B with the maximum number of mixtures $\overline{M} = 5$ and $c = 0.1$ for the variance constraint in Eq. (24). Online Appendix F.1 presents a sensitivity analysis for the choice of c.

We also consider five competing estimators as benchmarks: (1) the positive convolution approximation (PCA) of Bondarenko (2003), denoted by PCA(h^*) and PCA($\widetilde{h}$), where h^* is the bandwidth parameter that fixes the number of normal densities at 23 as recommended by Bondarenko (2003),[11] and $\widetilde{h}$ is an optimal bandwidth obtained by a grid search to minimize the root mean integrated squared error (RMISE) of the fitted RND defined in Eq. (26); (2) the local linear regression (LLR) of Aït-Sahalia and Duarte (2003) with RMISE-optimal bandwidth; (3) the sieve estimator of Lu and Qu (2021), denoted by Sieve(J^*) and Sieve($\widetilde{J}$), where J^* is the adaptive order of the Hermite expansion as recommended by Lu and Qu (2021), and $\widetilde{J}$ is the Hermite expansion order that minimizes the RMISE; (4) the smoothing implied volatility method of Bliss and Panigirtzoglou (2004), denoted by IV spline(λ^*) and IV spline($\widetilde{\lambda}$), where λ^* and $\widetilde{\lambda}$ are respectively the tuning parameters chosen by a 10-fold cross validation and the RMISE-optimal choice; (5) the 4-parameter Hermite Expansion (HE) method of Jarrow and Rudd (1982).[12] Implementation details of the competing estimators are described in Online Appendix D. We note that PCA($\widetilde{h}$), Sieve($\widetilde{J}$), LLR, and IV spline($\widetilde{\lambda}$) are using infeasible optimal tuning parameters that minimizes the RMISE based on all simulated prices. Therefore, simulation results of these benchmarks should be interpreted as upper bounds of the performance of the corresponding estimators in practice.

We highlight that the LWM method always produces RND estimates that satisfy Assumption 1. This property is shared by only the PCA method among the competitors. The sieve estimator, the IV spline and the HE method may produce negative RND estimates. The LLR and the IV spline only estimate the RND on the strike range, so that Assumption 1.(3) is not guaranteed. The LLR and the sieve method produce RND estimates without ensuring Assumption 1.(4), so that the expected value of the RND is not necessarily the spot forward price.

In order to examine the quality of an RND estimate, denoted by $\widehat{f}_t(x)$, we consider both the mean total variation distance (MTVD) and the RMISE to measure its overall performance, defined as follows:

$$\text{MTVD} = 100 \times \text{E}[\sup_{x \in \mathcal{K}} |\widehat{f}_t(x) - f_t^*(x)|], \tag{25}$$

$$\text{RMISE} = 100 \times \sqrt{\text{E}[\textstyle\int_{\mathcal{K}} (\widehat{f}_t(x) - f_t^*(x))^2 dx]}. \tag{26}$$

[11] Bondarenko (2003) (p.110) reported using between 21 and 25 normal densities in his simulation study to achieve good performance, so we use the average number of normal densities in our study.

[12] We thank the reviewers for recommending the last two methods.

Note that we compute the above statistics on the strike range $\mathcal{K}$ to facilitate comparison among all competing estimators. The RMISE can be further decomposed into the root integrated squared bias (RISB) and the root integrated variance (RIV) as $\text{RMISE}^2 = \text{RISB}^2 + \text{RIV}^2$, where:

$$\text{RISB} = 100 \times \sqrt{\int_{\mathcal{K}} (\text{E}[\widehat{f}_t(x)] - f_t^*(x))^2 dx}, \quad \text{RIV} = 100 \times \sqrt{\int_{\mathcal{K}} \text{E}[(\widehat{f}_t(x) - \text{E}[\widehat{f}_t(x)])^2] dx}. \tag{27}$$

Intuitively, MTVD measures the largest deviation of the RND estimate from the true density, RMISE measures the overall quality of the RND estimate in the (pointwise) mean squared error sense, RISB is a measure of the accuracy, and RIV is a measure of the stability.

We present the simulation results for these evaluation metrics of all RND estimators under the large error variance case in Table 4.2. Results for the small and medium error variance cases are qualitatively similar and are reported in Online Appendix F.2. Graphical illustrations of the RND estimates for selected models are shown in Figures 4.2-4.5. Focusing first on the performance of the LWM-based RND estimates, our findings clearly show that the LWM model can accurately recover various RNDs with different shapes implied by the highly flexible 3FSV models. For different choices of N and DGP, the best performing LWM models only use 2 to 3 (resp. 3 to 4) mixtures for unimodal (resp. bimodal) RNDs. In general, choosing a larger M tends to increase the RIV but decrease the RISB, as the LWM model becomes less misspecified with more parameters, except for DGP II where the 2-LWM fit is already highly accurate. Consistent with the motivation discussed earlier in Section 2.2, more Weibull components (smaller M_1) are needed to fit RNDs with heavier left tails (DGPs I and II), and vice versa for DGP III. For a bimodal RND in DGP IV formed by a linear combination of left- and right-skewed RNDs, the two components play a more balanced role.

As to the performance of the competing estimators, the infeasible PCA($\widetilde{h}$), Sieve($\widetilde{J}$), and IV spline($\widetilde{\lambda}$) estimators have overall good and comparable performance in terms of MTVD and RMISE to the best LWM estimator for all DGPs. However, their feasible versions, namely PCA(h^*), Sieve(J^*) and IV spline(λ^*), are considerably less precise (except for Sieve(J^*) in DGP III when the recommended and the optimal tuning parameters coincide) and are all outperformed by the best LWM estimator, which shows the sensitivity to tuning for the (S)NP methods.

The PCA(h^*), Sieve(J^*) and HE estimators are amongst the most biased RND estimators for all DGPs except DGP III due to their insufficient flexibility to capture the various RND shapes. The 4 parameter HE estimator is by construction restrictive and may produce negative density estimates, as observed in Figure 4.2. The recommended tuning parameter choices h^* and J^* are too conservative for the PCA and the sieve methods, resulting in highly biased RND estimates, especially for DGPs I and IV. Also, even the optimally tuned LLS method has a non-trivial RISB, which is likely caused

by its optimal global bandwidth selection[13] for the kernel function. It is worth noting that Sieve(J^*) suggests bimodal RND estimates for DGPs I and II that are unimodal RNDs with heavy left tails, which will also appear in our empirical application.

It can be seen that the performances of both the PCA and the sieve estimators improve substantially by increasing their flexibility, i.e., the number of parameters involved. In particular, using 2 to 3 times the recommended number of mixtures in Bondarenko (2000, 2003) for the PCA method for DGPs I and IV and 2 to 4 times the recommended Hermite expansion order in Lu and Qu (2021) for the sieve method for DGPs I, II and IV, both PCA($\widetilde{h}$) and Sieve($\widetilde{J}$) estimators perform substantially better and can to a good extent recover the features of the true RNDs.

Summarizing our discussions above, we show that the LWM method has comparable performances to the competing (S)NP RND estimators with infeasibly optimized tuning parameters but significantly outperforms the latter when adaptive/recommended tuning parameters are used. This not only demonstrates the statistical advantage of the LWM method over its (S)NP rivals but also highlights the strong reliance of the latter on the choice of the tuning parameters. The empirically optimal choice is likely to depend strongly on the underlying data, which raises questions about the reliability and generality of the recommended choices in the existing literature for (S)NP methods.

We continue with an analysis of the sequential test procedure to choose the number of mixtures for the LWM method as discussed in Section 3.2. For each sample path, we firstly select the number of mixtures M and the number of lognormal densities M_1 based on our sequential test procedure with p and q ranging from 0 to 3 and $\alpha = 0.05$. For each test specification, we report the average choice of M (M_1) and the RMISE of the RND estimates based on the chosen M (M_1). For brevity, we only present results for simulation scenarios under a large error variance ($4\times$) in Table 4.3, but note that results for other scenarios tell a qualitatively similar story.

Table 4.3 presents convincing evidence supporting the effectiveness of the sequential test procedure. For example, when $N = 100$, we on average choose $M = 3$ for DGP I and $M = 2$ for DGPs II and III for various combinations of (p,q), both of which give the smallest RMISE according to Table 4.2. Results for DGP IV — a bimodal RND — seem to suggest $M = 3$ should be chosen on average for $N = 100$, which produces the second smallest RMISE that is only slightly larger than the smallest, as shown in Table 4.2. With a small sample size ($N = 25$), the test tends to pick a smaller M than the optimal $M = 3$, which is possibly caused by a lack of power of our testing procedure in small sample. Nevertheless, the resulting RMISEs based on the chosen M are still better than most of the competing RND estimators.

[13] We note that an adaptive local bandwidth parameter choice for the kernel regression should deliver optimal results here. However, here we follow Aït-Sahalia and Duarte (2003) to implement a global bandwidth parameter for the LLR method. We thank a referee for pointing this out.

	Large error variance (4×)									
	$N = 25$					$N = 100$				
Estimator	MTVD	RMISE	RISB	RIV	M_1	MTVD	RMISE	RISB	RIV	M_1
Panel 1: DGP I										
2-LWM (EW)	**0.180**	**0.912**	0.747	0.524	0.00	0.157	0.786	0.744	0.252	0.00
3-LWM (EW)	0.393	1.475	0.394	1.422	0.50	0.178	0.717	0.484	0.529	0.11
4-LWM (EW)	0.654	2.478	0.289	2.461	1.10	0.255	0.962	0.455	0.847	0.59
5-LWM (EW)	0.865	2.927	**0.254**	2.916	1.67	0.486	1.531	**0.265**	1.508	1.26
PCA(h^*)	1.558	7.512	7.503	0.373	23.00	1.576	7.604	7.603	**0.072**	23.00
Sieve(J^*)	0.710	3.514	3.512	**0.116**	4.00	0.770	3.753	3.752	0.091	4.00
IV spline(λ^*)	0.461	3.430	1.513	3.078	-	0.388	3.689	1.102	3.520	-
HE	0.858	5.018	5.013	0.230	-	0.860	5.018	5.017	0.101	-
PCA($\widetilde{h}$)	0.187	1.069	0.648	0.850	79.17	0.134	0.769	0.454	0.621	85.08
Sieve($\widetilde{J}$)	0.257	1.495	0.845	1.234	15.00	0.187	1.137	0.641	0.939	19.00
IV spline($\widetilde{\lambda}$)	0.300	1.615	1.311	0.944	-	**0.122**	**0.694**	0.434	0.542	-
LLS	1.497	8.003	7.105	3.682	-	0.423	1.974	1.800	0.811	-
Panel 2: DGP II										
2-LWM (EW)	**0.054**	**0.334**	**0.068**	0.327	0.16	**0.030**	**0.179**	**0.093**	0.153	0.04
3-LWM (EW)	0.387	1.118	0.130	1.110	0.68	0.120	0.428	0.129	0.408	0.37
4-LWM (EW)	0.768	2.091	0.221	2.079	1.32	0.320	0.860	0.147	0.847	1.05
5-LWM (EW)	1.413	3.533	0.255	3.524	1.95	0.688	1.575	0.197	1.562	1.73
PCA(h^*)	0.090	0.606	0.523	0.307	23.00	0.099	0.654	0.641	**0.131**	23.00
Sieve(J^*)	0.166	1.049	0.995	0.332	4.00	0.164	1.091	1.075	0.191	4.00
IV spline(λ^*)	0.151	1.450	0.125	1.444	-	0.130	1.303	0.140	1.296	-
HE	0.195	1.521	1.515	**0.137**	-	0.195	1.570	1.516	0.408	-
PCA($\widetilde{h}$)	0.059	0.411	0.280	0.300	26.16	0.040	0.279	0.159	0.229	28.94
Sieve($\widetilde{J}$)	0.084	0.586	0.473	0.346	7.00	0.048	0.348	0.188	0.293	10.00
IV spline($\widetilde{\lambda}$)	0.055	0.355	0.217	0.281	-	0.039	0.255	0.205	0.151	-
LLS	0.309	1.578	1.077	1.153	-	0.128	0.718	0.526	0.489	-
Panel 3: DGP III										
2-LWM (EW)	0.011	0.108	0.064	0.087	1.88	0.008	0.075	0.063	0.042	1.99
3-LWM (EW)	0.042	0.298	0.046	0.294	2.57	0.012	0.113	0.038	0.106	2.82
4-LWM (EW)	0.101	0.602	**0.042**	0.600	3.26	0.029	0.216	0.029	0.214	3.58
5-LWM (EW)	0.210	1.148	0.060	1.146	3.91	0.066	0.402	**0.026**	0.401	4.24
PCA(h^*)	0.019	0.200	0.156	0.125	23.00	0.023	0.226	0.219	0.053	23.00
Sieve(J^*)	0.013	0.132	0.086	0.101	4.00	0.010	0.103	0.087	0.054	4.00
IV spline(λ^*)	0.087	1.914	0.164	1.907	-	0.051	0.716	0.081	0.712	-
HE	0.011	0.106	0.088	**0.060**	-	0.010	0.094	0.088	**0.031**	-
PCA($\widetilde{h}$)	**0.009**	**0.100**	0.042	0.091	26.09	**0.006**	**0.060**	0.029	0.052	26.50
Sieve($\widetilde{J}$)	0.013	0.132	0.086	0.101	4.00	0.010	0.103	0.087	0.054	4.00
IV spline($\widetilde{\lambda}$)	0.018	0.167	0.097	0.136	-	0.014	0.124	0.090	0.085	-
LLS	0.085	0.719	0.689	0.207	-	0.040	0.323	0.239	0.217	-
Panel 4: DGP IV										
2-LWM (EW)	0.067	0.711	0.555	0.444	0.99	0.065	0.693	0.551	0.420	1.00
3-LWM (EW)	**0.037**	0.296	0.172	0.241	1.41	0.029	0.229	0.163	0.161	1.24
4-LWM (EW)	0.076	0.513	0.126	0.498	1.66	**0.029**	**0.223**	0.129	0.182	1.37
5-LWM (EW)	0.172	0.883	**0.078**	0.880	2.18	0.069	0.376	**0.051**	0.373	1.87
PCA(h^*)	0.117	1.166	1.153	0.173	23.00	0.120	1.212	1.211	0.054	23.00
Sieve(J^*)	0.118	0.945	0.944	0.038	4.00	0.131	1.109	1.109	0.016	4.00
IV spline(λ^*)	0.125	2.523	0.520	2.469	-	0.127	2.379	0.349	2.353	-
HE	0.184	1.897	1.896	**0.025**	-	0.184	1.896	1.896	**0.013**	-
PCA($\widetilde{h}$)	0.046	0.414	0.302	0.283	42.52	0.033	0.297	0.214	0.207	47.72
Sieve($\widetilde{J}$)	0.037	**0.276**	0.234	0.148	8.00	0.032	0.253	0.223	0.121	10.00
IV spline($\widetilde{\lambda}$)	0.080	0.646	0.548	0.343	-	0.040	0.330	0.252	0.212	-
LLS	0.092	0.978	0.918	0.337	-	0.049	0.432	0.330	0.279	-

Table 4.2: Simulation results of the RND estimates based on the M-LWM model with equal weighting (EW) and five competing estimators under a **large error variance** (4× the estimated error variance of SPX options). The specification and choice of tuning parameters of the competing estimators are detailed in Appendix D. MTVD, RMISE, RISB and RIV respectively stands for mean total variation distance, root mean integrated squared error, root integrated squared bias, root integrated variance, and are defined in Eq. (25)-(27). M_1 reports (i) for LWM method, the average number of log normal densities chosen in each M-LWM model, (ii) for PCA method, the average number of normal mixtures, and (iii) for Sieve method, the average order of Hermite expansion (i.e. J^* or $\widetilde{J}$). For each simulation scenario, the statistics in the table are computed based on 1,000 simulated paths. The best and the top three statistics among all estimators are highlighted in **bold** and underlined, respectively.

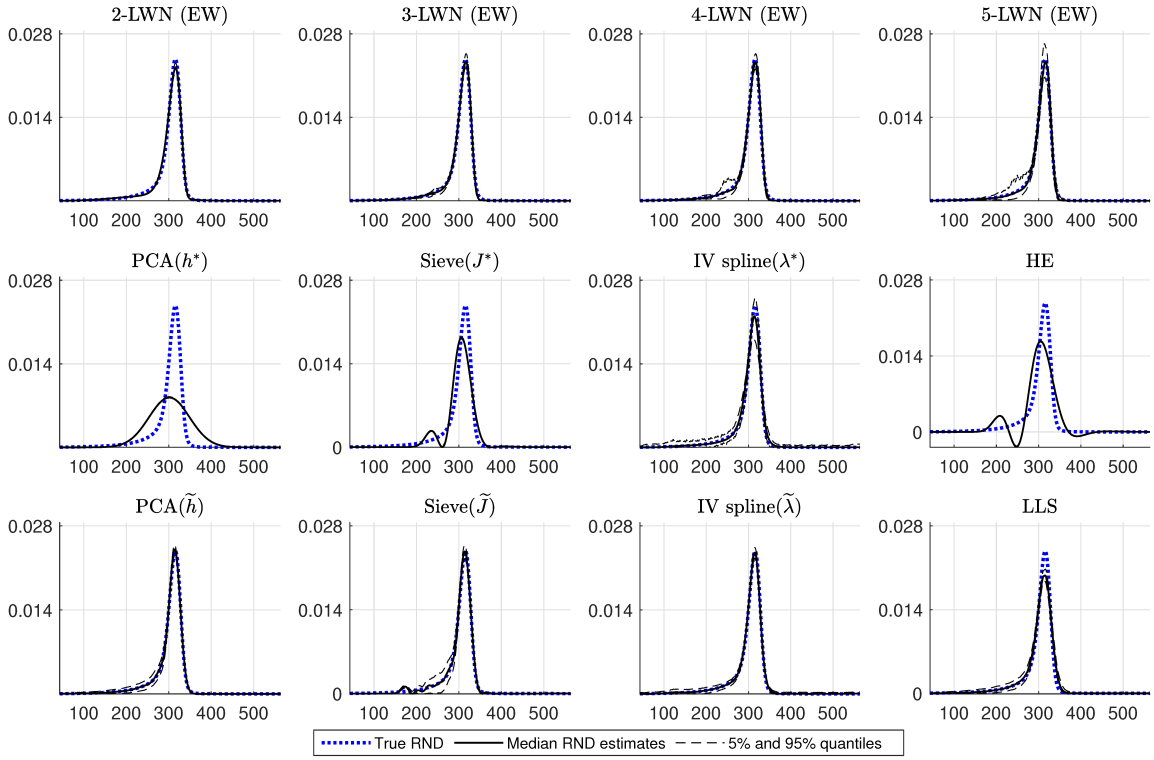

Figure 4.2: Comparison of RND estimates from selected models based on **DGP I** with a **large error variance** (4× **the estimated error variance of SPX options**) and $N = 100$. The median RND estimates and the 5% and 95% quantiles are computed pointwise based on 1,000 simulated paths.

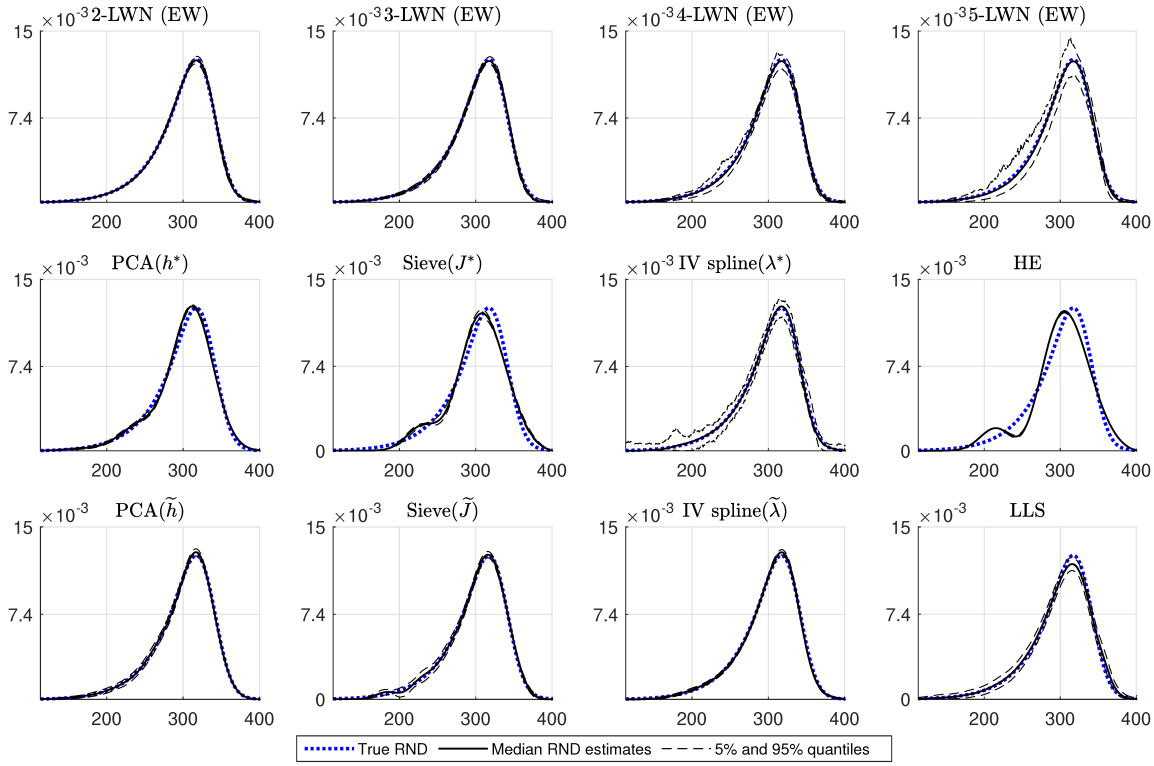

Figure 4.3: Comparison of RND estimates from selected models based on **DGP II** with a **large error variance** (4× **the estimated error variance of SPX options**) and $N = 100$. The median RND estimates and the 5% and 95% quantiles are computed pointwise based on 1,000 simulated paths.

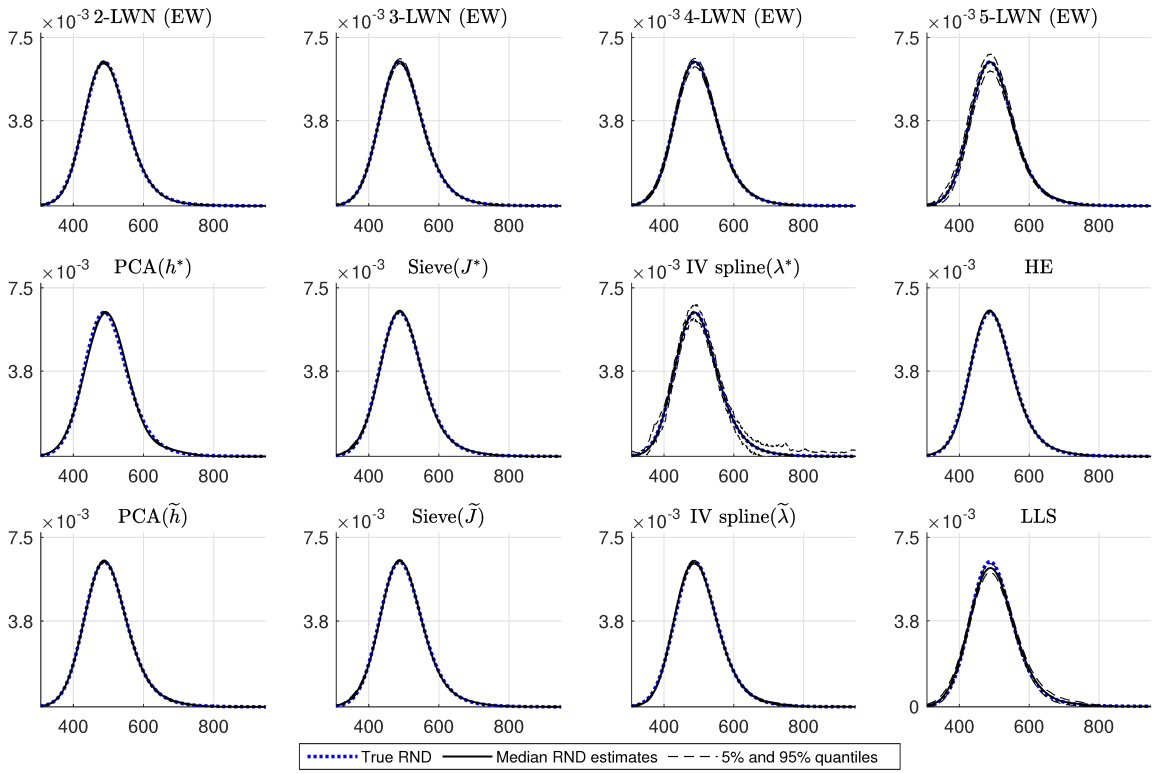

Figure 4.4: Comparison of RND estimates from selected models based on **DGP III** with a **large error variance** (4× **the estimated error variance of SPX options)** and $N = 100$. The median RND estimates and the 5% and 95% quantiles are computed pointwise based on 1,000 simulated paths.

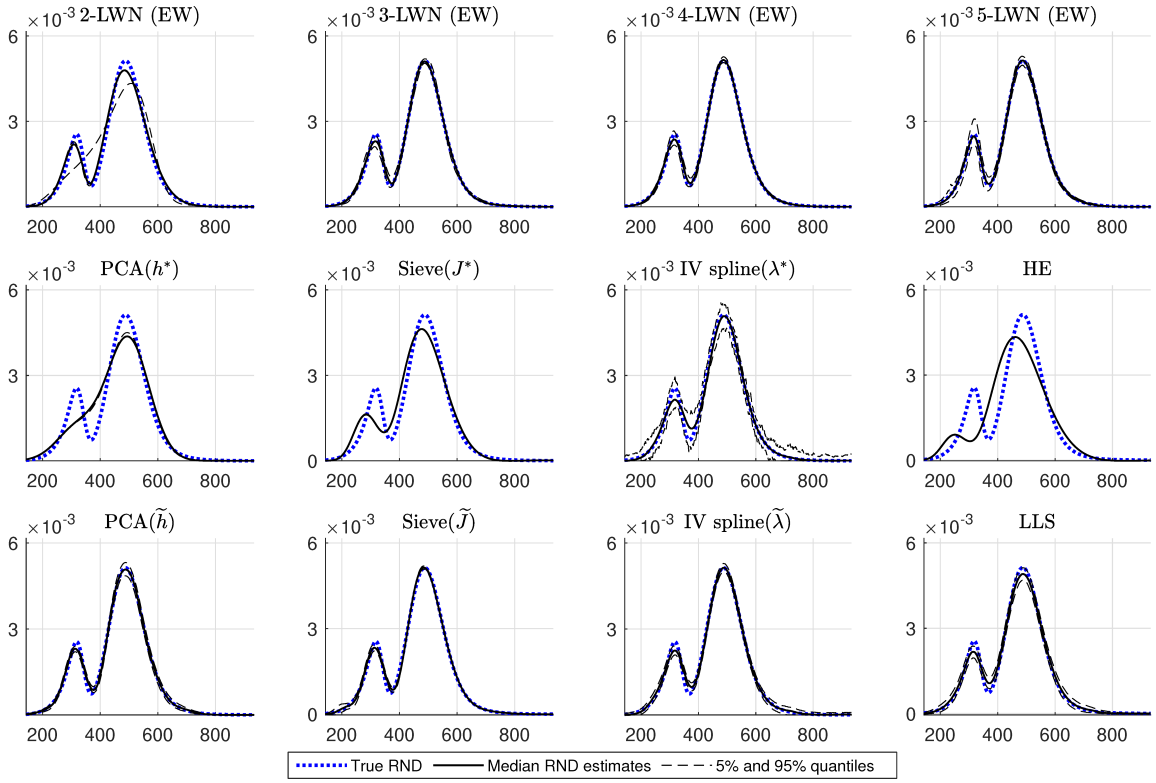

Figure 4.5: Comparison of RND estimates from selected models based on **DGP IV** with a **large error variance** (4× **the estimated error variance of SPX options)** and $N = 100$. The median RND estimates and the 5% and 95% quantiles are computed pointwise based on 1,000 simulated paths.

	Mean M (M_1) chosen				RMISE with chosen M (M_1)			
$p=$ \ $q=$	0	1	2	3	0	1	2	3
Panel 1: DGP I, $N = 25$								
0	2.50 (0.24)	2.00 (0.00)	2.00 (0.00)	2.00 (0.00)	1.144	**0.912**	**0.912**	**0.912**
1	2.03 (0.02)	2.00 (0.00)	2.00 (0.00)	2.00 (0.00)	0.936	**0.912**	**0.912**	**0.912**
2	2.00 (0.00)	2.00 (0.00)	2.00 (0.00)	2.00 (0.00)	**0.912**	**0.912**	**0.912**	**0.912**
3	2.00 (0.00)	2.00 (0.00)	2.00 (0.00)	2.00 (0.00)	**0.912**	**0.912**	**0.912**	**0.912**
Panel 2: DGP I, $N = 100$								
0	3.26 (0.23)	3.14 (0.21)	2.86 (0.12)	2.40 (0.05)	0.804	0.785	0.745	0.764
1	3.17 (0.20)	3.04 (0.14)	2.65 (0.08)	2.16 (0.03)	0.818	0.736	0.753	0.778
2	2.99 (0.13)	2.88 (0.10)	2.31 (0.03)	2.03 (0.00)	0.751	0.723	0.756	0.783
3	2.99 (0.12)	2.67 (0.08)	2.09 (0.01)	2.01 (0.00)	**0.716**	0.744	0.776	0.787
Panel 3: DGP II, $N = 25$								
0	2.02 (0.16)	2.03 (0.17)	2.06 (0.18)	2.07 (0.18)	0.349	0.340	0.366	0.370
1	2.01 (0.16)	2.04 (0.17)	2.07 (0.18)	2.07 (0.18)	**0.336**	0.348	0.374	0.366
2	2.02 (0.16)	2.04 (0.17)	2.07 (0.18)	2.08 (0.18)	0.339	0.347	0.380	0.378
3	2.02 (0.17)	2.06 (0.18)	2.07 (0.18)	2.08 (0.19)	0.343	0.362	0.378	0.395
Panel 4: DGP II, $N = 100$								
0	2.14 (0.08)	2.05 (0.05)	2.04 (0.05)	2.04 (0.05)	0.221	0.189	0.187	0.186
1	2.04 (0.05)	2.04 (0.05)	2.04 (0.05)	2.04 (0.05)	0.189	0.187	0.187	0.187
2	2.04 (0.05)	2.04 (0.05)	2.04 (0.05)	2.04 (0.05)	**0.183**	0.187	0.187	0.186
3	2.03 (0.04)	2.04 (0.05)	2.04 (0.05)	2.04 (0.05)	0.184	0.187	0.187	0.188
Panel 5: DGP III, $N = 25$								
0	2.02 (1.89)	2.00 (1.88)	2.00 (1.88)	2.00 (1.88)	0.111	**0.108**	**0.108**	0.109
1	2.00 (1.88)	2.00 (1.88)	2.00 (1.88)	2.00 (1.88)	**0.108**	**0.108**	**0.108**	0.109
2	2.00 (1.88)	2.00 (1.88)	2.00 (1.88)	2.00 (1.88)	**0.108**	**0.108**	**0.108**	0.109
3	2.00 (1.88)	2.00 (1.88)	2.00 (1.88)	2.00 (1.88)	**0.108**	**0.108**	**0.108**	0.109
Panel 6: DGP III, $N = 100$								
0	2.18 (2.13)	2.05 (2.03)	2.00 (2.00)	2.00 (1.99)	0.079	0.079	0.076	0.076
1	2.22 (2.17)	2.02 (2.01)	2.01 (2.00)	2.00 (1.99)	0.099	0.082	0.078	**0.075**
2	2.07 (2.05)	2.02 (2.00)	2.00 (2.00)	2.00 (1.99)	0.084	0.078	0.077	**0.075**
3	2.04 (2.02)	2.01 (2.00)	2.00 (1.99)	2.00 (1.99)	0.081	0.078	**0.075**	**0.075**
Panel 7: DGP IV, $N = 25$								
0	2.68 (1.24)	2.05 (1.02)	2.15 (1.05)	2.15 (1.05)	0.623	0.659	**0.528**	**0.528**
1	2.05 (1.01)	2.13 (1.04)	2.15 (1.05)	2.15 (1.05)	0.705	0.550	**0.528**	**0.528**
2	2.01 (1.00)	2.13 (1.04)	2.15 (1.05)	2.15 (1.05)	0.704	0.545	**0.528**	**0.528**
3	2.10 (1.03)	2.15 (1.05)	2.15 (1.05)	2.15 (1.05)	0.595	**0.528**	**0.528**	**0.528**
Panel 8: DGP IV, $N = 100$								
0	2.98 (1.18)	3.12 (1.24)	3.21 (1.26)	3.18 (1.26)	0.556	0.367	0.210	0.242
1	2.98 (1.19)	3.20 (1.26)	3.21 (1.26)	3.16 (1.25)	0.538	**0.210**	0.212	0.231
2	2.92 (1.19)	3.22 (1.26)	3.18 (1.26)	2.97 (1.20)	0.538	0.210	0.250	0.306
3	3.16 (1.26)	3.21 (1.26)	3.15 (1.25)	2.54 (1.12)	0.225	0.219	0.237	0.421

Table 4.3: Mean choice of M (M_1) by the sequential test procedure and the performance of the RND estimates based on the chosen M (M_1) under **large error variance** (4×) scenarios. All statistics are computed based on 1,000 simulated paths. Columns 2-5 reports the mean M (M_1) selected based on the $\text{Wald}_{p,q}$ statistic with $\alpha = 0.05$. Columns 6-9 reports the RMISE, defined in Eq (26), of the RND estimates based on the LWM model with M chosen by the sequential test. The best and top three test specifications are highlighted in **bold** and underlined, respectively. If more than one statistic are **bold** and underlined, they are of the same value.

The above findings provide useful guidance for the choice of p and q in practice. As the LWM model is more likely to be misspecified with real data, we recommend choosing at least $p = q = 1$ to better detect this unknown form of misspecification. For M (M_1) chosen based on $p, q \in \{1, 2\}$, the RMISE is quite robust to different DGPs and sample sizes and is close to the optimal RMISE presented in Table 4.2. Also, the conventional 5% significance level works well in rejecting misspecified cases for the $N = 100$ case. We thus use $p = q = 1$ with $\alpha = 5\%$ in our empirical analysis.

Last, we comment on the validity of the estimated confidence bounds for the LWM-based methods, with the relevant results presented in Online Appendix F.4. As the confidence bounds constructed based on Corollary 1 only hold in large samples with the correct model specification, we in general do not expect the feasible confidence bounds to have correct coverage rates in finite sample. However, our results show that, when $N = 100$ with a mild model misspecification, the estimated confidence bounds are conservative, i.e., the RND are more likely to be included in the confidence bounds than the nominal confidence level. We therefore believe that, although the estimated confidence bounds are likely invalid in theory, they can still provide some empirical guidance to the location of the true RND in practice.

4.2 Alternative Option Error Specifications

In addition to the baseline i.i.d. option error design discussed above, we consider three alternative error specifications that allow for heteroskedastic and/or strong autoregressive correlation in option errors. We re-examine the performance of RND estimators based on the LWM method and five other competitors. To save space, we summarize the main findings here, while relegating the details of the alternative error designs and key results to Online Appendix F.3.

Consistent with the baseline results, the LWM method works well in extracting different underlying RNDs with a small number of mixtures (2 or 3) even under heteroskedasticity and/or strong serial correlation. In detail, the LWM method provides accurate RND estimates with the smallest MTVD and RMISE amongst all feasible competitors and are comparable to the infeasibly optimal (S)NP methods. It is worth noting that when the error terms are homoscedastic on the implied volatility domain, a WNLS approach with inverse squared vega weights, which is approximately optimal in view of Christoffersen et al. (2018), can yield more precise RND estimates for relatively large M. This suggests that one can potentially improve the RND estimate by adopting a generalized nonlinear least square (GNLS) to solve Eq. (11). However, the performance of the RND estimates with the approximate optimal weighting scheme still does not convincingly beat the performance with equal weighting. This is possibly due to the fact that the efficiency gain can only be achieved in the asymptotic limit when the model is correctly specified, which does not necessarily hold true in finite

samples under a possibly misspecified model. As the equally weighted model provides more robust finite sample results and is easier to implement than a GNLS design which requires the knowledge of the dynamic structure of the option errors, we recommend equal weights for the task of RND estimation. Alternative weighting schemes can serve as robustness checks for the RND estimates, which is considered in our empirical analysis.

5 Empirical Illustrations

We use two examples to showcase the usefulness of our LWM approach in practice, which also point outs some caveats of implementing semi-nonparametric methods with real data. In the first example, we examine the one-month RND estimate of the S&P500 index on 20-Jun-2013. Lu and Qu (2021) show that this RND is bimodal based on their sieve estimator with the recommended tuning parameters. To replicate this result, we obtain the closing option quotes written on SPX with one month to expiration and the risk-free rate from OptionMetrics[14]. The dataset is then cleaned according to Online Appendix E to obtain a single cross-section of call option prices based on OTM options. It is worth noting that, different from Lu and Qu (2021), we only use the mid-quotes from OTM options as they are more liquidly traded with smaller spreads, and hence less noisy than the ITM options (Aït-sahalia and Lo, 1998). However, our findings based on the sieve estimator does not change qualitatively when using all puts and calls. The cross-section of call option prices is presented in Fig. 5.1.

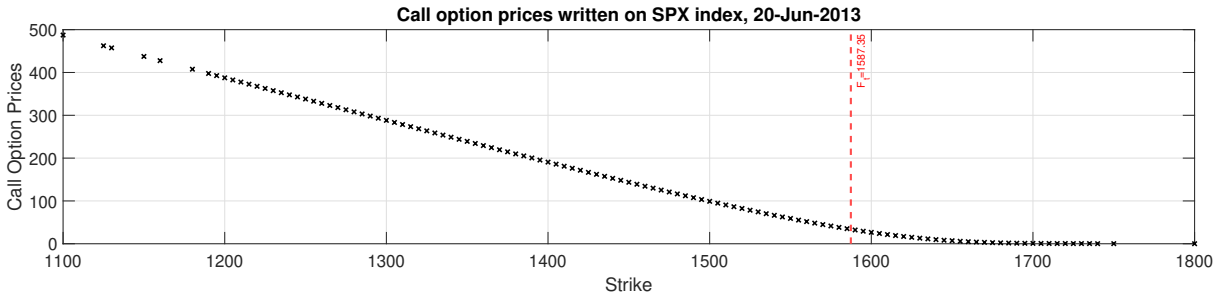

Figure 5.1: Call option prices written on SPX expiring in one month on 20-Jun-2013. The data is obtained from OptionMetrics which is cleaned according to the procedure in Online Appendix E. The dataset has $N = 119$ observations.

We consider several estimators in Section 4 to extract the RND from the option prices in Fig. 5.1. First, we estimate the equally weighted (EW) LWM-based RND with a maximum of 5 mixtures. We use the sequential test procedure with $p = q = 1$ and $\alpha = 5\%$ to choose the optimal number of mixtures. This gives a 3-LWM RND estimate with $M_1 = 0$, i.e., a pure mixture-of-Weibull density. To replicate the findings in Fig. 12 of Lu and Qu (2021), we estimate the Sieve(J^*) density without regularization[15],

[14] https://optionmetrics.com/

[15] This is achieved by setting $\xi = 0$, which is motivated by our simulation and the findings in Fig. 5.2 that the choice $J^* = 5$ is too restrictive, so regularization is not needed. Tuning ξ via a 10-fold cross-validation produces virtually identical RND estimates.

where $J^* = 5$ is the truncation order used in Fig. 12 of Lu and Qu (2021). We also consider the PCA estimator with $h^* = 70$ which produces 23 mixtures as recommended by Bondarenko (2003). As a fully non-parametric benchmark, we estimate RND by the IV spline of Bliss and Panigirtzoglou (2004) with the smoothing parameter λ selected by a ten-fold cross-validation. Due to the overall inferior performance of the LLS and the HE method in the simulation, we exclude them from the empirical analysis for conciseness. We present the RND estimates of different methods and the associated fitted residuals in Fig. 5.2.

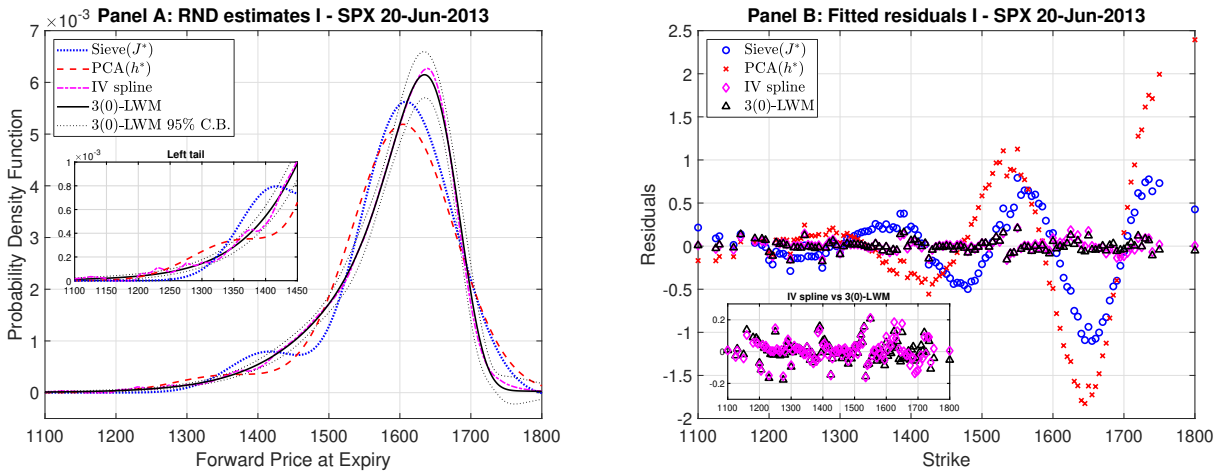

Figure 5.2: One-month-ahead RND estimates and the corresponding fitted residuals for the S&P500 index based on the sieve estimator, PCA, IV spline and the equally weighted $M(M_1)$-LWM method. C.B. stands for confidence bounds. The recommended tuning parameters used for the semi-nonparametric methods are: $J^* = 5$ and $\xi = 0$ for the Sieve estimator, and $h^* = 70$ (23 parameters) for PCA. For the IV spline method, we determine the smoothing parameter λ by a 10-fold cross-validation.

Panel A of Fig. 5.2 reveals a substantial discrepancy between the RND estimates based on different methods. Sieve(J^*) produces a bimodal RND estimate, which is consistent with the findings in Fig. 12 of Lu and Qu (2021) but different from PCA(h^*) which suggests a thick left tail. The sieve-based RND estimate produced has a thin left tail which does not agree with the other RND estimates. The IV spline and the 3-LWM produce very close RND estimates for the majority of the strike range, while some noisiness in the left tail of the estimated RND is observed for the IV spline method. The heavy left tail also explains the dominating Weibull component in the 3-LWM density, which is consistent with the intuition conveyed in Fig. 2.2.

Panel B of Fig. 5.2 presents further details on the goodness-of-fit of the various method to the observed option process. In detail, both Sieve(J^*) and PCA(h^*) underfit the observed option prices with evident and distinct sinusoidal patterns in the fitted residuals, which are much larger in magnitude than those from the other two methods. The fitted residuals for the IV spline and the 3-LWM method are of similar scales and without obvious patterns. This indicates that the IV spline and the 3-LWM method fit the observed option prices substantially better than the two SNP methods with recommended tuning parameters, while the 3-LWM method produces a smoother RND estimate than

the IV spline method.

In Fig. 5.3 we provide more details on the sequential test procedure and validate that $M = 3$ is the appropriate choice of the LWM model in Fig. 5.2. The figure shows that, with $M = 2$, the LWM model slightly underfits the observed options with a U-shaped pattern captured by our Fourier regression, which results in a strong rejection of the Wald test. This pattern vanishes after we increase M to 3 with a flat Fourier regression fit, accompanied by a sharp decrease in the SSE and an insignificant Wald test. Increasing M further only reduces the SSE slightly but does not seem to improve the goodness-of-fit further. This is in line with the intended outcome of the sequential testing procedure, which attempts to choose the smallest M that eliminates severe forms of misspecification. It is worth noting that the residuals for $M \geq 3$ do not appear heteroscedastic in the strike domain, so equal weights appear to be the appropriate weighting scheme here.

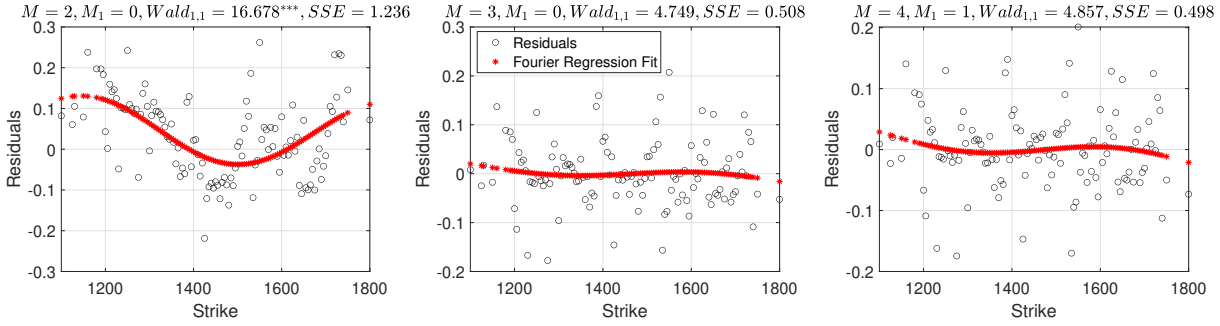

Figure 5.3: Fitted residual plots and the corresponding Fourier regression fit of the equally weighted LWM model with $M \in \{2, 3, 4\}$. The title of each plot state the choice of M and M_1, the Wald statistic constructed from Proposition 1 with $p = q = 1$, and the sum of squared errors (SSE). *, ** and *** represent significance at 10%, 5% and 1% significance level, respectively.

The above discussion reveals the tuning parameter choice problem for the (S)NP methods discussed in our simulation, as blindly adopting the recommended choices can lead to heavily biased RND estimates due to insufficient flexibility. To confirm this point, we adaptively choose the tuning parameters for the PCA and the sieve method by benchmarking on a smooth RND estimate, which is taken to be the 3(0)-LWM RND estimate in Fig. 5.2. Specifically, we first compute the fitted option prices $\{\widehat{O}_n\}_{n=1:N}$ from the estimated 3(0)-LWM model. We then fit the (non-penalized) sieve and the PCA models on $\{\widehat{O}_n\}_{n=1:N}$ with various choices[16] of J and h, and select $\widetilde{J}$ and $\widetilde{h}$ that minimize the total variation distance between the model-fitted RND and the 3(0)-LWM RND by a simple grid search. Following Lu and Qu (2021), we set the regularization parameter ξ of the sieve estimator by a 10-fold validation. Importantly, as $\widetilde{J}$ and $\widetilde{h}$ are computed from fitted instead of the original option prices, it only ensures that the PCA and the sieve model have a similar degree of smoothness as the 3(0)-LWM RND estimate, but it does not necessarily guarantee the convergence of the RND estimates based on the original data. We find $\widetilde{J} = 19$ and $\widetilde{h} = 28.1$ that corresponds to 53 mixtures for the PCA method, both of which suggest that the flexibility of the two models needs to be substantially

[16] We consider $J \leq 50$ and $h \in [20, 42]$, which is sufficiently large to contain the optimal choices.

increased to fit the benchmark RND.

Finally, we also consider an inverse Vega weighted LWM estimator (see footnote 4), where the option Vegas are computed from a smoothing cubic spline of observed IVs with a smoothing parameter of 10^{-4}. The sequential test suggests $M = 4$ with $M_1 = 2$. The RND estimates and the fitted residuals of these estimators are shown in Fig. 5.4.

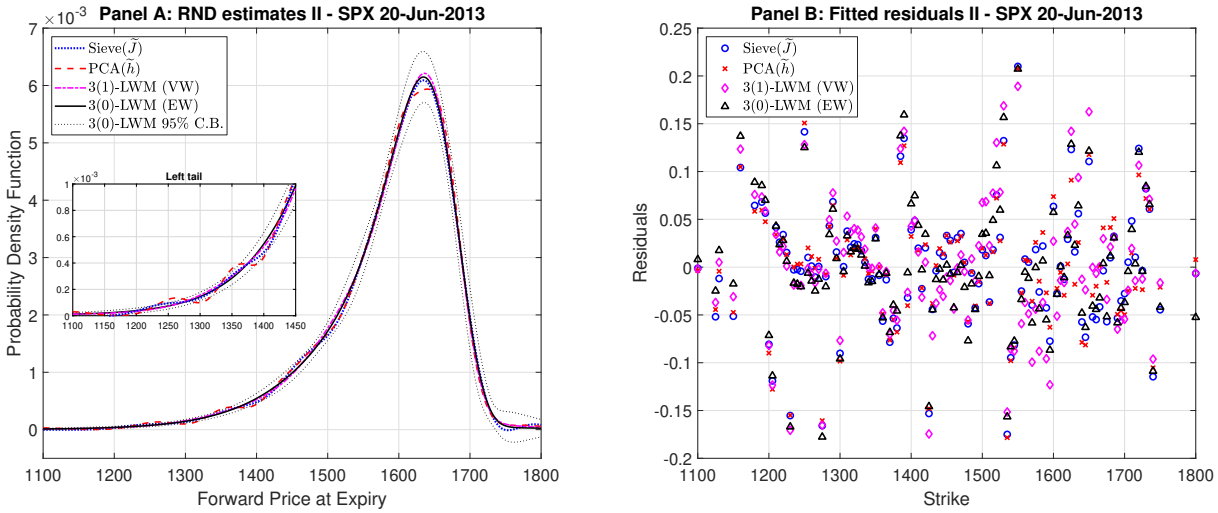

Figure 5.4: One-month-ahead RND estimates and the corresponding fitted residuals for the S&P500 index based on the sieve estimator, PCA, and the $M(M_1)$-LWM method with inverse Vega weights (VW) or equal weights (EW). C.B. stands for confidence bounds. The tuning parameters used for the semi-nonparametric methods are: $\widetilde{J} = 19$ and $\xi = 3.14 \times 10^{-4}$ for the Sieve estimator and $\widetilde{h} = 28.1$ (53 parameters) for PCA.

Panel A of Fig. 5.4 shows that, after increasing the flexibility of the SNP methods, the RND estimates of both Sieve($\widetilde{J}$) and PCA($\widetilde{h}$) are much closer to our 3-LWM estimates in Fig. 5.2. However, the left tails of the SNP-based RND estimates are less smooth than the LWM methods, which can be attributed to a potentially excessive flexibility of the SNP-based density. Although the Vega weighted LWM method suggests an additional lognormal component in the mixture, the estimated RND largely coincides with the equally weighted one. It is thus not surprising to see that all models have comparable goodness-of-fit to the observed option prices, as demonstrate in Panel B of the figure.

Our previous example demonstrates that the bimodality of the sieve-based RND estimate is likely spurious. The unimodal RND with a heavy left tail as suggested by the LWM method is more parsimonious with a satisfactory fit to the observed option prices, which leads to a smoother RND estimate than the fine-tuned (semi)-nonparametric contenders considered here. In our second example, we demonstrate that the LWM method allows us to detect genuine bimodality in the empirical RND driven by binary outcomes in political events which might have been overlooked in the literature.

Following Ferreira et al. (2022) and Kostakis et al. (2023), we examine the RNDs around the release of the Brexit referendum result on 24 June, 2016. In detail, we replicate Panel B of Fig. 3 and Panel C-D of Fig. 4 in Kostakis et al. (2023) by estimating the RND of the FTSE100 index on 15th

July 2016 using end-of-day option midquotes traded on 23 and 24 June, 2016.[17] We present the two cross-sections of option prices and the corresponding IV curves in Fig. 5.5. The figure shows a 2.7% drop of the forward price of the FTSE100 index after the Brexit referendum, which is likely caused by the 'leave' outcome. It is also worth noting that, unlike the usual volatility smirk observed on 24 June, the IV curve on 23 June has an unusual locally concave curvature in the ATM region, which motivates our analysis below.

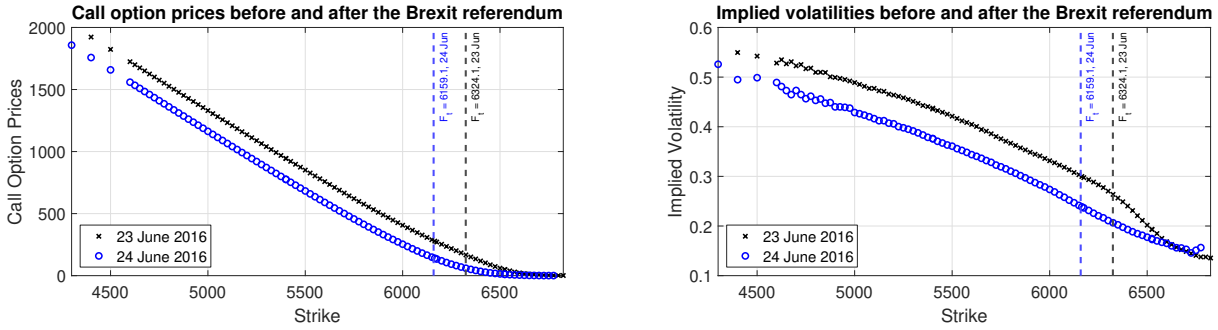

Figure 5.5: Call option prices written on FTSE100 index expiring on 15th July 2016. The original data is sourced from CME Datamine which is cleaned according to the procedure in Online Appendix E. The 23-Jun (resp. 24-Jun) dataset has $N = 95$ (resp. $N = 94$) observations. F_t in the plots are the corresponding forward prices of the FTSE100 index.

The local concavity of the IV curve requires some further investigation. In fact, Fig. 4 of Kostakis et al. (2023) shows that the IV curves of both the GBP/USD options and the FTSE100 options are locally concave on 23th June. The authors fit a quintic IV spline to the observed IV curve to extract the RND on the strike range, following Bliss and Panigirtzoglou (2004). Interestingly, in their Fig. 3, the authors only show a bimodal RND for the GBP/RND exchange rates while the FTSE100 RND is found to be unimodal with a heavy left tail. An inspection of Panel C of Fig. 4 reveals that the fitted IV curve completely ignores the concavity in the IV curve, which may be due to underfitting. Using our LWM approach, we shall verify this argument and show that the RND of the FTSE100 index prior to the Brexit referendum is also likely to be bimodal, which reflects the binary political event risk in the market.

To this end, we extract the RNDs using our LWM approach and the cubic IV spline method with the same implementation details as in our first example. As a benchmark, we also replicate the quintic smoothing IV spline implemented in Kostakis et al. (2023), whose flexibility depends on the tick size of the options contract. Findings on the sieve and the PCA methods are qualitatively similar to our first example, which we present in Online Appendix F.5. We present the RND estimates before and after the Brexit referendum in Fig. 5.6.

Panel A of Fig. 5.6 shows that the LWM method produces very similar RND estimates before and after the referendum under different weighting schemes. The RND estimates on 23-Jun is clearly

[17] Detailed description of the data can be found in Kostakis et al. (2023). We thank the authors for sharing an excerpt of their option data with us.

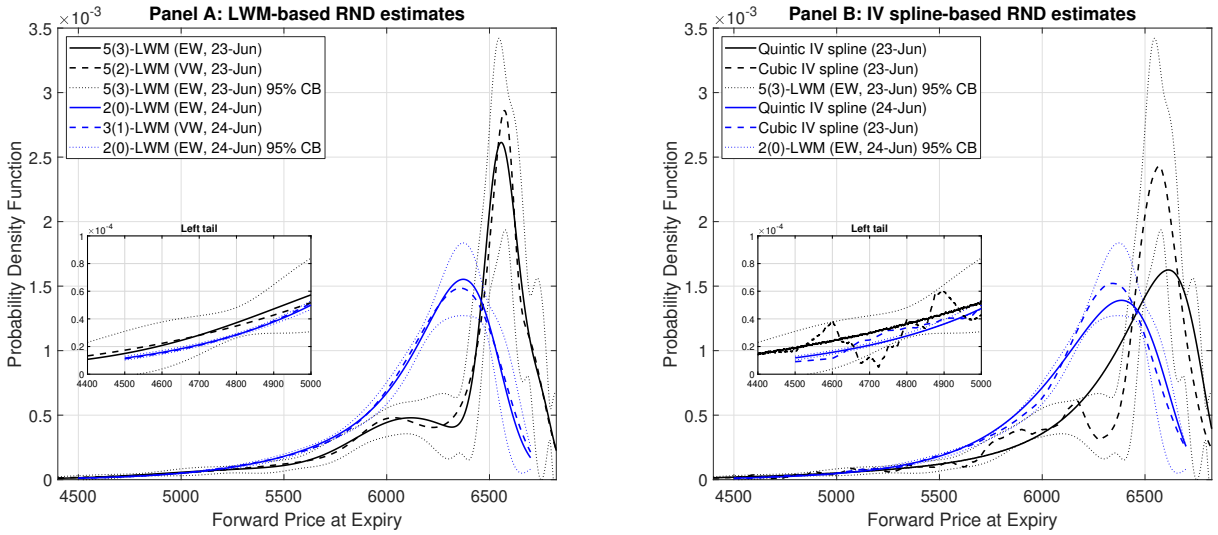

Figure 5.6: RND estimates before and after the Brexit referendum based on the option prices in Fig. 5.5. In each figure, the black and blue dotted lines are the 95% confidence bands of the equally weighted LWM-based RND estimates. The LWM and the cubic IV spline methods are implemented in the same way as in Fig. 5.2 and Fig. 5.4, and the quintic IV spline method on the strike range is implemented according to Kostakis et al. (2023).

bimodal, while the bimodality collapses on 24-Jun after the referendum. It is also intuitive that one needs more mixtures ($M = 5$) to fit a bimodal density than a unimodal one ($M = 3$). This is further confirmed by the cubic IV spline-based RND estimates in Panel B of the figure. However, the quintic IV spline-based RND estimates suggest unimodal RND estimates on both dates, which is identical to Panel B of Fig. 3 in Kostakis et al. (2023).

To evaluate the goodness-of-fit of the RND estimates in Fig. 5.6, in Fig. 5.7 we plot the corresponding model residuals in both the price and the IV domains. Further diagnostic information about the residuals are presented in Table 5.1. Fig. 5.7 shows that both price and IV residuals from the quintic IV spline method are large in magnitude with evident sinusoidal patterns, which clearly differs from the three competing estimators. Table 5.1 further confirms the poor fit of the quintic IV spline to the observed data, as its SSEs are hundreds of times larger than those of the other methods. It is also worth noting that the cubic IV spline has the smallest SSEs which results from the cross-validation, but the RND estimates can be quite noisy in the tails, as is shown in Fig. 5.6. As a sanity check, one can verify that the LWM (VW) model has better SSE in the IV domain than the LWM (EW) model and vice versa, as the inverse Vega weights allow us to approximately minimize the SSE in the IV domain as intended.

The above empirical findings suggest that the unimodal RND estimate on 23-June prior to the Brexit referendum in Kostakis et al. (2023) is likely due to an over-smoothed quintic IV spline. Instead, a bimodal RND as shown in Fig. 5.6 provides a much better fit to the empirically observed data, which can also be naturally explained by the market's anticipation of binary political event risk as pointed out by Hanke et al. (2018), Ferreira et al. (2022), Kostakis et al. (2023). Importantly, the LWM

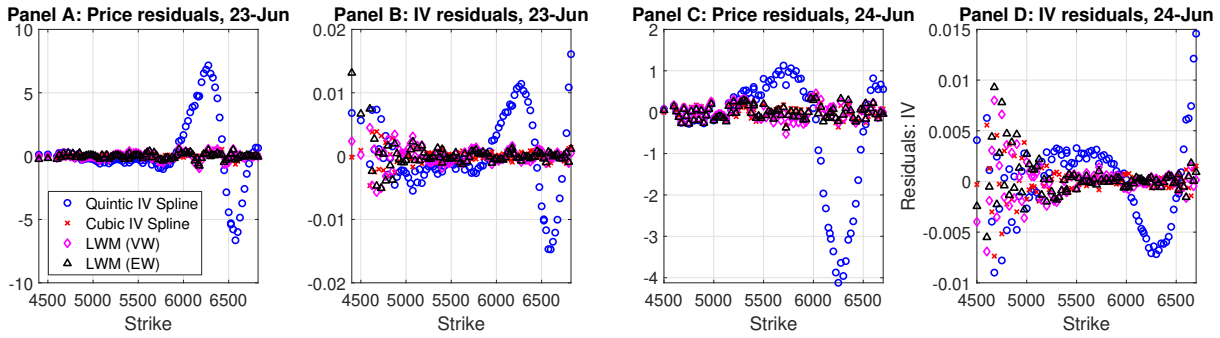

Figure 5.7: Option price and IV residuals associated with the RND estimates in Figure 5.6. For each estimator, the price residuals (resp. IV residuals) are defined as $O_n - \widehat{O}_n$ (resp. $IV_n - \widehat{IV}_n$), where O_n and $\widehat{O}_n$ (resp. IV_n and $\widehat{IV}_n$) denote the observed and the model-implied option prices (resp. implied volatilities) at strike K_n, respectively.

	23-Jun-2016					24-Jun-2016				
Estimators	M	M_1	SSE(O)	SSE(IV)$\times 10^3$	Wald$_{1,1}$	M	M_1	SSE(O)	SSE(IV)$\times 10^3$	Wald$_{1,1}$
LWM (EW)	2	0	181.943	4.514	16.781***	**2**	**0**	**7.285**	**0.699**	**6.879**
	3	1	24.192	2.175	15.028***	3	1	2.324	0.339	7.150
	4	2	19.363	0.798	10.220**	4	1	2.284	0.356	8.665*
	5	**3**	**4.569**	**0.411**	**8.018***					
LWM (VW)	2	0	261.985	1.458	10.261**	2	0	17.088	0.414	11.849**
	3	1	59.912	0.887	11.104**	**3**	**0**	**9.054**	**0.330**	**7.482**
	4	2	39.166	0.377	10.351**	4	1	11.639	0.314	21.514***
	5	**2**	**7.777**	**0.208**	**8.637***					
Cubic IV spline			3.473	0.140				2.101	0.244	
Quintic IV spline			702.134	3.450				174.485	1.548	

Table 5.1: Residual Diagnostics for the RND estimates in Figure 5.6. SSE(O) and SSE(IV) stands for sum of squared errors based on prices and IVs, respectively. For the LWM models, the chosen M based on the sequential test procedure is highlighted in bold.

approach produces smooth RND estimates that also provide a reasonably good fit to the observed data, which strike a desirable balance between the misspecified or over-fitting spline-based methods.

To sum up, the two examples demonstrate the flexibility and parsimony of the LMW method, as it can reliably recover complex empirical RNDs with heavy left tails or bimodality using up to 5 mixtures. The number of free parameters is typically much less than those of the SNP alternatives that achieve a comparable fit (see e.g., Fig. 5.2) after some careful tuning. We stress that the LWM method requires minimal tuning — one simply chooses the smallest M that eliminates obvious misspecification by our sequential test procedure. The additional confidence bounds for the LWM method also allow us to evaluate the precision of the RND estimates. Based on the above discussions, we argue that the LWM method provides a reliable parametric RND extraction technique which should be considered in practice. Our results also reveal that some recommended tuning parameters for the SNP methods can produce highly unreliable RND estimates due to a lack of flexibility, which can potentially be avoided by examining the fitted residuals. This highlights the importance of diagnostic testing for (semi)-nonparametric RND extraction tools which is largely overlooked in literature.

6 Concluding Remarks

This paper proposes a parametric RND extraction method based on a finite lognormal-Weibull mixture (LWM) density. We develop formal asymptotic theory for the estimator in a general misspecified framework. Based on the theoretical results, we propose a simple sequential test procedure to choose the number of mixtures M iteratively, which consequently determines the number of lognormal mixtures M_1. Both simulation and empirical analyses confirm that our LWM approach provides more reliable RND estimates that better fit the observed data and are more robust to tuning than popular (S)NP competitors. We argue that the precision gain mainly comes from the flexibility of the novel LWM density, which can approximate typical empirical RNDs with a much smaller number of parameters than the (S)NP approach. Therefore, despite its relatively higher computation costs, we believe that our LWM method provides a valuable parametric tool for practical RND extraction tasks.

Supplementary Material

The supplementary material of this paper contains: (1) the Online Appendix containing proofs and further results; (2) a MATLAB toolbox to implement the LWM method proposed in this paper, and the data used in our empirical analysis. It can be accessed from the publishers website.

References

Abadir, K. M. and Rockinger, M. (2003). Density Functionals, with an Option-Pricing Application. *Econometric Theory*, 19(5):778–811.

Aït-Sahalia, Y. and Duarte, J. (2003). Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, 116(1-2):9–47.

Aït-sahalia, Y. and Lo, A. W. (1998). Nonparametric Estimation of State-Price Densities Implict in Financial Asset Prices. *Journal of Finance*, LIII(2):499–547.

Aït-Sahalia, Y. and Lo, A. W. (2000). Nonparametric risk management and implied risk aversion. *Journal of Econometrics*, 94(1):9–51.

Andersen, T. G., Archakov, I., Grund, L. E., Hautsch, N., Li, Y., Nasekin, S., Nolte, I., Pham, M. C., Taylor, S. J., and Todorov, V. (2021a). A Descriptive Study of High-Frequency Trade and Quote Option Data. *Journal of Financial Econometrics*, 19(1):128–177.

Andersen, T. G., Fusari, N., and Todorov, V. (2015a). Parametric inference and dynamic state recovery from option panels. *Econometrica*, 83(3):1081–1145.

Andersen, T. G., Fusari, N., and Todorov, V. (2015b). The risk premia embedded in index options. *Journal of Financial Economics*, 117(3):558–584.

Andersen, T. G., Fusari, N., and Todorov, V. (2017). Short-Term Market Risks Implied by Weekly Options. *Journal of Finance*, 72(3):1335–1386.

Andersen, T. G., Fusari, N., Todorov, V., and Varneskov, R. T. (2021b). Spatial dependence in option observation errors. *Econometric Theory*, 37(2):205–247.

Andrews, D. W. (1999). Estimation when a parameter is on a boundary. *Econometrica*, 67(6):1341–1383.

Andrews, D. W. K. (1987). Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers. *Econometrica*, 55(6):1465–1471.

Andrews, D. W. K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica*, 59(3):817–858.

Banz, R. W. and Miller, M. H. (1978). Prices for State-Contingent Claims: Some Estimates and Applications. *The Journal of Business*, 51(4):653–672.

Birru, J. and Figlewski, S. (2012). Anatomy of a meltdown: The risk neutral density for the S&P 500 in the fall of 2008. *Journal of Financial Markets*, 15(2):151–180.

Black, F. (1976). The pricing of commodity contracts. *Journal of Financial Economics*, 3(1):167–179.

Bliss, R. R. and Panigirtzoglou, N. (2004). Option-Implied Risk Aversion Estimates. *The Journal of Finance*, 59(1):407–446.

Bondarenko, O. (2000). Recovering Risk-Neutral Densities: A New Nonparametric Approach. *Working Paper*.

Bondarenko, O. (2003). Estimation of risk-neutral densities using positive convolution approximation. *Journal of Econometrics*, 116(1-2):85–112.

Breeden, D. T. and Litzenberger, R. H. (1978). Prices of State-Contingent Claims Implicit in Option Prices. *Journal of Business*, 51(4):621–651.

Bu, R. and Hadri, K. (2007). Estimating option implied risk-neutral densities using spline and hypergeometric functions. *Econometrics Journal*, 10(2):216–244.

Christoffersen, P., Fournier, M., and Jacobs, K. (2018). The Factor Structure in Equity Options. *The Review of Financial Studies*, 31(2):595–637.

Christoffersen, P., Jacobs, K., and Chang, B. Y. (2013). Forecasting with Option-Implied Information. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2 of *Handbook of Economic Forecasting*, pages 581–656. Elsevier.

Clark, I. J. and Amen, S. (2018). Implied Distributions from Risk-Reversals and Brexit/Trump Predictions. In Glau, K., Linders, D., Min, A., Scherer, M., Schneider, L., and Zagst, R., editors, *Proceedings of the Innovations in Insurance, Risk- and Asset Management Conference*, pages 117–134.

Cuesdeanu, H. and Jackwerth, J. C. (2018). The pricing kernel puzzle in forward looking data. *Review of Derivatives Research*, 21(3):253–276.

Dalderop, J. (2020). Nonparametric filtering of conditional state-price densities. *Journal of Econometrics*, 214(2):295–325.

Domowitz, I. and White, H. (1982). Misspecified models with dependent observations. *Journal of Econometrics*, 20(1):35–58.

Eriksson, A., Ghysels, E., and Wang, F. (2009). The Normal Inverse Gaussian Distribution and the Pricing of Derivatives. *The Journal of Derivatives*, 16(3):23–37.

Feng, P. and Dang, C. (2016). Shape constrained risk-neutral density estimation by support vector regression. *Information Sciences*, 333:1–9.

Fengler, M. R. and Hin, L. Y. (2015). Semi-nonparametric estimation of the call-option price surface under strike and time-to-expiry no-arbitrage constraints. *Journal of Econometrics*, 184(2):242–261.

Ferreira, A. L., Gong, Y., and Gozluklu, A. E. (2022). Risk-Corrected Probabilities of a Binary Event. *SSRN Electronic Journal*.

Figlewski, S. (2010). Estimating the Implied Risk-Neutral Density for the US Market Portfolio. In Bollerslev, T., Russell, J., and Watson, M., editors, *Volatility and Time Series Econometrics: Essays in Honor of Robert Engle*, pages 323–353. Oxford University Press.

Figlewski, S. (2018). Risk-Neutral Densities: A Review. *Annual Review of Financial Economics*, 10:329—359.

Gallant, A. R. (1977). Testing a nonlinear regression specification: A nonregular case. *Journal of the American Statistical Association*, 72(359):523–530.

Geyer, C. J. (1994). On the Asymptotics of Constrained M-Estimation. *The Annals of Statistics*, 22(4):1993–2010.

Ghysels, E. and Wang, F. (2014). Moment-implied densities: Properties and applications. *Journal of Business and Economic Statistics*, 32(1):88–111.

Hanke, M., Poulsen, R., and Weissensteiner, A. (2018). Event-Related Exchange-Rate Forecasts Combining Information from Betting Quotes and Option Prices. *Journal of Financial and Quantitative Analysis*, 53(6):2663–2683.

Härdle, W. and Hlávka, Z. (2009). Dynamics of state price densities. *Journal of Econometrics*, 150(1):1–15.

Hathaway, R. J. (1985). A Constrained Formulation of Maximum-Likelihood Estimation for Normal Mixture Distributions. *The Annals of Statistics*, 13(2):795–800.

Jackwerth, J. C. (1999). Option-implied risk-neutral distributions and implied binomial trees: A literature review. *Journal of Derivatives*, 7(2):66–82.

Jackwerth, J. C. (2004). *Option-Implied Risk-Neutral Distributions and Risk Aversion*. Charlotteville: Research Foundation of AIMR.

Jarrow, R. and Rudd, A. (1982). Approximate option valuation for arbitrary stochastic processes. *Journal of Financial Economics*, 10(3):347–369.

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

Jondeau, E. and Rockinger, M. (2001). Gram-Charlier densities. *Journal of Economic Dynamics and Control*, 25(10):1457–1483.

Kostakis, A., Mu, L., and Otsubo, Y. (2023). Detecting political event risk in the option market. *Journal of Banking & Finance*, 146:106624.

Kundu, A., Kumar, S., and Tomar, N. K. (2019). Option Implied Risk-Neutral Density Estimation: A Robust and Flexible Method. *Computational Economics*, 54(2):705–728.

Liu, X., Shackleton, M. B., Taylor, S. J., and Xu, X. (2007). Closed-form transformations from risk-neutral to real-world distributions. *Journal of Banking & Finance*, 31(5):1501–1520.

Longstaff, F. A. (1995). Option Pricing and the Martingale Restriction. *The Review of Financial Studies*, 8(4):1091–1124.

Lu, J. and Qu, Z. (2021). Sieve estimation of option-implied state price density. *Journal of Econometrics*, 224(1):88–112.

Ludwig, M. (2015). Robust estimation of shape-constrained state price density surfaces. *Journal of Derivatives*, 22(3):56–72.

Markose, S. and Alentorn, A. (2011). The Generalized Extreme Value Distribution, Implied Tail Index, and Option Pricing. *The Journal of Derivatives*, 18(3):35–60.

Melick, W. R. and Thomas, C. P. (1997). Recovering an Asset's Implied PDF from Option Prices: An Application to Crude Oil during the Gulf Crisis. *Journal of Financial and Quantitative Analysis*, 32(1):91–115.

Monteiro, A. M. and Santos, A. A. (2020). Conditional risk-neutral density from option prices by local polynomial kernel smoothing with no-arbitrage constraints. *Review of Derivatives Research*, 23(1):41–61.

Newey, W. K. and West, K. D. (1987). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3):703.

Orosi, G. (2015). Estimating Option-Implied Risk-Neutral Densities: A Novel Parametric Approach. *The Journal of Derivatives*, 23(1):41–61.

Pötscher, B. M. and Prucha, I. R. (1991a). Basic structure of the asymptotic theory in dynamic nonlineaerco nometric models, part I: consistency and approximation concepts. *Econometric Reviews*, 10(2):125–216.

Pötscher, B. M. and Prucha, I. R. (1991b). Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part II: asymptotic normality. *Econometric Reviews*, 10(3):253–325.

Ritchey, R. J. (1990). Call option valuation for discrete normal mixtures. *Journal of Financial Research*, 13(4):285–296.

Rubinstein, M. (1994). Implied Binomial Trees. *The Journal of Finance*, 49(3):771–818.

Savickas, R. (2002). A Simple Option-Pricing Formula. *The Financial Review*, 37:207–226.

Shackleton, M. B., Taylor, S. J., and Yu, P. (2010). A multi-horizon comparison of density forecasts for the S&P 500 using index returns and option prices. *Journal of Banking and Finance*, 34(11):2678–2693.

Song, Z. and Xiu, D. (2016). A tale of two option markets: Pricing kernels and volatility risk. *Journal of Econometrics*, 190(1):176–196.

Taylor, S. J. (2005). *Asset Price Dynamics, Volatility, and Prediction*. Princeton: Princeton University Press.

Wang, J. (1996). Asymptotics of least-squares estimators for constrained nonlinear regression. *Annals of Statistics*, 24(3):1316–1326.

Wang, J. (2000). Approximate representation of estimators in constrained regression problems. *Scandinavian Journal of Statistics*, 27(1):21–33.

Wang, L. (2004). Asymptotics of estimates in constrained nonlinear regression with long-range dependent innovations. *Annals of the Institute of Statistical Mathematics*, 56(2):251.

White, H. (1980). Nonlinear Regression on Cross-Section Data. *Econometrica*, 48(3):721–746.

White, H. (1981). Consequences and Detection of Misspecified Nonlinear Regression Models. *Journal of the American Statistical Association*, 76(374):419–433.

White, H. and Domowitz, I. (1984). Nonlinear Regression with Dependent Observations. *Econometrica*, 52(1):143–162.

Wu, C.-F. (1981). Asymptotic theory of nonlinear least squares estimation. *Annals of Statistics*, 9(3):501–513.

Yatchew, A. and Härdle, W. (2006). Nonparametric state price density estimation using constrained least squares and the bootstrap. *Journal of Econometrics*, 133(2):579–599.

Yuan, M. (2009). State price density estimation via nonparametric mixtures. *The Annals of Applied Statistics*, 3(3):963–984.

Online Appendix to "Parametric Risk-Neutral Density Estimation via Finite Lognormal-Weibull Mixtures"

Yifan Li — University of Manchester
Ingmar Nolte — Lancaster University
Manh Cuong Pham — Lancaster University

Abstract

In the online appendix, we present all the proofs of the paper in Online Appendix A. Computation and implementation details of the LWM method are described in Online Appendix B. Online Appendix C explains how we simulate option prices from the three-factor stochastic volatility model. Implementation of the competing RND estimators is discussed in Online Appendix D. Data cleaning procedures for our empirical application can be found in Online Appendix E. Finally, Online Appendix F contains additional simulation and empirical results.

A Proofs

Proof to Theorem 1. We start with a useful lemma which characterizes some important properties of the space $\widetilde{\Theta}_M$:

Lemma A.1. *The following results hold:*

(a) The quotient map π is a continuous function. For every $g \in \mathcal{G}$ and its dual $\widetilde{g}$, it holds that $g = \widetilde{g} \circ \pi$. In addition, if g is continuous, then so is $\widetilde{g}$.

(b) If Θ_M is compact, then so is $\widetilde{\Theta}_M$.

(c) If Θ_M is a metric space equipped with some metric d, then $\widetilde{\Theta}_M$ is also a metric space equipped with the metric $\widetilde{d}$ defined as:

$$\widetilde{d}([\theta_1], [\theta_2]) := \inf_{\theta_1 \in [\theta_1], \theta_2 \in [\theta_2]} d(\theta_1, \theta_2). \tag{1}$$

(d) If a sequence of random variables θ_n satisfies $\theta_n \to \theta$ on Θ_M, then $[\theta_n] \to [\theta]$ on $\widetilde{\Theta}_M$ with the same mode of convergence. Conversely, if $[\theta_n] \to [\theta]$ on $\widetilde{\Theta}_M$, then for an arbitrary element $\theta^* \in [\theta]$, there

exists a sequence $\theta_n^* \to \theta^*$ on Θ_M with the same mode of convergence, and each θ_n^* is a permutation of θ_n. The modes of convergence are either almost surely, in probability or in distribution.

Proof to Lemma A.1. Claim (a) is a well-known result in topology (see e.g. Singh (2019) Ch. 6.1). Claim (b) is immediate from (a) as the image of a continuous function on a compact set is also compact.

For claim (c), we need to verify that $\widetilde{d}$ is indeed a metric. The following properties are trivial: $\widetilde{d}([\theta_1],[\theta_2]) = 0 \Leftrightarrow [\theta_1] = [\theta_2]$, $\widetilde{d}([\theta_1],[\theta_2]) = \widetilde{d}([\theta_2],[\theta_1])$. We now prove the triangle inequality, that is: $\widetilde{d}([\theta_1],[\theta_3]) \leq \widetilde{d}([\theta_1],[\theta_2]) + \widetilde{d}([\theta_2],[\theta_3])$. To this end, we note that by construction we can always find some (not necessarily unique) $\theta_1^* \in [\theta_1]$ and $\theta_2^* \in [\theta_2]$ such that $\widetilde{d}([\theta_1],[\theta_2]) = d(\theta_1^*, \theta_2^*)$. The triangle inequality thus follows from:

$$\widetilde{d}([\theta_1],[\theta_3]) = d(\theta_1^*,\theta_3^*) \leq d(\theta_1^*,\theta_2^*) + d(\theta_2^*,\theta_3^*) = \widetilde{d}([\theta_1],[\theta_2]) + \widetilde{d}([\theta_2],[\theta_3]). \tag{2}$$

This proves that $\widetilde{d}$ is a metric and thus claim (c) follows.

The first part of claim (d) is a direct result of the continuous mapping theorem. For the second part, define the function $\varphi([\theta];\theta^*) = \operatorname{argmin}_{\theta\in[\theta]} d(\theta,\theta^*)$, which is clearly continuous in the first argument and satisfies $\varphi([\theta];\theta) = \theta$. The desired result then follows from the continuous mapping theorem by setting $\theta_n^* = \varphi([\theta_n];\theta^*)$, and note that θ_n^* is by construction a permutation of θ_n. This completes the proof. □

It is easy to see that the corresponding free parameter space and the associated quotient space, namely Θ_M and $\widetilde{\Theta}_M$, also satisfy Lemma A.1. Now back to the proof of Theorem 1, which largely follows from Lemmas 2.2 of White (1980) and Theorem 2.3 of White and Domowitz (1984). We firstly show that a version of the uniform law of large numbers (ULLN) holds for the function $Q_N(\vartheta)$ on Θ_M. Consider the function:

$$q_n(\vartheta) := \omega_n(u_n + O^*(K_n) - O(K_n;\vartheta))^2/2. \tag{3}$$

Since the space Θ_M is compact and $O(K,\vartheta)$, ω_n are bounded by assumption, we clearly have:

$$|q_n(\vartheta)| \leq M(O^*(K_n) - O(K_n;\vartheta))^2 + Mu_n^2, \quad \forall n \in \{1,\ldots,N\}, \tag{4}$$

where M is an upper bound of $\omega(x)$, thus $\mathrm{E}[|q_n(\vartheta)|^2] < \infty$ by Assumptions 3 and 5. By the exponential mixing rate of u_n, the conditions of Theorem 2.3 of White and Domowitz (1984) are satisfied, which

leads to the following ULLN:

$$\left|N^{-1}\sum_{n=1}^{N}(q_n(\vartheta)-\mathrm{E}[q_n(\vartheta)])\right| = \left|Q_N(\vartheta)-\mathrm{E}[Q_N(\vartheta)]\right| \xrightarrow{a.s.} 0, \tag{5}$$

uniformly on Θ_M as $N \to \infty$. As uniform convergence is preserved under a uniform continuous mapping, the ULLN holds also on the quotient space in the following sense:

$$|\widetilde{Q}_N([\vartheta]) - \mathrm{E}[\widetilde{Q}_N([\vartheta])]| \xrightarrow{a.s.} 0, \tag{6}$$

uniformly on $\widetilde{\Theta}_M$ as $N \to \infty$. To use Lemma 2.2 of White (1980), we now prove that $\mathrm{E}[\widetilde{Q}_N([\vartheta])]$ has an identifiably unique minimum at $[\vartheta^*]$. To see this, note that:

$$\mathrm{E}[\widetilde{Q}_N([\vartheta])] = (2N)^{-1}\sum_{n=1}^{N}\omega_n(\widetilde{O}(K_n;[\vartheta]) - O^*(K_n))^2 + (2N)^{-1}\sum_{n=1}^{N}\omega_n\gamma_n^2. \tag{7}$$

Both terms are Riemann sums which converge to a definite integral over $\mathcal{K}$. We can derive the following result from Assumption 2:

$$\begin{aligned}\sup_{1\le n\le N}|K_n - K_{n-1}| &= \sup_{1\le n\le N}|\kappa^{-1}\big(\frac{n}{N}\big) - \kappa^{-1}\big(\frac{n-1}{N}\big) + o(N^{-1})| \\ &= \sup_{1\le n\le N}|\kappa^{-1}(c)/N + o(N^{-1})| = O(N^{-1}),\end{aligned} \tag{8}$$

where $c \in [(n-1)/N, n/N]$ which exists due to the mean value theorem, and note that $\kappa^{-1}(x)$ is the inverse function of $\kappa(x)$ which is bounded by assumption. This result shows that the mesh of the strike grid approaches zero uniformly on $\mathcal{K}$, and similarly: $\sup_{1\le n\le N}|\kappa(K_n) - \kappa(K_{n-1})| = 1/N + o(N^{-1})$. As a result, the first sum in Eq. (7) converges pointwise to:

$$\frac{1}{2}\sum_{n=1}^{N}\omega_n(\widetilde{O}(K_n;[\vartheta]) - O^*(K_n))^2(\kappa(K_n) - \kappa(K_{n-1}) + o(N^{-1})) \to \widetilde{Q}([\vartheta])/2, \tag{9}$$

which is uniform on $\widetilde{\Theta}_M$ by a standard subsequence argument. Similarly, the second sum in Eq. (7) converges to a constant which is irrelevant for optimization purposes. By the discussion below Lemma 2.2 of White (1980), $[\vartheta^*]$ is the identifiably unique minimum of $\mathrm{E}[\widetilde{Q}_N([\vartheta])]$ by Assumption 6. As $\widetilde{\Theta}_M$ is a metric space due to Lemma A.1(c), Lemma 2 of Jennrich (1969) is applicable. Then, Lemma 2.2 of White (1980) readily applies, which shows that a sequence $[\widehat{\vartheta}_N]$ that minimizes $\widetilde{Q}_N([\vartheta])$ exists, and $[\widehat{\vartheta}_N] \xrightarrow{a.s.} [\vartheta^*]$. This completes the proof. □

Proof of Theorem 2. Theorem 1 and Lemma A.1(d) ensure that, for the sequence of estimators $[\widehat{\vartheta}_N]$ which converges a.s. to $[\vartheta^*]$, $\widehat{\vartheta}_N$ must belong to a sequence which converges a.s. to an element in

$[\vartheta^*]$. With some abuse of notation we write $\widehat{\vartheta}_N \xrightarrow{a.s.} \vartheta^*$ noting that the sequence $\widehat{\vartheta}_N$ may be permuted accordingly for each N to facilitate the convergence. It suffices to study the asymptotic distribution of the permuted sequence $\widehat{\vartheta}_N$. We start with the following lemma, which plays a central role in the asymptotic normality result.

Lemma A.2. *Let $\psi : \mathcal{K} \mapsto \mathbb{R}^k$ denote a k-dimensional bounded and element-wise Hölder continuous function and write $\psi_n := \psi(K_n)$. As $N \to \infty$, for arbitrary sequence of weights $\{\omega_n\}_{n=1:N}$ that satisfy Assumption 5, the following results hold:*

1. *(Law of Large Numbers) $N^{-1}\sum_{n=1}^{N}\omega_n\psi_n u_n \xrightarrow{p} \mathbf{0}_{k\times 1}$.*
2. *We have:*
$$\mathrm{Var}[N^{-1/2}\sum_{n=1}^{N}\omega_n\psi_n u_n] \to \mathbf{\Sigma} := \sigma_Z^2\int_{\mathcal{K}}\omega^2(x)\gamma^2(x)\psi(x)\psi(x)^\intercal d\kappa(x). \tag{10}$$
3. *(Central Limit Theorem) $N^{-1/2}\sum_{n=1}^{N}\omega_n\psi_n u_n \xrightarrow{d} \mathcal{N}(0,\mathbf{\Sigma})$.*

Proof. To prove the first claim, we can take $\{\mathcal{F}_{n,N}\}$ with $\mathcal{F}_{n,N} = \sigma(u_{1,N}, u_{2,N}, \ldots, u_{n,N})$ as the filtration of the triangular array $(u_{n,N})$, then $(\omega_n\psi_n u_{n,N}, \mathcal{F}_{n,N})$ is an uniformly integrable L_1-mixingale in view of Andrews (1988), and claim 1 follows element-wise from Theorem 1 of Andrews (1988).

For the second claim, note that:

$$\begin{aligned}
&\mathrm{Var}[N^{-1/2}\sum_{n=1}^{N}\omega_n\psi_n u_n] \\
&= \frac{1}{N}\Big(\sum_{n=1}^{N}\omega_n^2\mathrm{E}[u_n^2]\psi_n\psi_n^\intercal + \sum_{j=1}^{N-1}\sum_{n=j+1}^{N}\omega_n\omega_{n-j}\mathrm{E}[u_n u_{n-j}](\psi_n\psi_{n-j}^\intercal + \psi_{n-j}\psi_n^\intercal)\Big) \\
&= \frac{1}{N}\Big(\sum_{n=1}^{N}\omega_n^2\gamma_n^2\psi_n\psi_n^\intercal + \sum_{j=1}^{N-1}\rho_j\sum_{n=j+1}^{N}\omega_n\omega_{n-j}\gamma_n\gamma_{n-j}(\psi_n\psi_{n-j}^\intercal + \psi_{n-j}\psi_n^\intercal)\Big),
\end{aligned} \tag{11}$$

where we use the notation $\rho_j := \mathrm{E}[Z_t Z_{t-j}]$. The following result follows from a Riemann sum approximation similar to Eq. (9):

$$\frac{1}{N}\sum_{n=1}^{N}\gamma_n^2\psi_n\psi_n^\intercal \to \int_{\mathcal{K}}\gamma^2(x)\psi(x)\psi(x)^\intercal d\kappa(x). \tag{12}$$

For the second term in Eq. (11), we note that the Hölder continuity of $\gamma(x)$ implies that, there exists some $\alpha \in (0,1]$ and some finite constant C such that:

$$|\gamma_{n-j} - \gamma_n| \le C|K_n - K_{n-j}|^\alpha = O((j/N)^\alpha), \tag{13}$$

where the last estimate follows from Eq. (8). This implies that $\gamma_{n-j} = \gamma_n + O((j/N)^\alpha)$. The same

result holds for ω_n and ψ_n with possibly different C and α due to the Hölder continuity assumption, and we shall take the largest C and the smallest α of all functions henceforward. We thus have:

$$\omega_n\omega_{n-j}\gamma_n\gamma_{n-j}(\psi_n\psi_{n-j}^\intercal + \psi_{n-j}\psi_n^\intercal) = 2\omega_n^2\gamma_n^2\psi_n\psi_n^\intercal + O((j/N)^\alpha) + o((j/N)^\alpha). \tag{14}$$

We therefore see that:

$$\begin{aligned}
&\sum_{j=1}^{N-1}\rho_j\sum_{n=j+1}^{N}\omega_n\omega_{n-j}\gamma_n\gamma_{n-j}(\psi_n\psi_{n-j}^\intercal + \psi_{n-j}\psi_n^\intercal)\\
=&\sum_{j=1}^{N-1}\rho_j\Big(2\sum_{n=1}^{N}\omega_n^2\gamma_n^2\psi_n\psi_n^\intercal - 2\sum_{n=1}^{j}\omega_n^2\gamma_n^2\psi_n\psi_n^\intercal + (N-j-1)O((j/N)^\alpha)\Big)\\
=&\sum_{j=1}^{N-1}\rho_j\Big(2\sum_{n=1}^{N}\omega_n^2\gamma_n^2\psi_n\psi_n^\intercal + O(j) + (N-j-1)O((j/N)^\alpha)\Big).
\end{aligned} \tag{15}$$

As a result, with $N\to\infty$ we find that:

$$\begin{aligned}
&\frac{1}{N}\sum_{j=1}^{N-1}\rho_j\sum_{n=j+1}^{N}\omega_n\omega_{n-j}\gamma_n\gamma_{n-j}(\psi_n\psi_{n-j}^\intercal + \psi_{n-j}\psi_n^\intercal)\\
=&\sum_{j=1}^{N-1}\rho_j\Big(\frac{2}{N}\sum_{n=1}^{N}\omega_n^2\gamma_n^2\psi_n\psi_n^\intercal + O(j/N) + O((j/N)^\alpha)\Big)\\
\to&2\int_{\mathcal{K}}\omega^2(x)\gamma^2(x)\psi(x)\psi(x)^\intercal d\kappa(x)\sum_{j=1}^{\infty}\rho_j + o(1).
\end{aligned} \tag{16}$$

The last convergence follows from the fact that $\sum_{j=1}^{N} j^\alpha\rho_j \to O(1)$ for any $\alpha > 0$ due to the exponential mixing rate of Z_t. The second claim therefore follows directly from Eqs. (12) and (16).

For claim 3, it suffices to notice that u_n is near-epoch dependent on Z_n with an arbitrarily small size due to its exponential mixing rate.[1] Then by claim 2 of the lemma, the exponential mixing rate and the finite fourth moment of Z_t, the central limit result follows from Theorem 4.2(b) of Pötscher and Prucha (1991). This completes the proof. □

The theorem now follows by an application of Theorem 2.4 of Domowitz and White (1982). We do so by verifying the conditions of the theorem. First, one should realize that the conditions in Theorem 2.2 of Domowitz and White (1982) is implied by the proof of Theorem 1, where the equicontinuity of the limiting quantity $Q(\vartheta)$ holds trivially as it is unique. Second, $Q_N(\vartheta)$ is at least twice differentiable with bounded derivatives on the compact parameter space Θ_M by the functional forms of C_{LN} and C_{WB}. Third, Eq. (11) and Assumption 6 ensure that $\widehat{\vartheta}_N \in \text{int}(\Theta_M)$ for all N. Write $\nabla Q_N(\vartheta) := \partial Q_N(\vartheta)/\partial\vartheta$

[1] For a definition of near-epoch dependence and its size, see e.g. Wooldridge and White (1988).

as the $(3M-2)$-by-1 gradient vector of $Q_N(\vartheta)$, which has the following explicit form:

$$\nabla Q_N(\vartheta) = -N^{-1}\Big(\sum_{n=1}^{N}\omega_n \nabla O(K_n;\vartheta)u_n + \sum_{n=1}^{N}\omega_n \nabla O(K_n;\vartheta)(O^*(K_n) - O(K_n;\vartheta))\Big), \tag{17}$$

where $\nabla O(K;\vartheta) := \partial O(K;\vartheta)/\partial\vartheta$. Let us denote ϑ_N^* as the minimizer of $\mathrm{E}[Q_N(\vartheta)]$ (which is tail equivalent to $\widehat{\vartheta}_N$), then the second term in Eq. (17) vanishes when we evaluate $\nabla Q_N(\vartheta)$ at ϑ_N^* by construction. A direct application of Lemma A.2 to $\nabla Q_N(\vartheta_N^*)$ yields:

$$\sqrt{N}\nabla Q_N(\vartheta_N^*) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{B}^*), \tag{18}$$

where:

$$\boldsymbol{B}^* := \sigma_Z^2 \int_{\mathcal{K}} \omega^2(x)\gamma^2(x)\nabla O(x;\vartheta^*)\nabla O(x;\vartheta^*)^\intercal d\kappa(x), \tag{19}$$

which is clearly positive definite. Finally, let $\nabla^2 Q_N(\vartheta)$ denote the Hessian of $Q_N(\vartheta)$, which has the following expression:

$$\nabla^2 Q_N(\vartheta) = N^{-1}\Big(\sum_{n=1}^{N}\omega_n \nabla O(K_n;\vartheta)\nabla O(K_n;\vartheta)^\intercal - \sum_{n=1}^{N}\omega_n \nabla^2 O(K_n;\vartheta)\varepsilon_n(\vartheta)\Big). \tag{20}$$

By the ULLN used in the proof of Theorem 1 and a Riemann sum argument, we see that $|\nabla^2 Q_N(\vartheta) - \nabla^2 Q(\vartheta)/2| \xrightarrow{a.s.} 0$ uniformly for all $\vartheta \in \Theta_M$, where $Q(\vartheta)$ is the counterpart of $\widetilde{Q}([\vartheta])$ defined on the original free parameter space with the following integral form:

$$\nabla^2 Q(\vartheta) := 2\int_{\mathcal{K}} \Big(\nabla O(x;\vartheta)\nabla O(x;\vartheta)^\intercal - \nabla^2 O(x;\vartheta)(O^*(x) - O(x;\vartheta))\Big)\omega(x)d\kappa(x), \tag{21}$$

which is trivially equicontinuous. Thus all the conditions of Theorem 2.4 are now verified. Writing $\boldsymbol{A}^* = \nabla^2 Q(\vartheta^*)/2$ which is positive definite in a neighbourhood of ϑ^* by Assumption 6, Theorem 2.4 of Domowitz and White (1982) deduces the following central limit result:

$$\sqrt{N}\boldsymbol{A}^*(\widehat{\vartheta}_N - \vartheta^*) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{B}^*) = \xi, \tag{22}$$

or equivalently, $\widehat{\vartheta}_N \xrightarrow{d} \vartheta^* + N^{-1/2}(\boldsymbol{A}^*)^{-1}\xi$, up to a suitable permutation of the LHS. Then $[\widehat{\vartheta}_N] \xrightarrow{d} [\vartheta^* + N^{-1/2}(\boldsymbol{A}^*)^{-1}\xi]$ follows directly by Lemma A.1(d) , and the proof is complete.

□

Proof of Corollary 1. Take the dual of g which is denoted by $\widetilde{g}$. The almost sure convergence follows directly by applying $\widetilde{g}$ to both side of the results in Theorem 1 using $\widetilde{g}([\vartheta]) = g(\vartheta)$. For the asymptotic

normality, Theorem 2 and the continuous mapping theorem implies that:

$$g(\widehat{\vartheta}_N) = \widetilde{g}([\widehat{\vartheta}_N]) \xrightarrow{d} \widetilde{g}([\vartheta^* + N^{-1/2}(\boldsymbol{A}^*)^{-1}\xi]) = g(\vartheta^* + N^{-1/2}(\boldsymbol{A}^*)^{-1}\xi). \tag{23}$$

The corollary thus follows in the same spirit as a multivariate delta method. This completes the proof. □

Proof of Proposition 1. Notice that $\widehat{\boldsymbol{\epsilon}} = \boldsymbol{W}_N^{1/2}\widehat{\boldsymbol{\varepsilon}}$, where $\boldsymbol{W}_N$ is the N-by-N dimensional matrix of regression weights, and $\widehat{\boldsymbol{\varepsilon}}$ is the N-by-1 vector of the fitted WNLS residuals. Starting with the standard expression of $\widehat{\boldsymbol{\beta}}$ under the null:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal\boldsymbol{W}_N^{1/2}\widehat{\boldsymbol{\varepsilon}} = (\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal\boldsymbol{W}_N^{1/2}(\boldsymbol{u} + \boldsymbol{O}(\vartheta^*) - \boldsymbol{O}(\widehat{\vartheta}_N)),$$

where $\boldsymbol{u} := \{u_n\}_{n=1:N}^\intercal$ and $\boldsymbol{O}(\vartheta) := \{O(K_n;\vartheta)\}_{n=1:N}^\intercal$ are the vectorized versions of the option errors and option prices, respectively. We use the following canonical asymptotic representation of the WNLS residuals:

$$\widehat{\boldsymbol{\varepsilon}} = (\boldsymbol{I}_N - \boldsymbol{J}_O(\vartheta^*)(\boldsymbol{J}_O(\vartheta^*)^\intercal\boldsymbol{W}_N\boldsymbol{J}_O(\vartheta^*))^{-1}\boldsymbol{J}_O(\vartheta^*)^\intercal\boldsymbol{W}_N)\boldsymbol{u} + o_p(N^{-1/2}), \tag{24}$$

which follows by a mean value expansion of $\boldsymbol{O}(\vartheta^*) - \boldsymbol{O}(x)$ around ϑ^*, and the $o_p(N^{-1/2})$ estimate is due to the $\sqrt{N}$ consistency of $\widehat{\vartheta}_N$. To see this, we note that, under the null hypothesis, by a Taylor expansion of the first order condition of Eq. (11) similar to the proof of Theorem 2.4 in Domowitz and White (1982), we have:

$$\begin{aligned}\sqrt{N}(\widehat{\vartheta}_N - \vartheta^*) &= -\sqrt{N}(\boldsymbol{A}^*)^{-1}\nabla Q_N(\vartheta^*) + o_p(1)\\ &= N^{-1/2}(\boldsymbol{A}^*)^{-1}\boldsymbol{J}_O(\vartheta^*)^\intercal\boldsymbol{W}_N\boldsymbol{u} + o_p(1),\end{aligned} \tag{25}$$

due to the relation $\nabla Q_N(\vartheta^*) = -N^{-1}\boldsymbol{J}_O(\vartheta^*)^\intercal\boldsymbol{W}_N\boldsymbol{u}$. Also, under the null hypothesis of correct model specification, we have $\lim_{N\to\infty}\boldsymbol{J}_O(\vartheta^*)^\intercal\boldsymbol{W}_N\boldsymbol{J}_O(\vartheta^*)/N \to \boldsymbol{A}^*$ which follows from Eq. (21) by a Riemann sum argument.

We proceed to derive the asymptotic distribution for $\widehat{\boldsymbol{\beta}}$. Write $\boldsymbol{X}_n$ as the nth row of $\boldsymbol{X}$, which satisfies $\boldsymbol{X}_n \equiv X(K_n)$, where $X : \mathcal{K} \mapsto \mathbb{R}^{p+2q+1}$ is simply the Fourier polynomial defined in Eq. (19) (up to a change of variable). One should verify that $X(K)$ is bounded and smooth on $\mathcal{K}$. Then the following (uniform) convergence holds by a Riemann sum argument:

$$\begin{aligned}&N^{-1}\boldsymbol{X}^\intercal\boldsymbol{X} \to \int_{\mathcal{K}} X(x)^\intercal X(x)d\kappa(x) := \boldsymbol{L}_1,\\ &N^{-1}\boldsymbol{X}^\intercal\boldsymbol{W}_N^{1/2}\boldsymbol{J}_O(\vartheta)(\boldsymbol{A}^*)^{-1} \to \int_{\mathcal{K}} \sqrt{\omega(x)}X(x)^\intercal\nabla O(x;\vartheta)(\boldsymbol{A}^*)^{-1}d\kappa(x) := \boldsymbol{L}_2(\vartheta),\end{aligned} \tag{26}$$

where $\boldsymbol{J}_O(\vartheta)$ is the N-by-$(3M-2)$ Jacobian matrix of $\boldsymbol{O}(\vartheta)$, $\boldsymbol{L}_1$ and $\boldsymbol{L}_2(\vartheta)$ are a finite matrix and a smooth matrix-valued function, respectively. Then by a standard Delta method argument and in view of Eq. (26):

$$(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal\boldsymbol{W}_N^{1/2}(\boldsymbol{O}(\vartheta^*)-\boldsymbol{O}(\widehat{\vartheta}_N)) = -N^{-1}\boldsymbol{L}_1^{-1}\boldsymbol{L}_2(\vartheta^*)\boldsymbol{J}_O(\vartheta^*)^\intercal\boldsymbol{W}_N\boldsymbol{u}+o_p(N^{-1/2}). \tag{27}$$

Substituting these into the expression of $\widehat{\boldsymbol{\beta}}$, we arrive at:

$$\widehat{\boldsymbol{\beta}} = N^{-1}\boldsymbol{L}_1^{-1}(\boldsymbol{X}^\intercal\boldsymbol{W}_N^{1/2}-\boldsymbol{L}_2(\vartheta^*)\boldsymbol{J}_O(\vartheta^*)^\intercal\boldsymbol{W}_N)\boldsymbol{u}+o_p(N^{-1/2}). \tag{28}$$

From here it should be clear that $\mathrm{E}[\widehat{\boldsymbol{\beta}}]\to 0$ and $\widehat{\boldsymbol{\beta}}$ is asymptotically equivalent to a weighted average of $\{u_n\}_{n=1:N}$. A direct application of Lemma A.2.2 thus gives:

$$\lim_{N\to\infty}\mathrm{Var}[\sqrt{N}\widehat{\boldsymbol{\beta}}]\to\boldsymbol{V}^* = \sigma_Z^2\boldsymbol{L}_1^{-1}\Big(\int_{\mathcal{K}}\gamma(x)^2\boldsymbol{l}(x;\vartheta^*)\boldsymbol{l}(x;\vartheta^*)^\intercal d\kappa(x)\Big)\boldsymbol{L}_1^{-1}, \tag{29}$$

where $\boldsymbol{l}(x;\vartheta)$ is the following smooth vector-valued function:

$$\boldsymbol{l}(x;\vartheta) = X(x)^\intercal\sqrt{\omega(x)}-\boldsymbol{L}_2(\vartheta)\nabla O(x;\vartheta)\omega(x). \tag{30}$$

Clearly, $\boldsymbol{V}^*$ is positive definite by construction. Lemma A.2.3 thus concludes that $\sqrt{N}\widehat{\boldsymbol{\beta}}\xrightarrow{d}\mathcal{N}(0,\boldsymbol{V}^*)$, and hence $N\widehat{\boldsymbol{\beta}}^\intercal(\boldsymbol{V}^*)^{-1}\widehat{\boldsymbol{\beta}}\xrightarrow{d}\chi^2_{p+2q+1}$. Finally, it suffices to notice that $\widehat{\boldsymbol{V}}$ is simply the Newey-West estimator of $\boldsymbol{V}^*$ with the limiting quantities $\boldsymbol{L}_1$ and $\boldsymbol{l}(x;\vartheta)$ replaced by their sample counterparts, whose consistency and positive semi-definiteness readily follows. This completes the proof.

□

B Computational Details of the LWM Method

This section documents the computational details of the M-LWM method used in our simulation and empirical analyses, including (1) a global search algorithm to estimate the model parameters; (2) various closed-form expressions for the derivatives of the WNLS regression.

B.1 The Global Search Algorithm

The global search algorithm is designed to solve the problem Eq. (11) for the branching step in our sequential testing procedure. We present the details of this algorithm below.

Step 1: Initialization. Here we present some details of the initialization step in our sequential test procedure. We solve Eq. (11) with both $M_1 = 0$ and $M_1 = 1$ to obtain the estimated scale

parameters $\widehat{k}^{(0)}$ and $\widehat{\sigma}^{(0)}$, and the optimized objective function values $Q_N^{(1)}(\widehat{k}^{(0)};0)$ and $Q_N^{(1)}(\widehat{\sigma}^{(0)};1)$. As this is a univariate optimization problem, obtaining parameter estimates is fairly straightforward. For example, one can set the initial guess of $\widehat{\sigma}^{(0)}$ to be σ_{ATM}, where σ_{ATM} is the at-the-money Black-implied volatility. This can be conveniently computed using the following closed-form expression (see Orlando and Taglialatela (2017)):

$$\sigma_{ATM} = \sqrt{\frac{8}{\tau}} \text{erf}^{-1}\left(\frac{O(F_t)}{F_t}\right), \tag{31}$$

where $\text{erf}^{-1}(x)$ is the inverse error function, and $O(F_t)$ is the at-the-money call option price, which can be computed from the observed option prices by interpolation. Let us denote $\overline{Q} = Q_N^{(1)}(\widehat{k}^{(0)};0) \wedge Q_N^{(1)}(\widehat{\sigma}^{(0)};1)$, which serves as an upper bound for the optimization problem when we add more mixtures.

Step 2: Branching. In this step, we detail the branching step in our sequential test procedure. To this end, we shall assume that we obtained the $(M-1)$-LWM parameter estimate $\widehat{\theta}_{M_1^{(M-1)}}$, where the subscript indicates both the number of mixtures and the number of lognormal components, and the dependence on N is suppressed for convenience. We shall also assume that the sequential test rejects the $M-1$ mixture case with $M < \overline{M}$, so it is necessary to estimate an M-LWM model.

In the branching step, we build a M-LWM model by adding either a lognormal or a Weibull component to the estimated $(M-1)$-LWM model. As both models nest the estimated $(M-1)$-LWM model, the minimized objective function value $\overline{Q} = Q_N(\widehat{\theta}_{M_1^{(M-1)}})$ serves as a natural upper bound for the M-mixture objective function. We discuss how to estimate this M-LWM model in the following substeps.

Substep 2.1: Generating Initial Guesses. We estimate the M-LWM model using a standard non-linear constrained programme minimizer (e.g., `fmincon` in MATLAB[2]). To this end, we need to provide an initial guess of the solution. Based on the evolutionary estimation of the MLN model discussed in Bondarenko (2003), we shall use the parameter estimates $\widehat{\theta}_{M_1^{(M-1)}}$ from the previous $(M-1)$-mixture as the starting values for the first $M-1$ mixtures. We thus only need to provide an initial guess to the parameters of the additional Mth density, which we denote as $F^{(M)}$, $\sigma^{(M)}$ (or $k^{(M)}$ for the Weibull case). We draw these uniformly from the following intervals:

$$F^{(M)} \in [e^{-3\sqrt{\tau}\sigma_{ATM}}F_t, e^{\sqrt{\tau}\sigma_{ATM}}F_t], \quad \sigma^{(M)} \in [\frac{\widehat{\sigma}^{(0)}}{2}, \frac{3\widehat{\sigma}^{(0)}}{2}], \quad k^{(M)} \in [\frac{\widehat{k}^{(0)}}{2}, \frac{3\widehat{k}^{(0)}}{2}]. \tag{32}$$

Intuitively, $F^{(M)}$ is drawn from the range that corresponds to a log-moneyness between -3 and 1, while $\sigma^{(M)}$ and $k^{(M)}$ are proportional to the parameter estimates in the $M=1$ case, which serves as

[2]Further computational details of the implementation of our LWM method can be found in the supplementary material of the paper.

benchmarks to determine the relative sizes of the initial guesses. The constants in the ranges above are only chosen to ensure that the additional density is not too ill-behaved, and different reasonable choices do not affect the optimization performance.

To obtain initial guesses for $w^{(i)}$ given the initial guesses for $F^{(i)}$, $\sigma^{(i)}$ and $k^{(i)}$, we simply optimize Eq. (11) by holding $F^{(i)}$, $\sigma^{(i)}$ and $k^{(i)}$ constant. In this case, Eq. (11) reduces to a quadratic programming problem, which can be efficiently solved by standard quadratic programming packages (e.g., `quadprog` in MATLAB). Combining the initial guesses for $w^{(i)}$ with $F^{(i)}$, $\sigma^{(i)}$ and $k^{(i)}$, we arrive at a random starting point denoted by θ_{ig}. To avoid a bad random draw of the additional density, we repeatedly draw θ_{ig} 100 times and choose the starting point as the best θ_{ig} with the smallest objective function of Eq. (11). Notice that θ_{ig} always improves over $\widehat{\theta}_{M_1^{(M-1)}}$ in the sense that $\overline{Q} > Q_N^{(M)}(\theta_{ig})$ by construction.

Substep 2.2: Local Optimization. Using θ_{ig} obtained from the previous step as starting points, we solve Eq. (11) twice to obtain a set of candidate solutions $\widehat{\theta}_{M_1^{(M^-)}}$ and $\widehat{\theta}_{M_1^{(M^+)}}$, where $M_1^{(M^-)} = M_1^{(M-1)}$ (so that the additional density is Weibull), while $M_1^{(M^+)} = M_1^{(M^-)} + 1$. We use the analytical gradient and Hessian in this procedure to improve the computational efficiency and stability, whose expressions can be found in Section B.2. Note that we discard the solution $\widehat{\theta}_{M_1^{(M^\pm)}}$ if $Q_N^{(M)}(\widehat{\theta}_{M_1^{(M^\pm)}}) > \overline{Q}$, or if the constraints on the parameters are not satisfied. These are infeasible solutions to Eq. (11) that should be avoided.

Substep 2.3: Initial Value Perturbation. We repeat substeps 2.1 and 2.2 to generate multiple solutions of $\widehat{\theta}_{M_1^{(M^\pm)}}$ which may be different due to the randomness in θ_{ig}, and we retain the solution with the smallest objective function value. In our simulation and empirical analysis, we pick the best solutions $\widehat{\theta}_{M_1^{(M^\pm)}}$ out of five candidate solutions to Eq. (11).

Substep 2.4: Choosing the LWM mixture. In view of Eq. (23), $\widehat{\theta}_{M_1^{(M^\pm)}}$ solves the optimization problem for ϑ_N for the given choice of $M_1 \in \{M_1^{(M^-)}, M_1^{(M^+)}\}$. The choice of $M_1^{(M)}$ that solves Eq. (23) is now obvious:

$$M_1^{(M)} = \underset{M_1 \in \{M_1^{(M^-)}, M_1^{(M^+)}\}}{\operatorname{argmin}} Q_N(\widehat{\theta}_{M_1}; M_1), \tag{33}$$

and the pair $(\widehat{\theta}_{M_1^{(M)}}, M_1^{(M)})$ is understood as the solution to Eq. (23).

Step 3: Testing and Iteration. With $(\widehat{\theta}_{M_1^{(M)}}, M_1^{(M)})$, we can proceed to the testing step in the sequential test procedure and check whether the current model is misspecified. If the test is rejected and $M < \overline{M}$, we go back to Step 2 and use $\widehat{\theta}_{M_1^{(M)}}$ as the prior parameter estimate which serves as the starting point of the $(M+1)$-mixture case. We also update $\overline{Q} = Q_N(\widehat{\theta}_{M_1^{(M)}}; M_1^{(M)})$ as the new upper bound for the numerical optimization. Otherwise, we terminate the sequential test procedure with the solution $(\widehat{\theta}_{M_1^{(M)}}, M_1^{(M)})$.

B.2 Analytical Expressions for the Gradients and Hessians

This section gives the analytical formulas for the gradients and Hessians of the optimization problem in Eq. (11), which are also used in the construction of the diagnostic tests. Before we state our results, we introduce some notation for the extended option Greeks based on a lognormal RND:

$$\begin{aligned}
\Delta_{LN} &:= \frac{\partial C_{LN}}{\partial F} = e^{-r_f\tau}\Phi(d_1), \quad \upsilon_{LN} := \frac{\partial C_{LN}}{\partial \sigma} = e^{-r_f\tau}F\phi(d_1)\sqrt{\tau}, \\
\Gamma_{LN} &:= \frac{\partial^2 C_{LN}}{\partial F^2} = \frac{e^{-r_f\tau}\phi(d_1)}{\sigma F\sqrt{\tau}}, \quad \varsigma_{LN} := \frac{\partial^2 C_{LN}}{\partial \sigma^2} = \upsilon d_1 d_2/\sigma, \\
\eta_{LN} &:= \frac{\partial^2 C_{LN}}{\partial \sigma \partial F} = -\frac{e^{-r_f\tau}\phi(d_1)d_2}{\sigma},
\end{aligned} \tag{34}$$

where C_{LN} corresponds to the Black (1976) model for call option in Eq. (9) and $\phi(\cdot)$ is the PDF of a standard normal distribution. The put option counterparts of the above Greeks can also be derived analogously, which is omitted for brevity.

We also present some option Greeks based on a Weibull RND, and for notational convenience, we shall also denote the scale parameter of the Weibull density using σ instead of k:

$$\begin{aligned}
\Delta_{WB} &:= \frac{\partial C_{WB}}{\partial F} = e^{-r_f\tau}\Upsilon_{1+\sigma}(g(K;F,\sigma)), \\
\Gamma_{WB} &:= \frac{\partial^2 C_{WB}}{\partial F^2} = e^{-r_f\tau}\frac{K^2}{F^2}f_{WB}(K;F,\sigma).
\end{aligned} \tag{35}$$

where C_{WB} corresponds to the call option pricing function in Eq. (6). The derivatives w.r.t. σ are denoted as υ_{WB}, ς_{WB} and η_{WB}, defined similar to those in Eq. (34). However, the expressions of these derivatives involve Meijer's G function, which is a special non-elementary function that is slow to compute in MATLAB. We therefore compute them numerically and do not report the analytical expressions here.

Let us fix some M and M_1. Consider the following N-by-M matrices, understood as the matrix of component-wise option prices and the elementwise derivatives:

$$\begin{aligned}
&\boldsymbol{O} = [O_n^{(i)}]_{n=1:N,i=1:M}, \boldsymbol{\Delta} = [\Delta_n^{(i)}]_{n=1:N,i=1:M}, \boldsymbol{\upsilon} = [\upsilon_n^{(i)}]_{n=1:N,i=1:M}, \\
&\boldsymbol{\Gamma} = [\Gamma_n^{(i)}]_{n=1:N,i=1:M}, \boldsymbol{\varsigma} = [\varsigma_n^{(i)}]_{n=1:N,i=1:M}, \boldsymbol{\eta} = [\eta_n^{(i)}]_{n=1:N,i=1:M},
\end{aligned} \tag{36}$$

where n is the row index and i is the column index. All the elements in the matrix depend on the type of mixture implicitly. For example, $O_n^{(i)}$ is the option price for the ith mixture with a strike price K_n, thus $O_n^{(i)} = O_{LN}(K_n; F^{(i)}, \sigma^{(i)})$ if $i \leq M_1$ and $O_n^{(i)} = O_{WB}(K_n; F^{(i)}, \sigma^{(i)})$ otherwise. Elements in $\boldsymbol{\upsilon}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\varsigma}$ and $\boldsymbol{\eta}$ are defined analogously.

We list the gradient vector and the Hessian matrix of $Q_N(\theta)$ in terms of the option Greeks in the

result below:

Proposition 1. *It holds that:*

$$\begin{aligned}\frac{\partial Q_N(\theta)}{\partial \theta} &= -\frac{1}{N}\boldsymbol{J}_O(\theta)^\intercal \boldsymbol{W}_N \boldsymbol{\varepsilon}(\theta),\\ \frac{\partial^2 Q_N(\theta)}{\partial \theta \partial \theta^\intercal} &= \frac{1}{N}\left[\boldsymbol{J}_O(\theta)^\intercal \boldsymbol{W}_N \boldsymbol{J}_O(\theta) - \sum_{n=1}^{N} \nabla_\theta^2 O(K_n;\theta)\omega_n \varepsilon_n(\theta)\right],\end{aligned} \tag{37}$$

where $\boldsymbol{W}_N$ *is the* N*-by-*N *diagonal matrix of regression weights,* $\boldsymbol{J}_O(\theta) = [\boldsymbol{O}, \boldsymbol{\Delta w}, \boldsymbol{\upsilon w}]$ *is the* N*-by-*$3M$ *Jacobian matrix, in which* $\boldsymbol{w}$ *is the* M*-by-*M *diagonal matrix with entries* $\{w^{(i)}\}_{i=1:M}$*. The* $3M$*-by-*$3M$ *Hessian matrix* $\nabla_\theta^2 O(K_n;\theta) := \frac{\partial^2 O(K_n;\theta)}{\partial\theta\partial\theta^\intercal}$ *has the following non-zero elements:*

$$\begin{aligned}&\frac{\partial^2 O(K_n;\theta)}{\partial \sigma^{(i)2}} = w^{(i)}\varsigma_n^{(i)}, \quad \frac{\partial^2 O(K_n;\theta)}{\partial \sigma^{(i)}\partial w^{(i)}} = \upsilon_n^{(i)}, \quad \frac{\partial^2 O(K_n;\vartheta)}{\partial \sigma^{(i)}\partial F^{(i)}} = w^{(i)}\eta_n^{(i)},\\ &\frac{\partial^2 O(K_n;\theta)}{\partial F^{(i)2}} = w^{(i)}\Gamma_n^{(i)}, \quad \frac{\partial^2 O(K_n;\theta)}{\partial F^{(i)}\partial w^{(i)}} = \Delta_n^{(i)}.\end{aligned} \tag{38}$$

Proof. The proposition follows from straightforward matrix calculus. □

The (partially) analytical derivatives can be used in solving the WNLS problem in Eq. (11) with two equality constraints. As a result, one also need to include two Lagrange functions in the optimization procedure, which is used in our local optimizer in the next section.

Alternatively, one can also optimize Eq. (11) in the free parameter space and consider the following

$$\begin{aligned}&\widehat{\vartheta}_N = \underset{\vartheta\in\Theta_M}{\operatorname{argmin}}\, Q_N(\vartheta) := \frac{1}{2N}\sum_{n=1}^{N}\omega_n\varepsilon_n(\vartheta)^2\\ \text{s.t.}\quad &w^{(i)}, F^{(i)}, \sigma^{(i)}, k^{(i)} > 0, \quad \sum_{i=1}^{M-1} w^{(i)} < 1, \quad \sum_{i=1}^{M-1} w^{(i)}F^{(i)} < F_t.\end{aligned} \tag{39}$$

The numerical solution to this problem is less stable that that of Eq. (11), but the analytical derivatives of $Q_N(\vartheta)$ can be used to construct confidence bands for various RND-related quantities (e.g., in Theorem 2 and Proposition 1), which are presented below:

Proposition 2. *It holds that:*

$$\begin{aligned}\frac{\partial Q_N(\vartheta)}{\partial \vartheta} &= -\frac{1}{N}\boldsymbol{J}_O(\vartheta)^\intercal \boldsymbol{W}_N \boldsymbol{\varepsilon}(\vartheta),\\ \frac{\partial^2 Q_N(\vartheta)}{\partial \vartheta \partial \vartheta^\intercal} &= \frac{1}{N}\left[\boldsymbol{J}_O(\vartheta)^\intercal \boldsymbol{W}_N \boldsymbol{J}_O(\vartheta) - \sum_{n=1}^{N} \nabla_\vartheta^2 O(K_n;\vartheta)\omega_n \varepsilon_n(\vartheta)\right],\end{aligned} \tag{40}$$

where $\boldsymbol{J}_O(\vartheta)$ *is the following* N*-by-*$(3M-2)$ *Jacobian matrix:*

$$\boldsymbol{J}_O(\vartheta) = [\widetilde{\boldsymbol{O}} - \boldsymbol{\Delta}^{(M)}\widetilde{\boldsymbol{F}}, (\widetilde{\boldsymbol{\Delta}} - \boldsymbol{\Delta}^{(M)})\widetilde{\boldsymbol{w}}, \boldsymbol{\upsilon w}], \tag{41}$$

in which $\widetilde{\boldsymbol{O}}$ *and* $\boldsymbol{O}^{(M)}$ *(resp.* $\widetilde{\boldsymbol{\Delta}}$ *and* $\boldsymbol{\Delta}^{(M)}$*) are the first* $M-1$ *and the last* M *columns of* $\boldsymbol{O}$ *(resp.* $\boldsymbol{\Delta}$*),* $\widetilde{\boldsymbol{F}}$ *is the* $M-1$ *dimensional diagonal matrix whose entries are given by* $\{F^{(i)}-F^{(M)}\}_{i=1:M-1}$*, and* $\widetilde{\boldsymbol{w}}$ *is the* $M-1$ *dimensional submatrix of* $\boldsymbol{w}$ *by dropping the last column and row.* $\nabla^2_{\vartheta}O(K_n;\vartheta) := \frac{\partial^2 O(K_n;\vartheta)}{\partial\vartheta\partial\vartheta^\intercal}$ *is the* $(3M-2)$*-by-*$(3M-2)$ *Hessian matrix with non-zero entries given by:*

$$
\begin{aligned}
&\frac{\partial^2 O(K_n;\vartheta)}{\partial\sigma^{(i)2}} = w^{(i)}\varsigma_n^{(i)}, \quad \frac{\partial^2 O(K_n;\vartheta)}{\partial\sigma^{(i)}\partial w^{(i)}} = \upsilon_n^{(i)}, \quad \frac{\partial^2 O(K_n;\vartheta)}{\partial\sigma^{(i)}\partial F^{(i)}} = w^{(i)}\eta_n^{(i)},\\
&\frac{\partial^2 O(K_n;\vartheta)}{\partial\sigma^{(i)}\partial w^{(i')}} = -\upsilon_n^{(M)}, \quad \frac{\partial^2 O(K_n;\vartheta)}{\partial\sigma^{(M)}\partial w^{(i)}} = \eta_n^{(M)}(F^{(M)}-F^{(i)}) - \upsilon_n^{(M)},\\
&\frac{\partial^2 O(K_n;\vartheta)}{\partial\sigma^{(M)}\partial F^{(i)}} = -\eta_n^{(M)}w^{(i)},\\
&\frac{\partial^2 O(K_n;\vartheta)}{\partial w^{(i)}\partial w^{(i')}} = \Gamma_n^{(M)}\frac{(F^{(M)}-F^{(i)})(F^{(M)}-F^{(i')})}{w^{(M)}},\\
&\frac{\partial^2 O(K_n;\vartheta)}{\partial w^{(i)}\partial F^{(i)}} = \Delta_n^{(i)} - \Delta_n^{(M)} - \frac{w^{(i)}}{w^{(M)}}(F^{(M)}-F^{(i)})\Gamma_n^{(M)},\\
&\frac{\partial^2 O(K_n;\vartheta)}{\partial w^{(i)}\partial F^{(i')}} = -\frac{w^{(i')}}{w^{(M)}}(F^{(M)}-F^{(i)})\Gamma_n^{(M)},\\
&\frac{\partial^2 O(K_n;\vartheta)}{\partial F^{(i)2}} = w^{(i)}\Big(\Gamma_n^{(i)} + \frac{w^{(i)}}{w^{(M)}}\Gamma_n^{(M)}\Big), \quad \frac{\partial^2 O(K_n;\vartheta)}{\partial F^{(i)}\partial F^{(i')}} = \frac{w^{(i)}w^{(i')}}{w^{(M)}}\Gamma_n^{(M)}.
\end{aligned}
\tag{42}
$$

where $i' \neq i \in \{1,\dots,M-1\}$.

Proof. The proposition follows from straightforward matrix calculus. □

The slightly more complicated forms are due to the dimension reduction for the parameter vector, which absorbs the two equality constraints and thus introduces dependence between the option price of the last mixture and the parameters in the first $M-1$ mixtures.

C Simulation of RND and Option Prices From Three-Factor Double Exponential Stochastic Volatility Models

We consider the three-factor double exponential stochastic volatility model of Andersen et al. (2015a,b) for the forward price F_t in the risk-neutral measure, which is specified as follows[3]:

$$
\begin{aligned}
\frac{dF_t}{F_{t-}} &= \sqrt{V_{1,t}}dW^{\mathbb{Q}}_{1,t} + \sqrt{V_{2,t}}dW^{\mathbb{Q}}_{2,t} + \eta\sqrt{U_t}dW^{\mathbb{Q}}_{3,t} + \int_{\mathbb{R}^2}(e^x-1)\widetilde{\mu}^{\mathbb{Q}}(dt,dx,dy),\\
dV_{1,t} &= \kappa_1(\bar{v}_1 - V_{1,t})dt + \sigma_1\sqrt{V_{1,t}}dB^{\mathbb{Q}}_{1,t} + \mu_v\int_{\mathbb{R}^2}x^2 1_{\{x<0\}}\mu(dt,dx,dy),\\
dV_{2,t} &= \kappa_2(\bar{v}_2 - V_{2,t})dt + \sigma_2\sqrt{V_{2,t}}dB^{\mathbb{Q}}_{2,t},\\
dU_t &= -\kappa_u U_t dt + \mu_u\int_{\mathbb{R}^2}[(1-\rho_u)x^2 1_{\{x<0\}} + \rho_u y^2]\mu(dt,dx,dy),
\end{aligned}
\tag{43}
$$

where $(W^{\mathbb{Q}}_{1,t}, W^{\mathbb{Q}}_{2,t}, W^{\mathbb{Q}}_{3,t}, B^{\mathbb{Q}}_{1,t}, B^{\mathbb{Q}}_{2,t})$ is a five-dimensional Brownian motion with $\mathrm{corr}(W^{\mathbb{Q}}_{1,t}, B^{\mathbb{Q}}_{1,t}) = \rho_1$, $\mathrm{corr}(W^{\mathbb{Q}}_{2,t}, B^{\mathbb{Q}}_{2,t}) = \rho_2$, and the remaining Brownian motions are mutually independent. Jumps in the

[3] As the model is specified in the forward price, it is assumed that both risk-free rate and dividend yield are 0.

forward price F and the state vector (V_1, V_2, U) are captured by an integer-valued counting measure μ. The corresponding (instantaneous) jump intensity, under the risk-neutral measure, is $dt \otimes v_t^{\mathbb{Q}}(dx, dy)$. The difference $\widetilde{\mu}^{\mathbb{Q}}(dt, dx, dy) = \mu(dt, dx, dy) - dt v_t^{\mathbb{Q}}(dx, dy)$ constitutes the associated martingale jump measure.

The jump specification includes two separate components, x and y. The former captures co-jumps that occur simultaneously in the forward price F_t, the first volatility factor $V_{1,t}$ and potentially in the third volatility factor U_t (if $\rho_u < 1$), while the latter represents independent shocks to the U_t factor, and potentially a jump in return volatility if $\eta > 0$. The jump compensator is given by

$$\frac{v_t^{\mathbb{Q}}(dx, dy)}{dxdy} = \begin{cases} c^-(t) \cdot 1_{\{x<0\}} \lambda_- e^{-\lambda_- |x|} + c^+(t) \cdot 1_{\{x>0\}} \lambda_+ e^{-\lambda_+ x} & \text{if } y = 0, \\ c^-(t) \lambda_- e^{-\lambda_- |y|} & \text{if } x = 0 \text{ and } y < 0. \end{cases} \tag{44}$$

The first term on the right-hand side, corresponding to the case $x \neq 0$ and $y = 0$, indicates co-jumps in both the forward price and volatility, while the second term, $x = 0$ and $y < 0$, represents the independent shocks to the U factor. The individual (strictly positive) jumps in U are either independent from V_1 (when $\rho_u = 1$) or proportional to the (instantaneous) jumps in V_1 (when $\rho_u = 0$). The jumps in the forward price are modelled as exponentially distributed with separate tail decay parameters λ_- and λ_+ for negative and positive jumps. For parsimony, the independent shocks to the U factor are distributed identically to the negative price jumps. The time-varying jump intensities are governed by the $c^-(t)$ and $c^+(t)$ coefficients, which evolve as affine functions of the state vector,

$$\begin{aligned} c^-(t) &= c_0^- + c_1^- V_{1,t,-} + c_2^- V_{2,t,-} + c_u^- U_{t,-}, \\ c^+(t) &= c_0^+ + c_1^+ V_{1,t,-} + c_2^+ V_{2,t,-} + c_u^+ U_{t,-}. \end{aligned} \tag{45}$$

Under the three-factor double exponential stochastic volatility model in Eq. (43), the spot diffusive variance of the forward return is given by

$$V_t = V_{1,t} + V_{2,t} + \eta^2 U_t. \tag{46}$$

In our simulation, we consider three scenarios, namely "Low Volatility", "Mid Volatility" and "High Volatility", whose common parameters are

$$\begin{aligned} &\rho_1 = -0.9818, \bar{v}_1 = 0.0084, \kappa_1 = 9.7196, \sigma_1 = 0.3924, \rho_2 = -0.8707, \\ &\bar{v}_2 = 0.0391, \kappa_2 = 0.1680, \sigma_2 = 0.1078, \mu_u = 0.9238, \kappa_u = 0.5967, \\ &\rho_u = 0.0005, c_0^- = 0, c_0^+ = 1.5713, c_1^- = 25.3536, c_1^+ = 92.4094, \\ &c_2^- = 0.8802, c_2^+ = 72.5628, c_u^- = 41.4017, c_u^+ = 0, \mu_v = 13.4143, \eta = 0, \end{aligned} \tag{47}$$

and other scenario-specific parameters are listed in Table C.1. Parameters corresponding to the "Mid Volatility" case, excluding the futures and spot volatility values, follow from Table VI of Andersen et al. (2015a), which generates a negatively skewed RND function shown in Fig. 4.1. The parameters for the "Low Volatility" and "High Volatility" cases are slightly modified from the "Mid Volatility" parameters to produce alternative RNDs with different symmetrical and dispersion properties, as depicted in Fig. 4.1. We also consider a bimodal RND, which is formed by linearly combining the RNDs in DGP II and III, with the respective weights of 0.2 and 0.8.

	DGP I	DGP II	DGP III
Parameter	"Low Volatility"	"Mid Volatility"	"High Volatility"
Jump parameters			
λ_-	8.7455	18.7455	58.7455
λ_+	58.2399	58.2399	18.2399
Futures and Spot Volatility values			
F_t	300	300	450
$V_{1,t}$	0.01	0.01	0.01
$V_{2,t}$	0.01	0.03	0.1
U_t	0.1	0.5	2

Table C.1: Parameters of three-factor double exponential stochastic volatility models. All other parameters are taken from Table VI of Andersen et al. (2015a) and detailed in Eq. (47).

Let $\phi(\omega; V_t, F_t, T) = \mathrm{E}[e^{i\omega \ln F_T}|\mathcal{F}_t]$ denote the conditional characteristic function of the three-factor double exponential stochastic volatility model specified in Eq. (43),[4] the RND of $\ln F_T$, denoted by $\breve{f}_t$, can be derived from the inverse Fourier transform of the characteristic function:

$$\breve{f}_t(x) = \frac{1}{2\pi}\int_0^\infty \phi(\omega; V_t, F_t, T)e^{-i\omega x}d\omega. \tag{48}$$

Consequently, the RND of F_T is given by $f_t(x) = \breve{f}_t(\ln x)/x$, and the price of an European call option with strike K which matures at time T is:

$$C(K) = e^{-r_f(T-t)}\int_0^\infty (F_T - K)^+ f_t(x)dx. \tag{49}$$

We evaluate $C(K)$ and $\breve{f}_t(x)$ using the Fourier cosine approximation of Fang and Oosterlee (2008). In detail, for some interval $[a, b]$ and a truncation of N terms of an infinite sum, $\breve{f}_t(x)$ can be approximated by:

$$\breve{f}_t(x) \approx \sum_{k=0}^{N} 2^{-1_{\{k=0\}}}\frac{2}{b-a}\Re\Big\{\phi\Big(\frac{k\pi}{b-a}; V_t, F_t, T\Big)\exp\Big(-i\frac{k\pi}{b-a}\Big)\Big\}\cos\Big(k\pi\frac{x-a}{b-a}\Big), \tag{50}$$

where $\Re\{x\}$ returns the real part of a complex number x. Here $[a, b]$ should cover most of the probability

[4] See Appendix C of Andersen et al. (2015b) for more details.

masses of $\ln F_T$, and a larger N always improves the precision of the approximation. In our result we use $a = 0, b = 10$ and $N = 2000$. The call option prices are generated using the following set of equations:

$$\begin{aligned}
C(K) &\approx e^{-r_f(T-t)} \sum_{k=0}^{N} 2^{-1_{\{k=0\}}} \Re\left\{ \phi\left(\frac{k\pi}{b-a}; V_t, F_t, T\right) \exp\left(ik\pi \frac{\ln\frac{F_t}{K} - a}{b-a}\right)\right\} U_k \\
U_k &= \frac{2}{b-a} K(\chi_k - \psi_k) \\
\chi_k &= \frac{1}{1+(\frac{k\pi}{b-a})^2}\left[\cos(k\pi)e^b - \cos\left(\frac{-ak\pi}{b-a}\right) + \frac{k\pi}{b-a}\left(\sin(k\pi)e^b - \sin\left(\frac{-ak\pi}{b-a}\right)\right)\right] \\
\psi_k &= \begin{cases} b, & k = 0 \\ \frac{b-a}{k\pi}\left(\sin(k\pi) - \sin\left(\frac{-ak\pi}{b-a}\right)\right), & k > 0 \end{cases}
\end{aligned} \tag{51}$$

Different from Eq. (50), the interval $[a, b]$ above should be chosen to cover the probability masses of $\ln\frac{F_t}{K}$. In our implementation we use $a = -3, b = 3, N = 2000$. We note that the above method is applicable to any affine stochastic volatility model with known characteristic function.

D Implementation of the Competing RND Estimators

D.1 Positive Convolution Approximation

We follow Bondarenko (2000) and Bondarenko (2003) closely to implement the PCA method. For a given bandwidth parameter h, we firstly set the grid size $\Delta z = 0.5h$ as recommended by Bondarenko (2003), which determines the number of normal mixtures M as $M = \lceil m(\mathcal{K})/\Delta z + 1\rceil$. The centres of the normal mixtures are therefore given by $\{z_j\}_{j=1:M}$ which is equidistantly placed on the strike range $\mathcal{K}$. The PCA method assumes that the true RND $f_t^*(x)$ can be approximated by a discrete mixture of normal densities $g_t(x; h)$ on $\mathcal{K}$ which takes the following form:

$$g_t(x; h) = \sum_{j=1}^{M} \frac{a_j}{h} \phi\Big(\frac{x - z_j}{h}\Big), \quad a_j \geq 0, x \in \mathcal{K}, \tag{52}$$

where the parameters to be estimated are $\{a_j\}_{j=1:M}$. Note that the left tail of $g_t(x; h)$ for $x < \underline{K}$ and the right tail of $g_t(x; h)$ for $x > \overline{K}$ are left unspecified in the above expression. Only the left tail plays a significant role in the estimation procedure, which is summarized as two additional parameters $b = \int_0^K g_t(x; h)dx$ and $B = \int_0^K x g_t(x; h)dx$. The following two linear constraints are imposed for

$g_t(x;h)$ to satisfy Assumption 1:

$$b+\sum_{j=1}^{M} a_j \leq 1, \quad B+\sum_{j=1}^{M} a_j z_j \leq F_t. \tag{53}$$

At the n-th strike on the strike grid $\{K_n\}_{n=1:N}$, the PCA-implied put option price takes the following form:

$$P_n(\theta;h) = -B + bK_n + \sum_{j=1}^{M} a_j \mathcal{Z}_j(K_n,h), \tag{54}$$

where $\mathcal{Z}_j(K_n,h) = h\phi(\frac{K_n-z_j}{h}) + (K_n - z_j)\Phi(\frac{K_n-z_j}{h})$ is the second-order integral of the j-th normal mixture with a bandwidth of h evaluated at K_n, and $\theta = \{a_1,\ldots,a_M,b,B\}$ denotes the parameter vector to be estimated. Denote the observed put option prices as $\{P_n\}_{n=1:N}$, we estimate θ by solving the following quadratic programming problem:

$$\begin{aligned} &\underset{\theta\in[0,\infty)^{M+2}}{\operatorname{argmin}} \sum_{n=1}^{N}(P_n - P_n(\theta;h))^2, \\ s.t.\ &b+\sum_{j=1}^{M} a_j \leq 1, \quad B+\sum_{j=1}^{M} a_j z_j \leq F_t, \\ &B - b(\underline{K} - 2.5h) \leq 0, \quad C - c(\overline{K} + 2.5h) \leq 0, \end{aligned} \tag{55}$$

where $c = 1 - b - \sum_{j=1}^{M} a_j$, $C = F_t - B - \sum_{j=1}^{M} a_j z_j$ and the latter two constraints follow from the smoothness constraints as in footnote 22 of Bondarenko (2000). The above problem is implemented in MATLAB R2019a using the Gurobi optimizer[5], which is highly efficient in solving quadratic programming problems.

For each simulation scenario, we optimize the bandwidth h based on a grid search in the range [3,70] to minimize the simulated RMISE defined in Eq. (26) of the estimated RND. Figure D.1 presents the grid search result and the choice of h for each simulation scenario under the baseline i.i.d. option error design and a large error variance ($4\times$), zooming in the local range of h that contains the optimal $\widetilde{h}$. The figure clearly shows the bias-variance trade-off for the choice of h, and the optimal bandwidth $\widetilde{h}$ is chosen to minimize the simulated RMISE, which is infeasible in practice.

D.2 Local Linear Regression

The local linear regression method is implemented based on Aït-Sahalia and Duarte (2003). Given a sample of observed call option prices $(C_n,K_n)_{n=1:N}$, the estimation consists of two steps. In the first step, we fit a constrained least square regression to obtain the smoothed option prices $\widetilde{C}_n$ based

[5]The optimizer is available at gurobi.com.

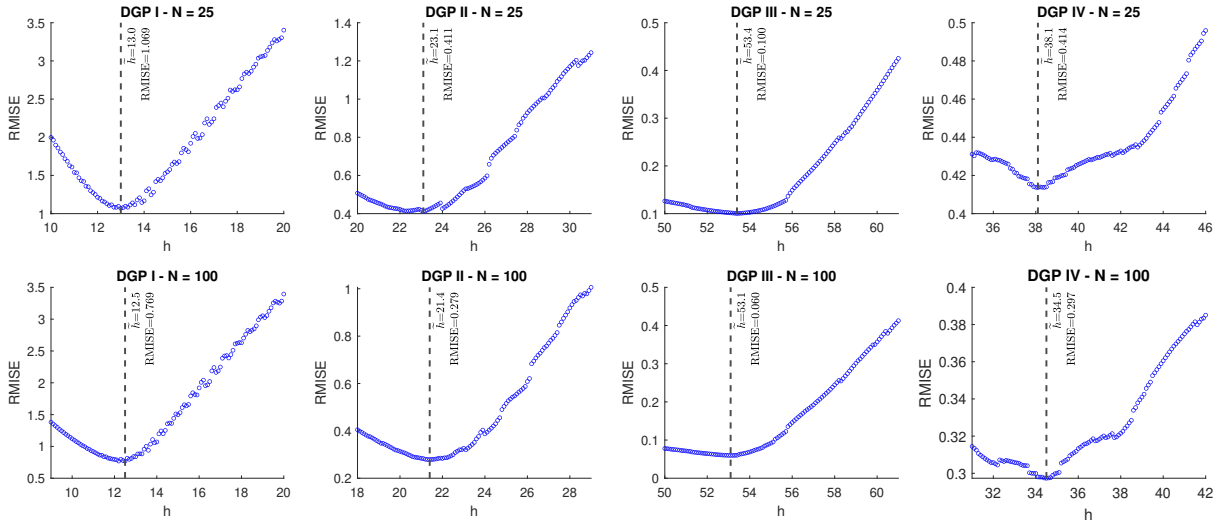

Figure D.1: Plot of the grid search of the optimal bandwidth parameter h of the PCA method for the four DGPs in our simulation. The vertical dashed line shows the optimal $\widetilde{h}$ which minimizes the simulated RMISE of the fitted PCA density against the true RND, and the minimized RMISE at $\widetilde{h}$ is reported.

on Eqs. (3.1) and (3.3) of Aït-Sahalia and Duarte (2003). This is a standard constrained quadratic programming problem which we implement using the Gurobi optimizer.

In the second step, we fit a local linear regression for the smoothed option prices at some point $x \in \mathcal{K}$ and extract the RND at x by differentiating the regression coefficient. In detail, we choose a Gaussian kernel $\phi(x) = (2\pi)^{-0.5} e^{-x^2/2}$ and let $\phi_h(x) = \phi(x/h)/h$. Choose some $x \in \mathcal{K}$ at which the RND is to be estimated, the weights applied to the observations on n-th strike grid is defined as $w_n = \phi_h(\widetilde{C}_n - x)$. Then according to Section 3.3 of Aït-Sahalia and Duarte (2003), the RND estimate at point x is given by

$$\widehat{f}_t(x) = e^{-r_f \tau} \frac{d\widehat{\beta}_1}{dx},$$

where $\widehat{\beta}_1$ is the slope coefficient of the local linear regression at point x which can be computed in closed form:

$$\widehat{\beta}_1 = \frac{S_{N,0} T_{N,1} - S_{N,1} T_{N,0}}{S_{N,2} S_{N,0} - S_{N,1}^2}, \tag{56}$$

where $S_{N,j} = \sum_{n=1}^{N} (\widetilde{C}_n - x)^j w_n$ and $T_{N,j} = \sum_{n=1}^{N} (\widetilde{C}_n - x)^j \widetilde{C}_n w_n$. The above expression can be found in Eq. (3.15) of Aït-Sahalia and Duarte (2003). In our simulation, we compute $\frac{d\widehat{\beta}_1}{dx}$ numerically using the central difference method.[6]

For each simulation scenario, we estimate $\widehat{f}_t(x)$ for x in the corresponding grid of N equidistant strikes and evaluate $\widehat{f}_t(x)$ with a general $x \in \mathcal{K}$, where $\mathcal{K}$ is specified in Table 4.1, by linear interpolation. The tuning parameter h is chosen based on a grid search in the range $[5, 45]$ which minimizes

[6] Although a closed-form derivative of $\widehat{\beta}_1$ can be derived based on Eq. (3.15) of Aït-Sahalia and Duarte (2003), we note that it requires the evaluation of two double sums with $O(N^2)$ terms, which is much slower than the central difference method that only requires several evaluation of a single sum of $O(N)$ terms.

the simulated RMISE defined in Eq. (26) of $\widehat{f}_t(x)$. Figure D.2 presents the grid search result and the choice of h for each simulation scenario under the basic i.i.d. option error design and a large error variance (4×), zooming in the local range of h that contains the optimal h^*. The figure clearly shows the bias-variance trade-off for the choice of h, and the optimal bandwidth h^* is chosen to minimize the simulated RMISE, which is infeasible in practice.

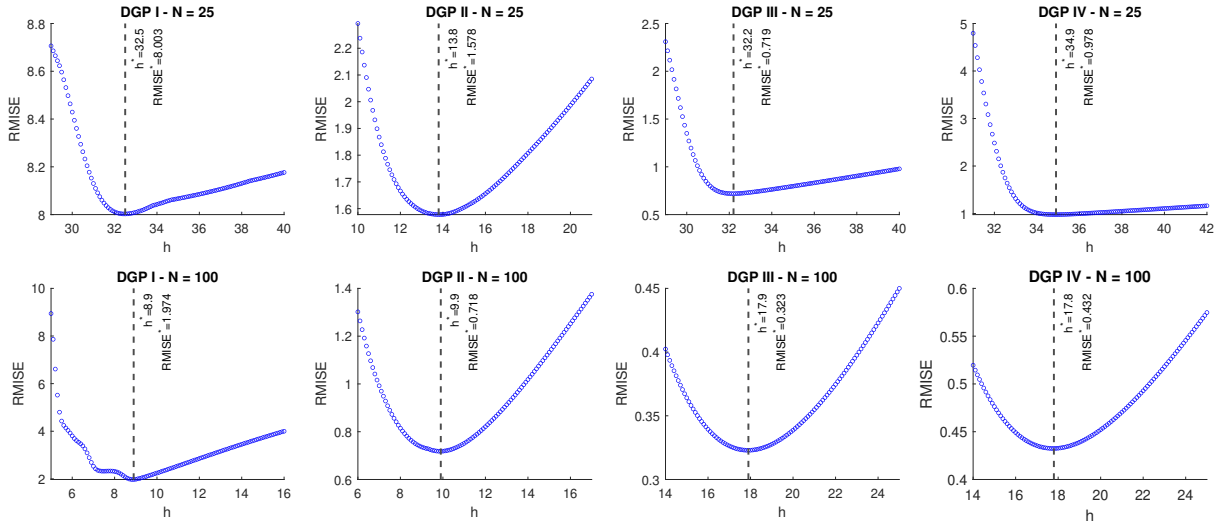

Figure D.2: Plot of the grid search of the optimal bandwidth parameter h^* of the LLS method for the four DGPs and two sample sizes under the basic i.i.d. option error design and a large error variance (4×) in our simulation. The vertical dashed line shows the optimal bandwidth h^* which minimizes the simulated RMISE of the RND estimates, based on 1,000 random draws of option prices. RMISE* denotes that minimized RMISE at h^*.

D.3 Sieve Estimator

The sieve estimator of Lu and Qu (2021) approximates the RND by a Jth order Hermite expansion. To implement their estimator, we firstly introduce two changes of variables:

$$s_T = \frac{\ln(F_T/F_t)}{\sigma_{ATM}\sqrt{\tau}}, \quad z_n = \frac{\ln(K_n/F_t)}{\sigma_{ATM}\sqrt{\tau}}, \tag{57}$$

where σ_{ATM} is the at-the-money Black-model implied volatility. Note that these are slightly different from Eqs. (3) and (4) in Lu and Qu (2021) as we work with forward prices. Consider the RND of s_T defined as $\breve{f}_t(x)$ which is related to the RND of F_T through $f_t(F_T) = \frac{1}{\sigma_{ATM}\sqrt{\tau}F_T}\breve{f}_t\big(\frac{\ln(F_T/F_t)}{\sigma_{ATM}\sqrt{\tau}}\big)$, Lu and Qu (2021) approximate $\breve{f}_t(x)$ by truncating its Gauss-Hermite expansion at the Jth term:

$$\breve{f}_t(x|\boldsymbol{\beta}) \approx \sum_{j=0}^{J} \beta_j h_j(x), \tag{58}$$

where $\boldsymbol{\beta} = \{\beta_j\}'_{j=0:J}$ are the $J+1$ free parameters to be estimated, and $h_j(x)$ is the jth order Hermite function (see Eqs. (8) and (9) of Lu and Qu (2021)). Consequently, the approximated RND of F_T is

denoted by $f_t(x|\boldsymbol{\beta})$.

For a particular choice of J and $\boldsymbol{\beta}$, the price of a call option at the strike grid K_n can be approximated by:

$$C_n \approx \sum_{j=0}^{J} \beta_j x_{n,j}, \tag{59}$$

where:

$$x_{n,j} := \int_{-\infty}^{\infty} e^{-r_f \tau} F_t\Big(e^{\sqrt{\tau}\sigma_{ATM}x} - e^{\sqrt{\tau}\sigma_{ATM}z_n}\Big) h_j(x)dx. \tag{60}$$

The quantities $x_{n,j}$ are independent of $\boldsymbol{\beta}$ which need to be evaluated prior to the estimation. To compute each $x_{n,j}$, we evaluate the integral numerically on the interval $[-m, m]$ with $m = 10 \vee \ln N$ following (I6) of Lu and Qu (2021). Denoting $\boldsymbol{x}_n = (x_{n,0}, \ldots, x_{n,J})'$, we estimate the parameters by solving the following constrained quadratic problem:

$$\begin{aligned} &\min_{\boldsymbol{\beta}} \sum_{n=1}^{N} (C_n - \boldsymbol{x}_n'\boldsymbol{\beta})^2 + \boldsymbol{\beta}'\boldsymbol{Q}_\alpha\boldsymbol{\beta}, \\ &s.t.\ \boldsymbol{\beta} \in \mathcal{H}_J := \{\boldsymbol{\beta} \in \mathbb{R}^{J+1} : \inf_{x\in\mathbb{R}} \sum_{j=0}^{J} \beta_j h_j(x) \geq \eta\}. \end{aligned} \tag{61}$$

The matrix $\boldsymbol{Q}_\alpha$ is a regularization matrix which can have different specifications, and α is the tuning parameter that controls the strength of the regularization. We set $\boldsymbol{Q}_\alpha$ according to Eqs. (22) and (23) of Lu and Qu (2021). We set $\eta = -10^{-3}$ in the constraint of $\boldsymbol{\beta}$ following (I1) of Lu and Qu (2021). Note that when implementing the constraint, we simply take 10,000 equidistant points on $\mathcal{K}$ and evaluate $\breve{f}_t(x)$ on these points, and $\boldsymbol{\beta} \in \mathcal{H}_J$ reduces to 10,000 linear constraints on $\boldsymbol{\beta}$ which is standard in quadratic programming problems. The problem is then solved using the Gurobi optimizer.

For the choice of J, we first consider $J^* = \lceil 2(N/\ln N)^{1/5} \rceil$ as recommended by Lu and Qu (2021) so that $J^* = 4$ when $N = 25$ and 100. For the choice of α, we firstly specify it as $\alpha = \xi\widehat{c}N^{1/3}$ according to (I4) of Lu and Qu (2021), where $\widehat{c}$ is the top left element in the $(J+1)$-by-$(J+1)$ matrix $N^{-1}\sum_{n=1}^{N} \boldsymbol{x}_n\boldsymbol{x}_n'$. For each simulation scenario and choice of J, we choose $\xi \in [0, 0.1]$ which minimizes the simulated RMISE of $f_t(x|\boldsymbol{\beta})$. We find that the optimal ξ's are zero for the majority of simulation scenarios, which suggests that the regularization is often not needed. Examples of grid search results for ξ with $J = J^* = 4$ for each simulation scenario under the baseline i.i.d. option error design and a large error variance ($4\times$) are presented in Figure D.3, which shows that five out of eight considered scenarios do not require any regularization. Nevertheless, given the large number of simulation scenarios considered in our study, several of which can benefit from some regularization, we still use optimally tuned ξ's via grid search for J^* to be consistent.

For each simulation scenario, we also consider a grid-search-based Hermite expansion order $\widetilde{J}$ over the range [1,30] that minimizes the simulated RMISE of the estimated RND. In our grid search, we

first fix the regularization parameter α (or effectively ξ) at 0 to tune J. Then for the chosen $J = \widetilde{J}$, we perform another grid search for $\xi \in [0, 0.1]$ to minimize RMISE. Although this sequential grid-search procedure is not as comprehensive as a full double grid search, it provides some optimal tuning for both J and ξ dimensions while avoiding the huge computational cost associated with the double grid search. Figures D.4 and D.5 respectively reveal the first and second-stage grid search results to find J and ξ for each simulation scenario under the basic i.i.d. option error design and a large error variance ($4\times$). While there is a clear bias-variance trade-off for the choice of J, the optimal value for ξ is 0 for most considered scenarios, suggesting that regularization is often not needed..

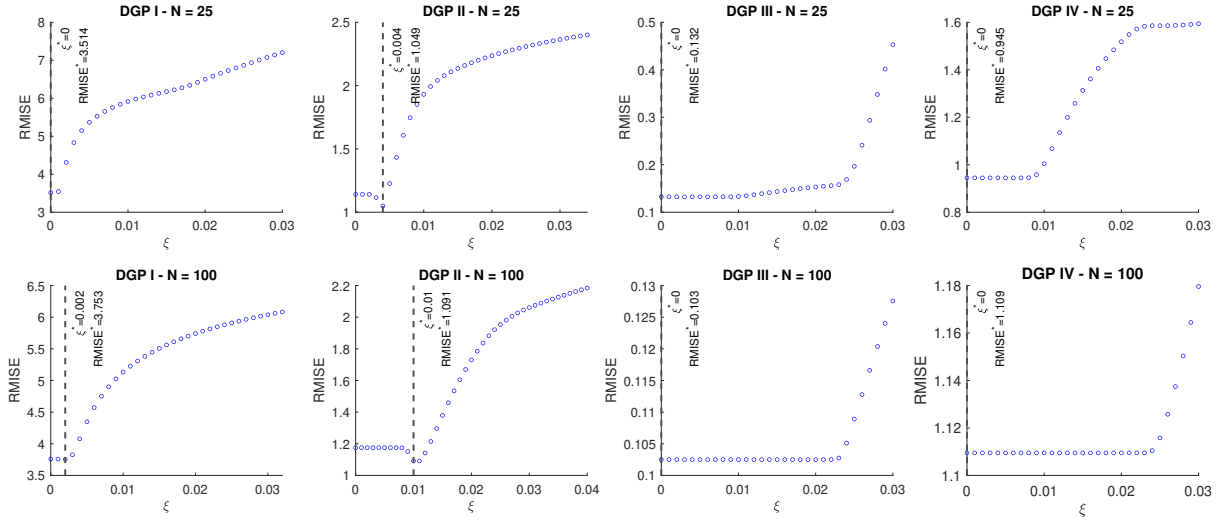

Figure D.3: Plot of the grid search of the optimal tuning parameter ξ^* of the Sieve method with $J = J^* = 4$ for the four DGPs and two sample sizes under the baseline i.i.d. option error design and a large error variance ($4\times$) in our simulation. The vertical dashed line shows the optimal tuning parameter ξ^* which minimizes the simulated RMISE of the RND estimates, based on 1,000 random draws of option prices. RMISE* denotes that minimized RMISE at ξ^*.

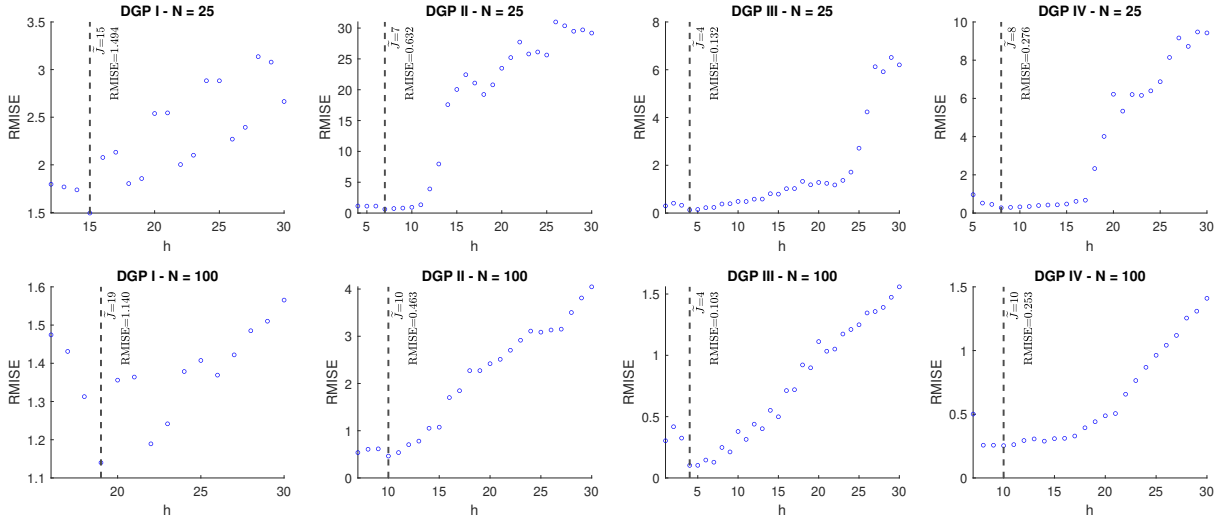

Figure D.4: Plot of the grid search of the optimal Hermite expansion order $\widetilde{J}$ of the Sieve method for the four DGPs and two sample sizes under the baseline i.i.d. option error design and a large error variance ($4\times$) in our simulation. The vertical dashed line shows the optimal $\widetilde{J}$ which minimizes the simulated RMISE of the RND estimates, based on 1,000 random draws of option prices. RMISE denotes that minimized RMISE at $\widetilde{J}$.

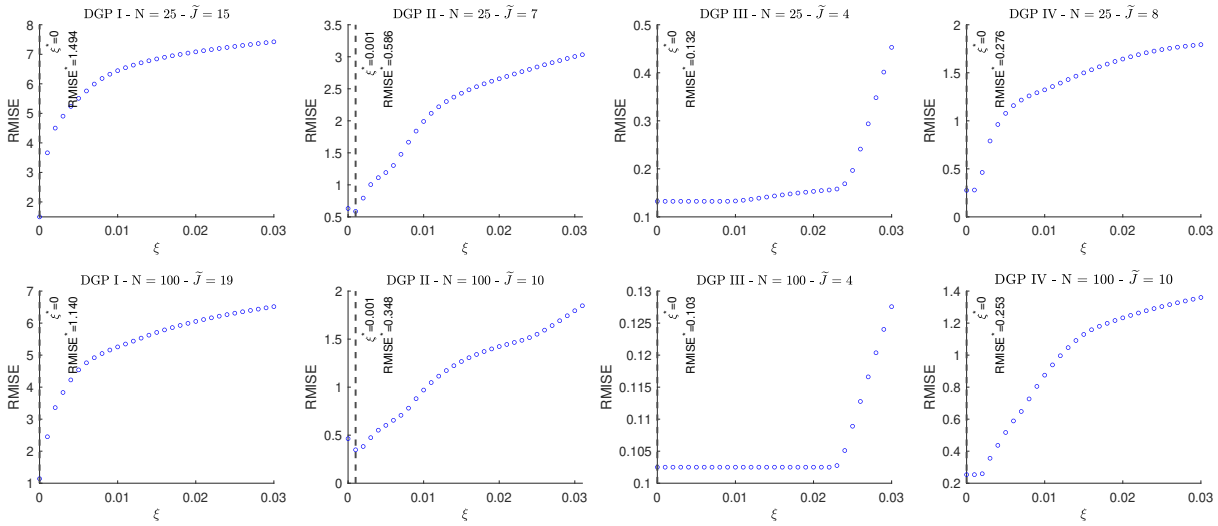

Figure D.5: Plot of the grid search of the optimal tuning parameter ξ^* of the Sieve method with first-stage RMISE-optimal Hermite expansion order $\widetilde{J}$ for the four DGPs and two sample sizes under the baseline i.i.d. option error design and a large error variance (4×) in our simulation. The vertical dashed line shows the optimal tuning parameter ξ^* which minimizes the simulated RMISE of the RND estimates, based on 1,000 random draws of option prices. RMISE* denotes that minimized RMISE at ξ^*.

D.4 Smoothing Implied Volatility

A popular method to estimate RND is by smoothing the implied volatility computed from the observed option prices (see e.g., Jackwerth (2000), Bliss and Panigirtzoglou (2004) and Kostakis et al. (2023)). To implement this method, we first convert the observed option price $\{O_n\}_{n=1:N}$ into the corresponding Black model implied volatilities $\{\sigma_n\}_{n=1:N}$. This is performed using the `blkimpv` function in MATLAB. We then fit a local cubic smoothing spline to $\{\sigma_n\}_{n=1:N}$ using `csaps` in MATLAB. Essentially, we find a local spline function $\widehat{\sigma}(x)$ that minimizes the following target function:

$$\lambda \sum_{n=1}^{N} \omega_n (\sigma_n - \sigma(K_n))^2 + (1-\lambda) \int_{\mathcal{K}} (\sigma(k)'')^2 dk, \tag{62}$$

where $\sigma(x)$ is a local cubic spline function defined on $\mathcal{K}$, and λ is a tuning parameter that penalizes the roughness of the function on $\mathcal{K}$, which is captured by the integrated squared second-order derivative of $\sigma(x)$ in the second term above. A smaller λ induces a more stringent penalty to the roughness of the cubic spline which results in a smoother fit, and vice versa.

From the estimated cubic spline function $\widehat{\sigma}(x)$, we can compute the fitted RND $\widehat{f}(x)$ as the numerical second derivative of the price function. For example, to compute the value of $\widehat{f}(K_0)$ at some $K_0 \in \mathcal{K}$, we choose a small value δ and compute the option prices $\widehat{O}(K_0)$, $\widehat{O}(K_0 - h)$ and $\widehat{O}(K_0 + h)$, where $\widehat{O}(K) \equiv O_{LN}(K; \widehat{\sigma}(K))$ is the Black model option price with the strike price K and the volatility parameter $\widehat{\sigma}(K)$ given by the cubic spline. From Assumption 1.2, we can approximate $\widehat{f}(K_0)$ as the

following central-different second-order numerical derivative:

$$\widehat{f}(K_0) \approx e^{r_f \tau} \frac{(\widehat{O}(K_0 - h) + \widehat{O}(K_0 + h) - 2\widehat{O}(K_0))}{h^2}. \tag{63}$$

In our simulation and empirical analysis, we compute $\widehat{f}(x)$ on an grid of $\mathcal{K}$ with 10,000 equidistant points to obtain an approximated RND estimate over the strike range.

Similar to the PCA, LLS and sieve methods, we optimally tune the penalty parameter λ based on a grid search in the range $[0, 0.1]$ which minimizes the RMISE (see Eq. (26)). Figure D.6 presents the grid search result and the choice of λ for each simulation scenario under the basic i.i.d. option error design and a large error variance ($4\times$), zooming in the local range of λ that contains the optimal $\widetilde{\lambda}$. The figure shows there is clear bias-variance trade-off for the choice of λ for the majority of scenarios, and the optimal tuning parameter $\widetilde{\lambda}$ is chosen to minimize the simulated RMISE, which is infeasible in practice.

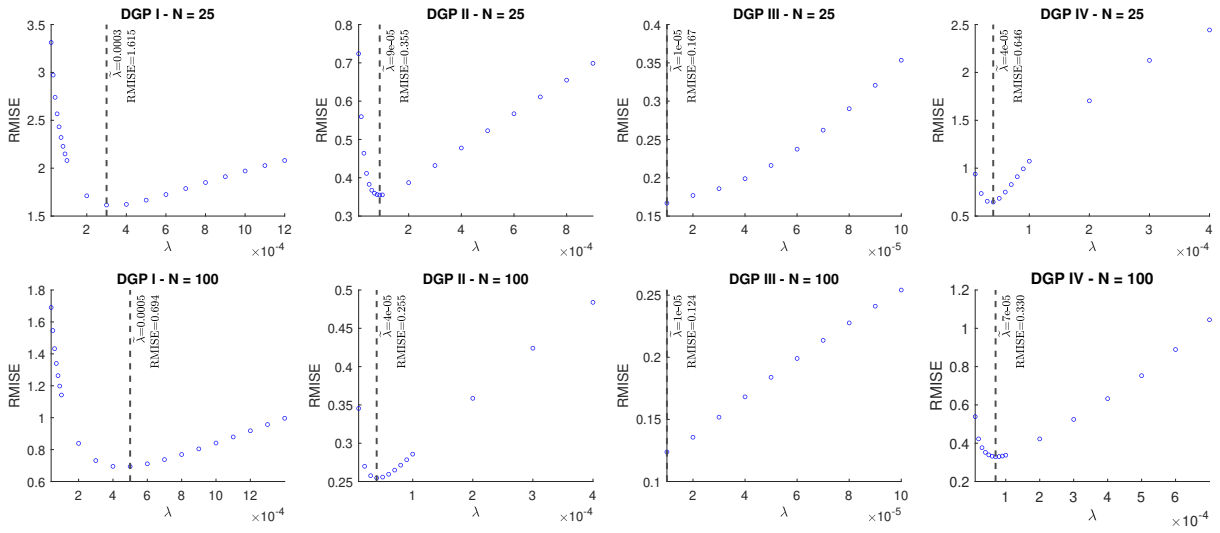

Figure D.6: Plot of the grid search of the optimal tuning parameter λ of the SIV method for the four DGPs and two sample sizes under the basic i.i.d. option error design and a large error variance ($4\times$) in our simulation. The vertical dashed line shows the optimal bandwidth $\widetilde{\lambda}$ which minimizes the simulated RMISE of the RND estimates, based on 1,000 random draws of option prices. RMISE denotes the minimized RMISE at $\widetilde{\lambda}$.

Besides implement the IV spline estimator with the infeasible optimally chosen tuning parameter $\widetilde{\lambda}$, we also consider a feasible version of the IV spline estimator where the tuning parameter λ is chosen by a 10-fold cross validation to minimize the objective function in Eq. (62), which can be implemented in practice.

D.5 Hermite Expansion Method

We implement the 4-parameter Hermite Expansion method of Jarrow and Rudd (1982) as a parametric benchmark RND estimator, which is also considered in Aït-Sahalia and Duarte (2003), Bondarenko

(2003), Lu and Qu (2021). The RND takes the following form:

$$f(x;\boldsymbol{\eta}) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-z^2/2}\Big(1 + \frac{\eta_3}{3!}(z^3 - 3z) + \frac{\eta_4}{4!}(z^4 - 6z^2 + 3)\Big), \quad z = \frac{\ln(x/\eta_1) + \eta_2^2/2}{\eta_2}, \tag{64}$$

where $\boldsymbol{\eta} = (\eta_i)_{i=1:4}^{\intercal}$ is the parameter vector. One should verify that $\int_0^\infty f(x;\boldsymbol{\eta}) = 1$ and $\int_0^\infty x f(x;\boldsymbol{\eta}) = \eta_1(1+\eta_3\eta_2^3/3!+\eta_4\eta_2^4/4!)$, where the second equality is used as a constraint in the optimization procedure to ensure that Assumption 1.(4) is satisfied. Note that $f(x;\boldsymbol{\eta})$ is not necessarily a well-defined density function as it can be negative for some $x \in \mathcal{K}$ under certain negative choices of η_3 and η_4.

This density implies the following pricing function for a call option written on a forward contract with strike price K:

$$\begin{aligned} O(K;\boldsymbol{\eta}) &= e^{-r_f\tau}(\eta_1\Phi(d_1) - K\Phi(d_2) + \eta_1\eta_3 F_3/3! + \eta_1\eta_4 F_4/4!), \\ F_3 &= \eta_2^3\Phi(d_1) - \phi(d_1)\eta_2(d_2 - \eta_2), \\ F_4 &= \eta_2^4\Phi(d_1) + \phi(d_1)(\eta_2(d_1^2 - 1) - 3\eta_2^2 d_2), \end{aligned} \tag{65}$$

where the quantities d_1 and d_2 are defined under Eq. (9) with $\eta_2 = \sigma\sqrt{\tau}$. The parameter vector $\boldsymbol{\eta}$ can be estimated by solving the following constrained nonlinear least square problem:

$$\begin{aligned} &\widehat{\boldsymbol{\eta}} = \underset{\boldsymbol{\eta}}{\operatorname{argmin}} \sum_{n=1}^{N} (O_n - O(K_n;\boldsymbol{\eta}))^2, \\ &\text{s.t.} \quad \eta_1, \eta_2 > 0, \eta_1(1 + \eta_3\eta_2^3/3! + \eta_4\eta_2^4/4!) = F_t. \end{aligned} \tag{66}$$

We solve the above problem using `fmincon` in MATLAB with the vector $(F_t, \sigma_{ATM}\sqrt{\tau}, 0, 0)^{\intercal}$ as the initial guess for $\widehat{\boldsymbol{\eta}}$.

E Data Cleaning Procedures

In this section, we document the data cleaning procedures used in our empirical analysis. For both examples, we firstly extract option quotes (best ask and bid) for both calls and puts with the same time-to-maturity (closest to one month in our examples) at the same point in time. For the first example, we take the closing quotes on 20-Jun-2013 from OptionMetrics, while for the second example, we extract the quotes prevailing at 13:30 and 14:30 ET on 18-Dec-2013 from the minute-by-minute LiveVol dataset. The risk-free rate r_f is obtained from OptionMetrics by a cubic spline interpolation of the zero coupon yield curve of the day, which is assumed to be constant throughout the trading day.

As a preliminary filter, we first remove all entries with a zero bid price to alleviate the problem caused by rounding for deep OTM options. The remainining entries are denoted by $\{C(K_n^C)\}_{n=1:N^C}$

and $\{P(K_n^P)\}_{n=1:N^P}$, where $C(K_n^C)$ and $P(K_n^P)$ are the midquotes of calls and puts at strike price K_n^C and K_n^P, respectively, and N^C and N^P denotes the number of observations for the calls and puts. Note that N^C and N^P are in general not the same, and for a given strike price K_C^i, its put counterpart on the same strike price is not necessarily quoted.

We proceed to match the calls with the puts by only keeping the strikes with both calls and puts quoted, denoted by $\{C(K_n^M), P(K_n^M)\}_{n=1:N^M}$, where the superscript stands for 'Matched'. The main purpose of this step is to compute the put-call parity implied forward price F_t, on which we elaborate as follows: For a given strike price K, the put-call parity implied forward price can be computed as:

$$F_t(K) = e^{\tau r_f}(C(K) - P(K)) + K. \tag{67}$$

As there are M matched quotes, we obtain $\{F_t(K_n^M)\}_{n=1:N^M}$ implied forward prices. We then set $F_t = \text{median}\{F_t(K^{(1)}), F_t(K^{(2)}), F_t(K^{(3)})\}$, where $\{K^{(n)}\}_{n=1:N^M}$ is a sorted version of $\{K_n^M\}_{n=1:N^M}$ based on:

$$|C(K^{(1)}) - P(K^{(1)})| \leq |C(K^{(2)}) - P(K^{(2)})| \leq \cdots \leq |C(K^{(N^M)}) - P(K^{(N^M)})|.$$

Intuitively, we compute the implied F_t from put-call parities around at-the-money options. This allows us to align the OTM call options with the ITM call options implied by the OTM put options and construct a smooth cross-section of option prices as a function of the strike.

As the last step, we collect the OTM calls and puts from the unmatched datasets $\{C(K_n^C)\}_{n=1:N^C}$ and $\{P(K_n^P)\}_{n=1:N^P}$, then convert the OTM puts to ITM calls using Eq. (67) to obtain the cross-section of call options $\{C_n, K_n\}_{n=1:N}$ used in our empirical analysis. Intuitively, for $K_n \geq F_t$, $C(K_n)$ coincides with $C(K_n^C)$, while for $K_n \leq F_t$, $P(K_n)$ coincides with $\{P(K_n^P)\}$.

F Additional results

F.1 Simulation results for different values of c

Table F.1 presents the impact of different choices of c for DGP I (RND with a small volatility and a long left tail) under difference sizes of error variances. The table clearly shows that, both the RISB and the RIV explode as M increases when the variance constraint is not imposed with $c = 0$. In this case, the RND estimates contain spurious spikes that inflate both the RISB and the RIV of the RND estimates. Unsurprisingly, the issue becomes more severe as the error variance increases. The introduction of the variance constraint for $c \in [0.05, 0.15]$ evidently stabilizes the RND estimates for M up to 5 under both error variance sizes. The performances of the LWM-based RND estimates for different choices of c in this range are qualitatively unchanged, which shows the robustness of the LWM approach to the choices of c. These findings are also consistent across different DGPs and observation error specifications. To conserve space, we do not report these results here, which are available upon request.

	$N = 25$					$N = 100$				
Estimator	MTVD	RMISE	RISB	RIV	M_1	MTVD	RMISE	RISB	RIV	M_1
Panel 1: Large Error Variance (4×)										
Panel 1A: $c = 0$										
2-LWM (EW)	0.181	0.916	0.751	0.525	0.01	0.157	0.786	0.744	0.252	0.00
3-LWM (EW)	0.900	2.024	0.413	1.981	0.71	192.601	968.900	30.788	968.410	0.28
4-LWM (EW)	1011.969	4728.096	228.081	4722.592	1.33	8027.451	20187.547	1023.667	20161.576	0.83
5-LWM (EW)	310.101	1516.495	69.132	1514.918	1.95	16872.675	28353.891	1262.107	28325.787	1.49
Panel 1B: $c = 0.05$										
2-LWM (EW)	0.180	0.912	0.747	0.524	0.00	0.157	0.786	0.744	0.252	0.00
3-LWM (EW)	0.565	1.728	0.410	1.679	0.48	0.324	0.942	0.502	0.797	0.19
4-LWM (EW)	0.800	3.009	0.351	2.988	1.08	0.258	0.975	0.506	0.834	0.65
5-LWM (EW)	1.019	3.533	0.272	3.522	1.72	0.576	1.711	0.239	1.694	1.35
Panel 1C: $c = 0.1$										
2-LWM (EW)	0.180	0.912	0.747	0.524	0.00	0.157	0.786	0.744	0.252	0.00
3-LWM (EW)	0.393	1.475	0.394	1.422	0.50	0.178	0.717	0.484	0.529	0.11
4-LWM (EW)	0.654	2.478	0.289	2.461	1.10	0.255	0.962	0.455	0.847	0.59
5-LWM (EW)	0.865	2.927	0.254	2.916	1.67	0.486	1.531	0.265	1.508	1.26
Panel 1D: $c = 0.15$										
2-LWM (EW)	0.180	0.912	0.747	0.524	0.00	0.157	0.786	0.744	0.252	0.00
3-LWM (EW)	0.367	1.420	0.440	1.350	0.74	0.195	0.767	0.487	0.592	0.54
4-LWM (EW)	0.568	2.226	0.268	2.210	1.26	0.256	1.024	0.279	0.985	1.04
5-LWM (EW)	0.873	3.288	0.291	3.275	1.82	0.408	1.452	0.194	1.439	1.62
Panel 2: Medium Error Variance (2×)										
Panel 2A: $c = 0$										
2-LWM (EW)	0.180	0.898	0.819	0.369	0.00	0.164	0.802	0.781	0.184	0.00
3-LWM (EW)	0.612	1.378	0.515	1.278	0.51	2.659	16.856	0.749	16.840	0.09
4-LWM (EW)	646.286	4245.303	147.006	4242.757	1.13	1442.103	3380.572	175.944	3375.990	0.65
5-LWM (EW)	321.305	1271.044	72.558	1268.972	1.72	8434.418	23009.729	789.800	22996.170	1.28
Panel 2B: $c = 0.05$										
2-LWM (EW)	0.180	0.898	0.819	0.369	0.00	0.164	0.802	0.781	0.184	0.00
3-LWM (EW)	0.356	1.106	0.520	0.976	0.36	0.190	0.712	0.529	0.476	0.08
4-LWM (EW)	0.596	2.404	0.373	2.375	0.93	0.194	0.784	0.470	0.627	0.58
5-LWM (EW)	0.716	2.754	0.234	2.744	1.57	0.413	1.353	0.172	1.342	1.31
Panel 2C: $c = 0.1$										
2-LWM (EW)	0.180	0.898	0.819	0.369	0.00	0.164	0.802	0.781	0.184	0.00
3-LWM (EW)	0.264	0.996	0.510	0.856	0.36	0.144	0.623	0.525	0.334	0.05
4-LWM (EW)	0.488	1.967	0.326	1.940	0.98	0.207	0.827	0.399	0.725	0.61
5-LWM (EW)	0.608	2.391	0.218	2.381	1.57	0.331	1.177	0.185	1.163	1.29
Panel 2D: $c = 0.15$										
2-LWM (EW)	0.180	0.898	0.819	0.369	0.00	0.164	0.802	0.781	0.184	0.00
3-LWM (EW)	0.268	0.993	0.559	0.820	0.66	0.140	0.647	0.522	0.383	0.44
4-LWM (EW)	0.419	1.806	0.235	1.791	1.25	0.194	0.827	0.243	0.791	0.99
5-LWM (EW)	0.555	2.352	0.254	2.338	1.81	0.271	1.090	0.164	1.078	1.61
Panel 3: Small Error Variance (1×)										
Panel 3A: $c = 0$										
2-LWM (EW)	0.182	0.886	0.845	0.266	0.00	0.167	0.807	0.797	0.126	0.00
3-LWM (EW)	0.358	1.051	0.556	0.891	0.31	0.135	0.590	0.533	0.252	0.07
4-LWM (EW)	92.239	655.097	20.753	654.768	0.91	359.081	1461.803	61.758	1460.498	0.59
5-LWM (EW)	218.713	875.504	49.411	874.109	1.49	2171.104	5764.977	276.838	5758.326	1.15
Panel 3B: $c = 0.05$										
2-LWM (EW)	0.182	0.886	0.845	0.266	0.00	0.167	0.807	0.797	0.126	0.00
3-LWM (EW)	0.209	0.812	0.566	0.582	0.23	0.131	0.579	0.523	0.249	0.07
4-LWM (EW)	0.462	2.083	0.311	2.060	0.87	0.167	0.702	0.362	0.602	0.54
5-LWM (EW)	0.550	2.400	0.252	2.387	1.51	0.247	0.941	0.121	0.933	1.22
Panel 3C: $c = 0.1$										
2-LWM (EW)	0.182	0.886	0.845	0.266	0.00	0.167	0.807	0.797	0.126	0.00
3-LWM (EW)	0.184	0.774	0.565	0.529	0.21	0.134	0.588	0.537	0.239	0.02
4-LWM (EW)	0.409	1.869	0.289	1.847	0.89	0.174	0.730	0.262	0.681	0.66
5-LWM (EW)	0.469	2.093	0.234	2.080	1.50	0.226	0.894	0.118	0.886	1.37
Panel 3D: $c = 0.15$										
2-LWM (EW)	0.182	0.886	0.845	0.266	0.00	0.167	0.807	0.797	0.126	0.00
3-LWM (EW)	0.189	0.809	0.597	0.547	0.50	0.115	0.580	0.535	0.223	0.35
4-LWM (EW)	0.339	1.576	0.284	1.550	1.17	0.157	0.714	0.200	0.686	0.93
5-LWM (EW)	0.451	2.125	0.315	2.102	1.78	0.176	0.777	0.129	0.767	1.60

Table F.1: Simulation results of the RND estimates based on the M-LWM model with equal weighting (EW) under **DGP I** with a **large error variance** (4× the estimated error variance of SPX options), **medium error variance** (2×) and **small error variance** (1×), for different values of c, which is the lower bound for the ratio of the smallest component standard deviation to the largest one in an M-LWM mixture, defined in Eq. (24). M_1 reports the average number of log normal densities chosen in each M-LWM model. For each simulation scenario, the statistics in the table are computed based on 1,000 simulated paths.

F.2 Additional Simulation Results for the Baseline Option Error Specification

	Small error variance (1×)									
	$N = 25$					$N = 100$				
Estimator	MTVD	RMISE	RISB	RIV	M_1	MTVD	RMISE	RISB	RIV	M_1
Panel 1: DGP I										
2-LWM (EW)	0.182	0.886	0.845	0.266	0.00	0.167	0.807	0.797	0.126	0.00
3-LWM (EW)	0.184	**0.774**	0.565	0.529	0.21	0.134	0.588	0.537	0.239	0.02
4-LWM (EW)	0.409	1.869	0.289	1.847	0.89	0.174	0.730	0.262	0.681	0.66
5-LWM (EW)	0.469	2.093	**0.234**	2.080	1.50	0.226	0.894	**0.118**	0.886	1.37
PCA(h^*)	1.560	7.522	7.514	0.359	23.00	1.577	7.606	7.606	0.071	23.00
Sieve(J^*)	0.709	3.510	3.509	**0.057**	4.00	0.769	3.755	3.755	**0.034**	4.00
IV spline(λ^*)	0.415	2.388	1.520	1.843	-	0.328	3.616	0.987	3.479	-
HE	0.861	5.001	5.000	0.117	-	0.861	5.000	5.000	0.052	-
PCA($\widetilde{h}$)	**0.144**	0.811	0.591	0.555	79.42	**0.099**	**0.569**	0.328	0.465	87.17
Sieve($\widetilde{J}$)	0.187	1.149	0.805	0.819	15.00	0.131	0.843	0.592	0.599	19.00
IV spline($\widetilde{\lambda}$)	0.221	1.283	0.962	0.849	-	0.155	0.727	0.656	0.313	-
LLS	1.308	6.451	5.988	2.401	-	0.347	1.555	1.442	0.583	-
Panel 2: DGP II										
2-LWM (EW)	**0.031**	**0.186**	0.094	0.160	0.05	**0.022**	**0.128**	0.100	0.080	0.01
3-LWM (EW)	0.194	0.542	**0.089**	0.535	0.58	0.095	0.285	0.107	0.265	0.44
4-LWM (EW)	0.373	0.999	0.089	0.995	1.24	0.179	0.480	0.086	0.472	1.11
5-LWM (EW)	0.645	1.720	0.120	1.715	1.88	0.297	0.745	0.080	0.741	1.78
PCA(h^*)	0.087	0.589	0.535	0.246	23.00	0.097	0.644	0.637	0.092	23.00
Sieve(J^*)	0.160	1.008	0.977	0.247	4.00	0.172	1.081	1.071	0.143	4.00
IV spline(λ^*)	0.091	0.644	0.101	0.636	-	0.065	0.581	**0.067**	0.578	-
HE	0.193	1.516	1.515	**0.075**	-	0.191	1.514	1.513	**0.036**	-
PCA($\widetilde{h}$)	0.038	0.269	0.140	0.230	28.35	0.025	0.172	0.093	0.145	29.57
Sieve($\widetilde{J}$)	0.073	0.487	0.458	0.167	7.00	0.032	0.241	0.187	0.153	10.00
IV spline($\widetilde{\lambda}$)	0.035	0.226	0.095	0.205	-	0.023	0.144	0.108	0.096	-
LLS	0.230	1.070	0.673	0.832	-	0.092	0.512	0.376	0.347	-
Panel 3: DGP III										
2-LWM (EW)	0.008	0.071	0.058	0.041	1.98	0.007	0.061	0.058	0.020	2.00
3-LWM (EW)	0.022	0.156	0.024	0.154	2.66	0.008	0.071	0.019	0.068	2.74
4-LWM (EW)	0.043	0.275	**0.018**	0.274	3.31	0.013	0.098	0.012	0.097	3.34
5-LWM (EW)	0.080	0.501	0.026	0.500	3.97	0.026	0.173	**0.009**	0.172	3.96
PCA(h^*)	0.019	0.200	0.166	0.112	23.00	0.023	0.230	0.224	0.052	23.00
Sieve(J^*)	0.009	0.098	0.083	0.051	4.00	0.009	0.088	0.085	0.025	4.00
IV spline(λ^*)	0.039	0.617	0.070	0.613	-	0.036	0.521	0.056	0.518	-
HE	0.010	0.092	0.087	**0.030**	-	0.009	0.089	0.087	**0.015**	-
PCA($\widetilde{h}$)	**0.006**	**0.059**	0.025	0.054	26.20	**0.004**	**0.037**	0.020	0.030	26.53
Sieve($\widetilde{J}$)	0.009	0.098	0.083	0.051	4.00	0.007	0.068	0.049	0.047	7.00
IV spline($\widetilde{\lambda}$)	0.011	0.106	0.056	0.090	-	0.008	0.070	0.048	0.052	-
LLS	0.066	0.545	0.515	0.180	-	0.031	0.246	0.181	0.166	-
Panel 4: DGP IV										
2-LWM (EW)	0.064	0.676	0.542	0.403	1.00	0.064	0.689	0.552	0.411	1.00
3-LWM (EW)	**0.029**	**0.228**	0.167	0.156	1.21	0.026	0.201	0.155	0.127	1.11
4-LWM (EW)	0.039	0.281	0.101	0.262	1.41	**0.020**	**0.145**	0.108	0.098	1.26
5-LWM (EW)	0.076	0.440	**0.049**	0.438	1.85	0.034	0.181	**0.040**	0.177	1.68
PCA(h^*)	0.117	1.173	1.162	0.155	23.00	0.120	1.214	1.213	0.052	23.00
Sieve(J^*)	0.118	0.946	0.945	0.020	4.00	0.131	1.111	1.111	0.008	4.00
IV spline(λ^*)	0.117	2.309	0.495	2.255	-	0.075	0.754	0.302	0.691	-
HE	0.184	1.896	1.896	**0.012**	-	0.184	1.896	1.896	**0.006**	-
PCA($\widetilde{h}$)	0.034	0.303	0.214	0.215	46.92	0.023	0.215	0.143	0.160	52.44
Sieve($\widetilde{J}$)	0.033	0.243	0.231	0.074	8.00	0.026	0.218	0.157	0.152	16.00
IV spline($\widetilde{\lambda}$)	0.061	0.417	0.391	0.145	-	0.033	0.242	0.200	0.137	-
LLS	0.076	0.734	0.708	0.195	-	0.038	0.330	0.248	0.217	-

Table F.2: Simulation results of the RND estimates based on the M-LWM model with equal weighting (EW) and five competing estimators under a **small error variance** (same as (1×) the estimated error variance of SPX options). The specification and choice of tuning parameters of the competing estimators are detailed in Online Appendix D. MTVD, RMISE, RISB and RIV respectively stands for mean total variation distance, root mean integrated squared error, root integrated squared bias, root integrated variance, and are defined in Eq. (25)-(27). M_1 reports (i) for LWM method, the average number of log normal densities chosen in each M-LWM model, (ii) for PCA method, the average number of normal mixtures, and (iii) for Sieve method, the average order of Hermite expansion (i.e. J^* or $\widetilde{J}$). For each simulation scenario, the statistics in the table are computed based on 1,000 simulated paths. The best and the top three statistics among all estimators are highlighted in **bold** and underlined, respectively.

	Medium error variance (2×)									
	$N = 25$					$N = 100$				
Estimator	MTVD	RMISE	RISB	RIV	M_1	MTVD	RMISE	RISB	RIV	M_1
Panel 1: DGP I										
2-LWM (EW)	0.180	**0.898**	0.819	0.369	0.00	0.164	0.802	0.781	0.184	0.00
3-LWM (EW)	0.264	0.996	0.510	0.856	0.36	0.144	**0.623**	0.525	0.334	0.05
4-LWM (EW)	0.488	1.967	0.326	1.940	0.98	0.207	0.827	0.399	0.725	0.61
5-LWM (EW)	0.608	2.391	**0.218**	2.381	1.57	0.331	1.177	**0.185**	1.163	1.29
PCA(h^*)	1.559	7.520	7.511	0.367	23.00	1.577	7.605	7.604	0.075	23.00
Sieve(J^*)	0.710	3.512	3.511	**0.082**	4.00	0.769	3.754	3.753	**0.059**	4.00
IV spline(λ^*)	0.433	2.901	1.487	2.491	-	0.283	1.494	1.037	1.076	-
HE	0.860	5.010	5.008	0.148	-	0.861	5.009	5.008	0.071	-
PCA($\widetilde{h}$)	**0.164**	0.930	0.605	0.707	79.35	**0.113**	0.660	0.341	0.565	87.13
Sieve($\widetilde{J}$)	0.220	1.317	0.813	1.037	15.00	0.155	0.972	0.597	0.767	19.00
IV spline($\widetilde{\lambda}$)	0.289	1.745	1.271	1.197	-	0.122	0.630	0.471	0.419	-
LLS	1.245	7.243	5.467	4.750	-	0.399	1.810	1.694	0.636	-
Panel 2: DGP II										
2-LWM (EW)	**0.040**	**0.253**	**0.073**	0.242	0.12	**0.025**	**0.144**	**0.097**	0.106	0.01
3-LWM (EW)	0.253	0.748	0.089	0.742	0.61	0.116	0.353	0.121	0.332	0.42
4-LWM (EW)	0.533	1.497	0.124	1.492	1.29	0.228	0.610	0.117	0.599	1.08
5-LWM (EW)	0.904	2.370	0.186	2.363	1.92	0.449	1.056	0.122	1.049	1.75
PCA(h^*)	0.088	0.592	0.525	0.273	23.00	0.097	0.645	0.636	0.107	23.00
Sieve(J^*)	0.163	1.026	0.985	0.287	4.00	0.174	1.084	1.073	0.155	4.00
IV spline(λ^*)	0.114	0.798	0.103	0.792	-	0.087	0.825	0.097	0.819	-
HE	0.194	1.517	1.514	**0.101**	-	0.191	1.514	1.513	**0.049**	-
PCA($\widetilde{h}$)	0.049	0.344	0.204	0.277	27.23	0.031	0.218	0.111	0.187	29.47
Sieve($\widetilde{J}$)	0.077	0.530	0.465	0.253	7.00	0.039	0.290	0.193	0.217	10.00
IV spline($\widetilde{\lambda}$)	0.044	0.289	0.175	0.230	-	0.030	0.192	0.143	0.128	-
LLS	0.277	1.311	0.843	1.004	-	0.108	0.607	0.442	0.417	-
Panel 3: DGP III										
2-LWM (EW)	0.009	0.084	0.060	0.059	1.95	0.007	0.066	0.060	0.027	2.00
3-LWM (EW)	0.028	0.205	0.034	0.202	2.67	0.010	0.088	0.027	0.084	2.78
4-LWM (EW)	0.064	0.407	**0.029**	0.405	3.33	0.019	0.140	0.017	0.139	3.43
5-LWM (EW)	0.127	0.753	0.037	0.752	4.00	0.040	0.257	**0.015**	0.256	4.07
PCA(h^*)	0.019	0.201	0.164	0.116	23.00	0.023	0.228	0.222	0.052	23.00
Sieve(J^*)	0.011	0.110	0.083	0.071	4.00	0.009	0.094	0.087	0.037	4.00
IV spline(λ^*)	0.077	3.239	0.194	3.233	-	0.043	0.720	0.069	0.717	-
HE	0.010	0.098	0.088	**0.043**	-	0.010	0.090	0.088	**0.020**	-
PCA($\widetilde{h}$)	**0.007**	**0.077**	0.032	0.070	26.20	**0.004**	**0.044**	0.023	0.038	26.51
Sieve($\widetilde{J}$)	0.011	0.110	0.083	0.071	4.00	0.008	0.089	0.071	0.054	5.00
IV spline($\widetilde{\lambda}$)	0.014	0.133	0.077	0.109	-	0.010	0.092	0.066	0.065	-
LLS	0.086	0.732	0.695	0.230	-	0.035	0.281	0.208	0.188	-
Panel 4: DGP IV										
2-LWM (EW)	0.066	0.696	0.552	0.423	1.00	0.064	0.686	0.552	0.407	1.00
3-LWM (EW)	**0.032**	**0.250**	0.166	0.187	1.31	0.027	0.213	0.160	0.141	1.16
4-LWM (EW)	0.053	0.368	0.108	0.352	1.53	**0.023**	**0.179**	0.117	0.135	1.28
5-LWM (EW)	0.120	0.684	**0.065**	0.681	2.00	0.051	0.273	**0.042**	0.270	1.73
PCA(h^*)	0.117	1.165	1.154	0.163	23.00	0.120	1.213	1.211	0.055	23.00
Sieve(J^*)	0.118	0.945	0.945	0.027	4.00	0.131	1.111	1.111	0.012	4.00
IV spline(λ^*)	0.117	1.956	0.512	1.888	-	0.137	2.099	0.332	2.072	-
HE	0.184	1.896	1.896	**0.018**	-	0.184	1.896	1.896	**0.009**	-
PCA($\widetilde{h}$)	0.038	0.355	0.237	0.265	46.29	0.027	0.259	0.163	0.201	51.72
Sieve($\widetilde{J}$)	0.034	0.256	0.232	0.106	8.00	0.031	0.241	0.223	0.092	10.00
IV spline($\widetilde{\lambda}$)	0.063	0.452	0.390	0.228	-	0.035	0.281	0.205	0.192	-
LLS	0.086	0.902	0.858	0.278	-	0.043	0.378	0.285	0.248	-

Table F.3: Simulation results of the RND estimates based on the M-LWM model with equal weighting (EW) and five competing estimators under a **medium error variance** (2× the estimated error variance of SPX options). The specification and choice of tuning parameters of the competing estimators are detailed in Online Appendix D. MTVD, RMISE, RISB and RIV respectively stands for mean total variation distance, root mean integrated squared error, root integrated squared bias, root integrated variance, and are defined in Eq. (25)-(27). M_1 reports (i) for LWM method, the average number of log normal densities chosen in each M-LWM model, (ii) for PCA method, the average number of normal mixtures, and (iii) for Sieve method, the average order of Hermite expansion (i.e. J^* or $\widetilde{J}$). For each simulation scenario, the statistics in the table are computed based on 1,000 simulated paths. The best and the top three statistics among all estimators are highlighted in **bold** and underlined, respectively.

F.3 Alternative Option Error Specifications

In this section, we consider three alternative error designs that allow for heteroskedastic and/or serially correlated option errors. In the first 'Alternative error design I', option errors are added to the Black model implied volatility of the theoretical option prices, as inspired by Christoffersen et al. (2012). In particular, for each strike K_n, we transform the theoretical call option price C_n^* to the Black model implied volatility IV_n^*, and add errors to the latter equal to $u_{\text{IV},n} = \gamma_{\text{IV}} \times Z_{\text{IV},n}$, where $Z_{\text{IV},n} \sim i.i.d.\, U(-1,1)$. The observed call price C_n is then computed by transforming back the observed IV_n (=max(IV_n^*+$u_{\text{IV},n}$,0)) to option prices using the Black model. Due to the nonlinear tranformation of the Black model, the option errors $u_n = C_n - C_n^*$ are heteroskedastic, despite being independent.

In our 'Alternative error design II', we allow for strong autogressive serial correlations in the option errors u_n, while still assuming that they are homoskedastic. In particular, u_n has the following structure:

$$u_n = \gamma_u \times Z_n, \quad Z_n = 2\Phi(\eta_n) - 1, \quad \eta_n = 0.8\eta_{n-1} + e_n, \quad e_n \sim i.i.d.\mathcal{N}(0, 0.36),$$

in which $\Phi(x)$ is the cumulative density function of a standard normal distribution, and consequently $\{Z_n\}_{n=1:N}$ is an AR(1) dependent time series which is unconditionally and uniformly distributed on $[-1,1]$. The first order autocorrelation of Z_n, and also u_n, is about 0.8, allowing for potentially much stronger serial correlation in option errors than documented empirically in Andersen et al. (2021).

Our final 'Alternative error design III' modifies the design I above to allow for strong autoregressive but homoskedastic errors in the IV domain, which translate into strong autoregressive and heteroskedastic errors in the price domain. In particular, we consider the following structure for $u_{\text{IV},n}$, while keeping other transformations unchanged.

$$u_{\text{IV},n} = \gamma_{\text{IV}} \times Z_{\text{IV},n}, \quad Z_{\text{IV},n} = 2\Phi(\eta_{\text{IV},n}) - 1, \quad \eta_{\text{IV},n} = 0.8\eta_{\text{IV},n-1} + e_{\text{IV},n}, \quad e_{\text{IV},n} \sim i.i.d.\mathcal{N}(0, 0.36).$$

Similar to the basic i.i.d. option error specification discussed earlier, the proportionality factors of the alternative error designs (i.e. γ_{IV} in designs I and III, and γ_u in design II) are chosen such that the variance of u_n is approximately the same as, twice and four times the estimated SPX options' error variance in our empirical application. Table F.4 tabulates the values of these factors used in our simulation, and we note that while γ_u is the same for all four DGPs for each error variance size, γ_{IV} is different for different DGPs.

Simulation results for alternative option error designs are presented in Tables F.5 to F.10. For the LWM method, in addition to the equal weighting scheme, we also consider an alternative Vega

Error factor	Variance factor	DGP I	DGP II	DGP III	DGP IV
		Low Volatility	Mid Volatility	High Volatility	Bimodal
γ_{IV}	1×	0.0099	0.0071	0.0046	0.0048
	2×	0.0142	0.0100	0.0066	0.0067
	4×	0.0198	0.0142	0.0093	0.0094
γ_u	1×		0.1212		
	2×		0.1715		
	4×		0.2425		

Table F.4: Error proportionality factors (γ_{IV} in alternative error designs I and III, and γ_u in design II) under small and large error variance cases (2× and 4× the estimated SPX options' error variance, respectively) for the four DGPs in our simulation.

weighting (VW) scheme suggested by Christoffersen et al. (2018) that sets $\omega_n = 1/v(K_n)^2$, where $v(K_n)$ is the true option Vega at strike K_n, to examine if it improves the performance of the LWM estimator under heteroskedastic errors.

Similar to previous results, the LWM method works well in recovering different RNDs with just a small number of mixtures (i.e. 2 or 3) even when option errors are heteroskedastic (designs I & III) and/or strongly serially correlated (designs II & III). In the majority of cases, the best LWM-based RND estimator are amongst the top three candidates in terms of MTVD and RMISE, having comparable performance to infeasibly optimally tuned SNP estimators, namely PCA($\widetilde{h}$), Sieve($\widetilde{J}$), and IV spline($\widetilde{\lambda}$). However, the best LWM estimator significantly outperforms the SNP rivals that employ feasible adaptive/recommended tuning parameters.

While the Vega-weighted LWM-based RND estimates are consistently outperformed by the equally-weighted ones under autocorrelated but homoskedastic option errors (design II), the former, surprisingly, do not perform as well as the latter even when option errors are heteroskedastic (designs I and III), except under DGP II and when 4 or 5 mixtures are considered, potentially due to model misspecification. When the LWM model is misspecified, the inverse Vega weights, which are low near-the-money, heavily distort the model fit of the near-the-money option prices, resulting in more biased RND estimates than those based on a simple equal weighting scheme. This suggests that the equal weight scheme provides a more robust choice for RND estimation in practice, which we recommend in empirical applications.

	$N = 100$ & Small error variance ($1\times$)														
	Alternative Error Design I					Alternative Error Design II					Alternative Error Design III				
Estimator	MTVD	RMISE	RISB	RIV	M_1	MTVD	RMISE	RISB	RIV	M_1	MTVD	RMISE	RISB	RIV	M_1
Panel 1: DGP I															
2-LWM (EW)	0.171	0.840	0.800	0.256	0.00	0.167	0.823	0.789	0.236	0.00	0.173	0.864	0.758	0.415	0.00
2-LWM (VW)	0.612	2.832	2.829	0.122	0.00	0.749	4.117	3.216	2.571	0.25	0.609	2.820	2.814	0.191	0.00
3-LWM (EW)	0.152	0.689	0.506	0.467	0.03	0.135	0.616	0.521	0.329	0.16	0.166	0.784	0.491	0.611	0.07
3-LWM (VW)	0.441	1.990	1.985	0.138	0.00	0.568	2.987	2.176	2.046	0.85	0.438	1.988	1.972	0.246	0.00
4-LWM (EW)	0.523	1.888	0.318	1.862	0.87	0.153	0.683	0.244	0.638	0.73	0.358	1.401	0.212	1.385	0.86
4-LWM (VW)	0.250	1.156	0.996	0.587	0.05	0.491	2.602	1.750	1.926	1.50	0.267	1.265	0.847	0.940	0.17
5-LWM (EW)	0.780	2.610	0.326	2.589	1.61	0.174	0.734	**0.084**	0.729	1.34	0.474	1.767	0.199	1.756	1.57
5-LWM (VW)	0.174	0.810	0.631	0.509	0.41	0.481	2.605	1.694	1.979	2.05	0.197	0.930	0.451	0.813	0.58
PCA(h^*)	1.583	7.641	7.641	**0.000**	23.00	1.565	7.549	7.543	0.311	23.00	1.583	7.641	7.641	**0.001**	23.00
Sieve(J^*)	0.769	3.754	3.753	0.067	4.00	0.770	3.754	3.753	**0.095**	4.00	0.770	3.757	3.754	0.152	4.00
IV spline(λ^*)	0.155	0.786	**0.303**	0.726	-	1.135	17.256	2.155	17.121	-	0.323	1.449	**0.149**	1.441	-
HE	0.861	4.962	4.962	0.074	-	0.861	5.003	5.002	0.122	-	0.860	4.964	4.961	0.173	-
PCA($\widetilde{h}$)	0.150	0.823	0.487	0.663	83.00	**0.103**	**0.586**	0.335	0.481	84.72	0.170	0.912	0.451	0.793	85.00
Sieve($\widetilde{J}$)	0.210	1.188	0.931	0.738	15.00	0.128	0.823	0.501	0.654	22.00	0.221	1.228	0.635	1.051	19.00
IV spline($\widetilde{\lambda}$)	**0.127**	**0.620**	0.447	0.430	-	0.282	1.387	1.207	0.684	-	**0.132**	**0.690**	0.416	0.551	-
LLS	0.358	1.589	1.199	1.043	-	1.341	6.633	6.288	2.113	-	0.308	1.433	0.976	1.050	-
Panel 2: DGP II															
2-LWM (EW)	0.024	0.139	0.103	0.093	0.00	0.035	**0.213**	0.089	**0.193**	0.09	0.042	0.268	0.076	0.257	0.06
2-LWM (VW)	0.036	0.217	0.204	0.075	0.00	0.124	0.765	0.575	0.504	0.43	0.037	0.229	0.196	0.119	0.00
3-LWM (EW)	0.246	0.581	0.076	0.576	0.63	0.134	0.422	0.086	0.413	0.65	0.321	0.886	0.070	0.883	0.58
3-LWM (VW)	0.032	0.182	0.170	0.066	0.51	0.106	0.626	0.308	0.545	1.18	0.043	0.228	0.160	0.163	0.47
4-LWM (EW)	0.613	1.331	0.108	1.326	1.30	0.232	0.674	**0.066**	0.671	1.31	0.577	1.460	0.103	1.456	1.19
4-LWM (VW)	0.065	0.313	0.069	0.306	1.23	0.122	0.647	0.170	0.624	1.87	0.112	0.494	0.074	0.488	1.19
5-LWM (EW)	1.038	2.309	0.152	2.304	1.99	0.331	0.950	0.066	0.948	1.97	0.817	2.103	0.128	2.099	1.82
5-LWM (VW)	0.101	0.449	0.065	0.444	1.90	0.155	0.744	0.096	0.737	2.52	0.167	0.696	0.079	0.692	1.82
PCA(h^*)	0.102	0.672	0.670	0.053	23.00	0.091	0.605	0.563	0.222	23.00	0.106	0.686	0.674	0.126	23.00
Sieve(J^*)	0.173	1.087	1.086	0.043	4.00	0.174	1.114	1.067	0.319	4.00	0.173	1.091	1.085	0.106	4.00
IV spline(λ^*)	0.077	0.472	**0.020**	0.471	-	1.141	12.496	1.517	12.404	-	0.361	1.637	**0.063**	1.636	-
HE	0.194	1.530	1.530	**0.034**	-	0.212	1.948	1.507	1.234	-	0.197	1.533	1.531	**0.085**	-
PCA($\widetilde{h}$)	0.032	0.211	0.119	0.175	28.00	0.040	0.280	0.137	0.244	27.49	0.053	0.350	0.181	0.300	27.00
Sieve($\widetilde{J}$)	0.032	0.244	0.168	0.178	10.00	0.049	0.352	0.178	0.304	10.00	0.056	0.403	0.172	0.364	10.00
IV spline($\widetilde{\lambda}$)	**0.017**	**0.103**	0.054	0.087	-	**0.034**	0.232	0.086	0.215	-	**0.028**	**0.183**	0.111	0.146	-
LLS	0.104	0.535	0.404	0.351	-	0.167	0.831	0.577	0.598	-	0.120	0.649	0.459	0.458	-
Panel 3: DGP III															
2-LWM (EW)	0.007	0.062	0.056	0.027	2.00	0.008	0.075	0.057	0.049	1.98	0.009	0.082	0.055	0.060	2.00
2-LWM (VW)	0.017	0.148	0.148	**0.011**	2.00	0.036	0.339	0.314	0.129	1.62	0.017	0.150	0.147	**0.027**	2.00
3-LWM (EW)	0.024	0.157	0.041	0.152	2.75	0.014	0.118	0.013	0.117	2.60	0.038	0.239	0.032	0.237	2.73
3-LWM (VW)	0.008	0.070	0.052	0.047	3.00	0.029	0.266	0.141	0.226	2.35	0.012	0.107	0.042	0.099	2.94
4-LWM (EW)	0.051	0.312	0.019	0.311	3.38	0.023	0.169	0.015	0.168	3.22	0.065	0.394	0.013	0.394	3.33
4-LWM (VW)	0.008	0.067	0.009	0.066	3.56	0.036	0.302	0.072	0.294	3.16	0.014	0.123	0.011	0.122	3.60
5-LWM (EW)	0.083	0.524	0.021	0.524	3.95	0.035	0.237	**0.012**	0.236	3.84	0.091	0.550	0.024	0.550	3.91
5-LWM (VW)	0.011	0.094	**0.005**	0.094	4.34	0.045	0.366	0.050	0.363	3.80	0.021	0.178	**0.010**	0.178	4.37
PCA(h^*)	0.025	0.251	0.251	0.013	23.00	0.021	0.215	0.192	0.098	23.00	0.026	0.253	0.251	0.031	23.00
Sieve(J^*)	0.009	0.091	0.085	0.031	4.00	0.010	0.104	0.087	0.056	4.00	0.011	0.112	0.084	0.073	4.00
IV spline(λ^*)	0.019	0.280	0.011	0.280	-	1.500	26.839	3.024	26.668	-	0.223	1.202	0.035	1.202	-
HE	0.009	0.090	0.088	0.021	-	0.010	0.096	0.088	**0.037**	-	0.011	0.102	0.088	0.051	-
PCA($\widetilde{h}$)	**0.005**	0.048	0.023	0.042	26.00	**0.006**	**0.063**	0.027	0.057	25.25	**0.009**	0.086	0.045	0.073	25.00
Sieve($\widetilde{J}$)	0.008	0.074	0.048	0.056	7.00	0.009	0.103	0.069	0.077	5.00	0.011	0.112	0.084	0.073	4.00
IV spline($\widetilde{\lambda}$)	0.005	**0.041**	0.014	0.038	-	0.011	0.103	0.028	0.099	-	0.010	**0.080**	0.016	0.079	-
LLS	0.040	0.275	0.208	0.180	-	0.048	0.422	0.397	0.144	-	0.043	0.306	0.212	0.220	-
Panel 4: DGP IV															
2-LWM (EW)	0.065	0.691	0.555	0.412	1.00	0.066	0.706	0.556	0.435	1.00	0.065	0.698	0.551	0.428	1.00
2-LWM (VW)	0.155	1.574	1.562	0.199	1.00	0.157	1.545	1.473	0.464	0.75	0.154	1.571	1.557	0.206	1.00
3-LWM (EW)	0.026	0.203	0.154	0.131	1.10	0.030	0.229	0.170	0.154	1.21	**0.028**	**0.224**	0.157	0.159	1.11
3-LWM (VW)	0.126	1.148	1.105	0.310	1.85	0.139	1.426	1.306	0.572	1.45	0.125	1.141	1.096	0.317	1.85
4-LWM (EW)	0.027	**0.191**	0.093	0.167	1.43	0.027	**0.194**	0.108	0.161	1.40	0.033	0.247	0.094	0.229	1.45
4-LWM (VW)	0.061	0.485	0.422	0.238	2.74	0.106	1.015	0.827	0.589	2.19	0.061	0.480	0.417	0.239	2.72
5-LWM (EW)	0.045	0.316	**0.045**	0.312	1.85	0.038	0.235	**0.042**	0.231	1.86	0.051	0.362	0.057	0.358	1.85
5-LWM (VW)	0.047	0.345	0.272	0.211	3.12	0.091	0.815	0.597	0.555	2.96	0.047	0.354	0.270	0.230	3.14
PCA(h^*)	0.121	1.239	1.239	**0.006**	23.00	0.118	1.191	1.183	0.132	23.00	0.121	1.239	1.239	**0.015**	23.00
Sieve(J^*)	0.131	1.114	1.114	0.009	4.00	0.131	1.112	1.111	0.020	4.00	0.131	1.114	1.114	0.023	4.00
IV spline(λ^*)	0.052	0.414	0.090	0.404	-	1.445	25.105	2.747	24.954	-	0.190	1.126	**0.048**	1.125	-
HE	0.184	1.896	1.896	0.007	-	0.184	1.896	1.896	**0.016**	-	0.184	1.897	1.896	0.018	-
PCA($\widetilde{h}$)	0.030	0.268	0.189	0.190	48.00	**0.027**	0.247	0.148	0.197	50.78	0.035	0.315	0.206	0.238	47.00
Sieve($\widetilde{J}$)	**0.024**	0.213	0.156	0.145	16.00	0.031	0.247	0.223	0.107	10.00	0.033	0.253	0.223	0.120	10.00
IV spline($\widetilde{\lambda}$)	0.033	0.246	0.193	0.153	-	0.062	0.459	0.420	0.185	-	0.034	0.272	0.193	0.192	-
LLS	0.045	0.370	0.283	0.238	-	0.051	0.468	0.430	0.184	-	0.045	0.378	0.268	0.267	-

Table F.5: Simulation results of the RND estimates based on the M-LWM model and five competing estimators with $N = 100$ and **small error variance** ($1\times$) under **alternative option error designs**. The specification and choice of tuning parameters of the competing estimators are detailed in Online Appendix D. For the LWM method, EW (VW) refers to the equal (Vega) weighting scheme. MTVD, RMISE, RISB and RIV respectively stands for mean total variation distance, root mean integrated squared error, root integrated squared bias, root integrated variance, and are defined in Eq. (25)-(27). M_1 reports (i) for LWM method, the average number of log normal densities chosen in each M-LWM model, (ii) for PCA method, the average number of normal mixtures, and (iii) for Sieve method, the average order of Hermite expansion (i.e. J^* or $\widetilde{J}$). The statistics in the table are computed based on 1,000 simulated paths. The best and the top three statistics among all estimators are highlighted in **bold** and underlined, respectively.

	$N = 100$ & Medium error variance (2×)														
	Alternative Error Design I					Alternative Error Design II					Alternative Error Design III				
Estimator	MTVD	RMISE	RISB	RIV	M_1	MTVD	RMISE	RISB	RIV	M_1	MTVD	RMISE	RISB	RIV	M_1
Panel 1: DGP I															
2-LWM (EW)	0.173	0.858	0.783	0.350	0.00	0.168	0.844	0.779	0.326	0.00	0.189	0.948	0.736	0.598	0.00
2-LWM (VW)	0.611	2.829	2.826	0.121	0.00	0.838	4.513	3.570	2.761	0.37	0.610	2.832	2.817	0.292	0.00
3-LWM (EW)	0.222	0.916	0.496	0.770	0.12	0.161	**0.695**	0.502	0.480	0.27	0.249	1.084	0.486	0.969	0.23
3-LWM (VW)	0.440	1.988	1.981	0.157	0.00	0.620	3.273	2.347	2.281	0.96	0.434	1.984	1.949	0.373	0.00
4-LWM (EW)	0.744	2.534	0.418	2.500	0.93	0.214	0.873	0.361	0.795	0.79	0.514	1.940	0.270	1.921	1.01
4-LWM (VW)	0.270	1.261	0.848	0.934	0.12	0.556	2.930	1.955	2.183	1.53	0.311	1.469	0.762	1.256	0.34
5-LWM (EW)	1.162	3.675	0.423	3.650	1.68	0.259	1.009	**0.174**	0.994	1.39	0.727	2.561	0.237	2.550	1.76
5-LWM (VW)	0.201	0.952	0.479	0.823	0.51	0.526	2.822	1.877	2.108	2.07	0.255	1.201	0.322	1.157	0.86
PCA(h^*)	1.583	7.641	7.641	**0.001**	23.00	1.563	7.541	7.532	0.372	23.00	1.583	7.641	7.641	**0.002**	23.00
Sieve(J^*)	0.769	3.753	3.752	0.094	4.00	0.772	3.753	3.750	**0.158**	4.00	0.770	3.756	3.750	0.214	4.00
IV spline(λ^*)	0.174	0.903	**0.375**	0.821	-	1.333	20.958	2.857	20.763	-	0.451	1.992	**0.150**	1.986	-
HE	0.861	4.964	4.963	0.105	-	0.860	5.008	5.005	0.173	-	0.861	4.969	4.962	0.250	-
PCA($\widetilde{h}$)	0.177	0.981	0.551	0.812	83.00	**0.124**	0.709	0.360	0.610	84.64	0.211	1.138	0.573	0.984	83.00
Sieve($\widetilde{J}$)	0.238	1.373	0.941	0.999	15.00	0.161	0.981	0.601	0.775	19.00	0.266	1.479	0.954	1.131	15.00
IV spline($\widetilde{\lambda}$)	**0.146**	**0.732**	0.490	0.545	-	0.309	1.609	1.322	0.916	-	**0.164**	**0.866**	0.497	0.709	-
LLS	0.399	1.798	1.332	1.208	-	1.548	7.639	7.465	1.623	-	0.358	1.701	1.150	1.253	-
Panel 2: DGP II															
2-LWM (EW)	0.029	0.171	0.099	0.139	0.01	0.043	**0.274**	**0.077**	0.263	0.15	0.059	0.379	**0.049**	0.376	0.15
2-LWM (VW)	0.036	0.218	0.202	0.083	0.00	0.129	0.781	0.596	0.505	0.44	0.041	0.260	0.200	0.166	0.01
3-LWM (EW)	0.369	0.844	0.077	0.841	0.64	0.182	0.585	0.082	0.579	0.70	0.483	1.292	0.085	1.289	0.68
3-LWM (VW)	0.033	0.190	0.167	0.090	0.47	0.114	0.683	0.311	0.609	1.24	0.063	0.293	0.153	0.250	0.50
4-LWM (EW)	0.851	1.841	0.141	1.836	1.32	0.305	0.921	0.089	0.917	1.36	0.840	2.114	0.120	2.111	1.29
4-LWM (VW)	0.092	0.413	0.072	0.406	1.18	0.158	0.785	0.198	0.760	1.93	0.176	0.712	0.090	0.707	1.21
5-LWM (EW)	1.474	3.164	0.184	3.159	1.96	0.449	1.300	0.106	1.295	1.96	1.228	3.052	0.170	3.047	1.91
5-LWM (VW)	0.161	0.635	0.067	0.631	1.83	0.198	0.913	0.138	0.903	2.58	0.265	1.043	0.094	1.039	1.82
PCA(h^*)	0.103	0.676	0.672	0.072	23.00	0.095	0.628	0.580	**0.240**	23.00	0.108	0.696	0.670	0.188	23.00
Sieve(J^*)	0.173	1.087	1.086	0.059	4.00	0.175	1.127	1.067	0.361	4.00	0.174	1.096	1.085	0.155	4.00
IV spline(λ^*)	0.083	0.527	**0.028**	0.526	-	1.170	13.994	1.717	13.888	-	0.502	2.278	0.100	2.276	-
HE	0.195	1.532	1.531	**0.045**	-	0.223	2.127	1.515	1.493	-	0.199	1.534	1.529	**0.126**	-
PCA($\widetilde{h}$)	0.040	0.264	0.166	0.205	27.00	0.051	0.352	0.191	0.296	26.65	0.064	0.430	0.252	0.348	26.00
Sieve($\widetilde{J}$)	0.041	0.302	0.171	0.248	10.00	0.082	0.551	0.476	0.277	7.00	0.088	0.569	0.462	0.333	7.00
IV spline($\widetilde{\lambda}$)	**0.021**	**0.130**	0.083	0.100	-	**0.042**	0.286	0.128	0.256	-	**0.036**	**0.236**	0.106	0.210	-
LLS	0.118	0.617	0.468	0.402	-	0.194	1.004	0.682	0.737	-	0.136	0.765	0.551	0.530	-
Panel 3: DGP III															
2-LWM (EW)	0.007	0.067	0.056	0.037	2.00	0.010	0.094	0.060	0.072	1.91	**0.011**	**0.105**	0.054	0.090	1.97
2-LWM (VW)	0.017	0.149	0.148	**0.015**	2.00	0.034	0.327	0.295	0.139	1.60	0.017	0.151	0.146	**0.038**	2.00
3-LWM (EW)	0.036	0.217	0.044	0.213	2.76	0.021	0.162	0.020	0.161	2.57	0.058	0.346	0.041	0.343	2.72
3-LWM (VW)	0.009	0.082	0.047	0.067	2.97	0.032	0.296	0.148	0.256	2.38	0.017	0.157	0.041	0.151	2.89
4-LWM (EW)	0.083	0.480	0.039	0.478	3.43	0.037	0.256	**0.017**	0.255	3.22	0.100	0.593	0.024	0.592	3.35
4-LWM (VW)	0.010	0.089	0.007	0.089	3.57	0.047	0.405	0.086	0.396	3.18	0.021	0.182	0.023	0.180	3.55
5-LWM (EW)	0.145	0.857	0.026	0.857	4.01	0.059	0.369	0.020	0.369	3.83	0.144	0.849	0.033	0.848	3.90
5-LWM (VW)	0.017	0.148	**0.006**	0.147	4.33	0.063	0.516	0.065	0.512	3.81	0.033	0.272	**0.013**	0.271	4.32
PCA(h^*)	0.025	0.251	0.250	0.018	23.00	0.021	0.213	0.184	0.106	23.00	0.026	0.256	0.252	0.044	23.00
Sieve(J^*)	0.009	0.096	0.085	0.046	4.00	0.011	0.114	0.083	0.078	4.00	0.013	0.135	0.087	0.103	4.00
IV spline(λ^*)	0.019	0.284	0.012	0.283	-	2.392	37.466	5.701	37.030	-	0.320	1.721	0.061	1.720	-
HE	0.010	0.092	0.087	0.030	-	0.011	0.103	0.088	**0.054**	-	0.012	0.115	0.089	0.073	-
PCA($\widetilde{h}$)	**0.006**	0.062	0.037	0.050	25.00	**0.008**	**0.083**	0.037	0.074	25.17	0.011	0.112	0.062	0.094	25.00
Sieve($\widetilde{J}$)	0.008	0.095	0.071	0.064	5.00	0.011	0.114	0.083	0.078	4.00	0.013	0.135	0.087	0.103	4.00
IV spline($\widetilde{\lambda}$)	0.007	**0.057**	0.014	0.055	-	0.015	0.139	0.044	0.132	-	0.014	0.114	0.014	0.113	-
LLS	0.047	0.324	0.244	0.214	-	0.043	0.382	0.310	0.224	-	0.050	0.368	0.258	0.263	-
Panel 4: DGP IV															
2-LWM (EW)	0.064	0.684	0.545	0.414	1.00	0.067	0.714	0.551	0.454	0.99	0.067	0.715	0.546	0.461	0.99
2-LWM (VW)	0.154	1.568	1.553	0.216	1.00	0.158	1.533	1.441	0.523	0.75	0.154	1.567	1.552	0.217	1.00
3-LWM (EW)	**0.027**	**0.211**	0.156	0.142	1.11	**0.032**	0.251	0.172	0.182	1.31	**0.031**	**0.257**	0.160	0.201	1.16
3-LWM (VW)	0.125	1.144	1.093	0.337	1.81	0.143	1.451	1.335	0.571	1.51	0.125	1.150	1.095	0.351	1.81
4-LWM (EW)	0.034	0.244	0.095	0.225	1.44	0.034	**0.247**	0.117	0.217	1.52	0.048	0.352	0.097	0.338	1.49
4-LWM (VW)	0.061	0.487	0.420	0.246	2.66	0.108	1.036	0.842	0.603	2.24	0.062	0.504	0.420	0.279	2.66
5-LWM (EW)	0.080	0.510	**0.061**	0.507	1.83	0.057	0.335	**0.051**	0.331	1.96	0.085	0.549	0.067	0.545	1.91
5-LWM (VW)	0.049	0.362	0.267	0.245	3.09	0.094	0.828	0.618	0.551	2.95	0.050	0.379	0.264	0.273	3.10
PCA(h^*)	0.121	1.239	1.239	**0.008**	23.00	0.118	1.185	1.175	0.150	23.00	0.121	1.239	1.239	**0.022**	23.00
Sieve(J^*)	0.131	1.113	1.113	0.012	4.00	0.131	1.111	1.110	0.028	4.00	0.131	1.114	1.114	0.032	4.00
IV spline(λ^*)	0.059	0.499	0.112	0.486	-	2.262	34.118	3.886	33.896	-	0.247	1.479	**0.057**	1.478	-
HE	0.184	1.896	1.896	0.010	-	0.184	1.896	1.896	**0.024**	-	0.184	1.896	1.896	0.026	-
PCA($\widetilde{h}$)	0.034	0.307	0.211	0.223	47.00	0.032	0.308	0.167	0.258	49.88	0.040	0.371	0.226	0.295	47.00
Sieve($\widetilde{J}$)	0.031	0.241	0.223	0.093	10.00	0.036	0.264	0.245	0.099	8.00	0.036	0.270	0.243	0.118	8.00
IV spline($\widetilde{\lambda}$)	0.037	0.285	0.205	0.197	-	0.075	0.569	0.548	0.153	-	0.039	0.321	0.215	0.238	-
LLS	0.051	0.421	0.325	0.267	-	0.055	0.512	0.418	0.295	-	0.051	0.437	0.309	0.308	-

Table F.6: Simulation results of the RND estimates based on the M-LWM model and five competing estimators with $N = 100$ and **medium error variance** (2×) under **alternative option error designs**. The specification and choice of tuning parameters of the competing estimators are detailed in Online Appendix D. For the LWM method, EW (VW) refers to the equal (Vega) weighting scheme. MTVD, RMISE, RISB and RIV respectively stands for mean total variation distance, root mean integrated squared error, root integrated squared bias, root integrated variance, and are defined in Eq. (25)-(27). M_1 reports (i) for LWM method, the average number of log normal densities chosen in each M-LWM model, (ii) for PCA method, the average number of normal mixtures, and (iii) for Sieve method, the average order of Hermite expansion (i.e. J^* or $\widetilde{J}$). The statistics in the table are computed based on 1,000 simulated paths. The best and the top three statistics among all estimators are highlighted in **bold** and underlined, respectively.

	$N = 100$ & Large error variance (4×)														
	Alternative Error Design I					Alternative Error Design II					Alternative Error Design III				
Estimator	MTVD	RMISE	RISB	RIV	M_1	MTVD	RMISE	RISB	RIV	M_1	MTVD	RMISE	RISB	RIV	M_1
Panel 1: DGP I															
2-LWM (EW)	0.181	0.911	0.760	0.502	0.00	0.174	0.886	0.752	0.470	0.00	0.227	1.149	0.727	0.890	0.01
2-LWM (VW)	0.613	2.837	2.833	0.154	0.00	0.946	4.963	4.050	2.869	0.49	0.611	2.849	2.823	0.381	0.00
3-LWM (EW)	0.353	1.293	0.476	1.203	0.31	0.209	**0.852**	0.463	0.715	0.39	0.412	1.632	0.478	1.561	0.40
3-LWM (VW)	0.442	2.003	1.992	0.211	0.00	0.706	3.731	2.722	2.552	1.08	0.435	2.020	1.953	0.515	0.00
4-LWM (EW)	1.057	3.454	0.452	3.425	1.10	0.277	1.083	0.397	1.008	0.88	0.786	2.825	0.418	2.794	1.18
4-LWM (VW)	0.311	1.464	0.726	1.271	0.23	0.637	3.385	2.285	2.497	1.60	0.374	1.746	0.699	1.600	0.48
5-LWM (EW)	1.539	4.690	0.402	4.673	1.88	0.386	1.340	**0.240**	1.318	1.43	1.075	3.603	0.342	3.587	1.95
5-LWM (VW)	0.252	1.212	**0.355**	1.159	0.68	0.643	3.449	2.148	2.699	2.09	0.338	1.572	0.333	1.536	1.06
PCA(h^*)	1.583	7.641	7.641	**0.001**	23.00	1.566	7.551	7.545	0.293	23.00	1.583	7.641	7.641	**0.002**	23.00
Sieve(J^*)	0.770	3.755	3.752	0.131	4.00	0.775	3.755	3.749	**0.211**	4.00	0.772	3.761	3.749	0.305	4.00
IV spline(λ^*)	0.203	1.057	0.488	0.938	-	1.550	23.635	2.798	23.468	-	0.612	2.708	**0.218**	2.699	-
HE	0.861	4.966	4.963	0.145	-	0.859	5.016	5.010	0.246	-	0.862	4.970	4.957	0.356	-
PCA($\widetilde{h}$)	0.221	1.175	0.839	0.822	78.00	**0.152**	0.869	0.408	0.768	84.80	0.254	1.372	0.855	1.073	78.00
Sieve($\widetilde{J}$)	0.276	1.591	0.985	1.249	15.00	0.201	1.182	0.637	0.996	19.00	0.319	1.765	0.975	1.471	15.00
IV spline($\widetilde{\lambda}$)	**0.174**	**0.885**	0.567	0.679	-	0.287	1.426	1.231	0.720	-	**0.203**	**1.075**	0.663	0.846	-
LLS	0.440	2.010	1.498	1.340	-	1.346	6.753	6.403	2.145	-	0.413	2.011	1.360	1.481	-
Panel 2: DGP II															
2-LWM (EW)	0.039	0.249	0.074	0.238	0.06	0.067	0.390	**0.077**	0.382	0.22	0.082	0.535	**0.051**	0.532	0.23
2-LWM (VW)	0.036	0.225	0.199	0.106	0.00	0.144	0.947	0.564	0.760	0.50	**0.045**	**0.294**	0.192	0.223	0.01
3-LWM (EW)	0.633	1.407	0.090	1.404	0.70	0.265	0.820	0.084	0.816	0.76	0.755	1.920	0.124	1.916	0.82
3-LWM (VW)	0.044	0.214	0.161	0.141	0.47	0.167	0.955	0.309	0.904	1.26	0.106	0.499	0.141	0.479	0.53
4-LWM (EW)	1.314	2.821	0.189	2.815	1.35	0.482	1.347	0.125	1.342	1.41	1.315	3.172	0.140	3.169	1.41
4-LWM (VW)	0.150	0.620	0.081	0.614	1.20	0.237	1.165	0.276	1.132	1.96	0.268	1.029	0.094	1.025	1.22
5-LWM (EW)	2.160	4.551	0.257	4.544	2.00	0.783	2.003	0.161	1.997	2.02	1.861	4.444	0.206	4.439	2.04
5-LWM (VW)	0.285	1.023	0.100	1.018	1.86	0.278	1.285	0.215	1.267	2.60	0.386	1.474	0.083	1.472	1.82
PCA(h^*)	0.104	0.682	0.674	0.101	23.00	0.098	0.649	0.578	**0.295**	23.00	0.112	0.718	0.668	0.264	23.00
Sieve(J^*)	0.173	1.089	1.086	0.084	4.00	0.178	1.150	1.071	0.419	4.00	0.175	1.110	1.090	0.210	4.00
IV spline(λ^*)	0.104	0.738	**0.026**	0.738	-	1.526	15.625	1.933	15.505	-	0.714	3.214	0.096	3.213	-
HE	0.196	1.533	1.532	**0.064**	-	0.233	2.243	1.525	1.644	-	0.202	1.539	1.529	**0.173**	-
PCA($\widetilde{h}$)	0.049	0.336	0.184	0.281	27.00	0.063	0.442	0.228	0.379	26.56	0.077	0.518	0.334	0.396	25.00
Sieve($\widetilde{J}$)	0.078	0.509	0.464	0.210	7.00	0.092	0.629	0.493	0.391	7.00	0.100	0.656	0.464	0.463	7.00
IV spline($\widetilde{\lambda}$)	**0.028**	**0.166**	0.085	0.142	-	**0.054**	**0.362**	0.194	0.306	-	0.046	0.298	0.160	0.251	-
LLS	0.133	0.715	0.541	0.467	-	0.218	1.155	0.786	0.846	-	0.159	0.901	0.661	0.611	-
Panel 3: DGP III															
2-LWM (EW)	0.008	**0.077**	0.055	0.054	2.00	0.012	0.118	0.065	0.098	1.84	0.015	0.156	0.053	0.147	1.92
2-LWM (VW)	0.017	0.149	0.148	**0.020**	2.00	0.033	0.334	0.269	0.198	1.56	0.017	0.153	0.144	**0.052**	2.00
3-LWM (EW)	0.053	0.313	0.044	0.310	2.77	0.030	0.230	0.027	0.228	2.49	0.081	0.491	0.028	0.490	2.66
3-LWM (VW)	0.013	0.127	0.045	0.119	2.95	0.038	0.343	0.161	0.303	2.35	0.026	0.235	0.051	0.229	2.88
4-LWM (EW)	0.129	0.718	0.027	0.717	3.43	0.056	0.375	**0.026**	0.374	3.15	0.153	0.878	0.030	0.877	3.25
4-LWM (VW)	0.015	0.137	0.015	0.136	3.59	0.067	0.566	0.119	0.553	3.13	0.034	0.285	0.027	0.284	3.55
5-LWM (EW)	0.254	1.335	0.056	1.334	3.97	0.096	0.568	0.027	0.567	3.77	0.231	1.287	0.037	1.287	3.75
5-LWM (VW)	0.027	0.229	**0.013**	0.228	4.37	0.078	0.639	0.083	0.633	3.74	0.054	0.424	**0.013**	0.424	4.28
PCA(h^*)	0.026	0.252	0.251	0.023	23.00	0.021	0.213	0.180	0.114	23.00	0.027	0.259	0.251	0.062	23.00
Sieve(J^*)	0.010	0.107	0.086	0.064	4.00	0.014	0.141	0.088	0.110	4.00	0.016	0.172	0.093	0.145	4.00
IV spline(λ^*)	0.025	0.423	0.016	0.422	-	2.393	35.597	5.077	35.233	-	0.439	2.401	0.088	2.400	-
HE	0.010	0.097	0.088	0.042	-	0.012	0.117	0.090	**0.075**	-	**0.014**	**0.139**	0.090	0.106	-
PCA($\widetilde{h}$)	**0.008**	0.080	0.042	0.067	25.00	**0.010**	**0.108**	0.049	0.096	25.07	0.015	0.147	0.081	0.124	25.00
Sieve($\widetilde{J}$)	0.010	0.107	0.086	0.064	4.00	0.014	0.141	0.088	0.110	4.00	0.016	0.172	0.093	0.145	4.00
IV spline($\widetilde{\lambda}$)	0.010	0.078	0.013	0.077	-	0.019	0.179	0.058	0.170	-	0.019	0.165	0.017	0.164	-
LLS	0.052	0.371	0.282	0.242	-	0.059	0.548	0.459	0.300	-	0.058	0.436	0.306	0.311	-
Panel 4: DGP IV															
2-LWM (EW)	0.066	0.704	0.551	0.437	0.99	0.068	0.717	0.538	0.474	0.99	0.067	0.695	0.534	0.446	0.99
2-LWM (VW)	0.154	1.569	1.555	0.213	1.00	0.162	1.576	1.463	0.586	0.75	0.154	1.566	1.549	0.229	1.00
3-LWM (EW)	**0.029**	**0.234**	0.158	0.173	1.15	**0.034**	**0.275**	0.157	0.226	1.34	**0.036**	0.317	0.175	0.264	1.29
3-LWM (VW)	0.125	1.136	1.090	0.320	1.84	0.145	1.461	1.337	0.589	1.59	0.125	1.154	1.091	0.374	1.81
4-LWM (EW)	0.060	0.415	0.095	0.404	1.48	0.043	0.318	0.116	0.296	1.55	0.075	0.533	0.118	0.519	1.63
4-LWM (VW)	0.061	0.481	0.416	0.241	2.72	0.114	1.092	0.880	0.647	2.33	0.064	0.534	0.426	0.323	2.62
5-LWM (EW)	0.160	0.871	**0.066**	0.869	1.91	0.076	0.447	**0.049**	0.444	2.04	0.150	0.864	**0.088**	0.860	2.06
5-LWM (VW)	0.054	0.389	0.261	0.289	3.18	0.102	0.894	0.656	0.608	2.99	0.054	0.418	0.271	0.319	3.11
PCA(h^*)	0.121	1.239	1.239	**0.011**	23.00	0.118	1.187	1.177	0.149	23.00	0.121	1.240	1.239	**0.032**	23.00
Sieve(J^*)	0.131	1.114	1.114	0.017	4.00	0.131	1.110	1.110	0.038	4.00	0.131	1.115	1.114	0.044	4.00
IV spline(λ^*)	0.066	0.600	0.139	0.583	-	2.028	32.757	4.070	32.503	-	0.338	2.053	0.115	2.049	-
HE	0.184	1.897	1.897	0.013	-	0.184	1.896	1.895	**0.034**	-	0.184	1.897	1.897	0.035	-
PCA($\widetilde{h}$)	0.040	0.358	0.273	0.232	43.00	0.040	0.369	0.232	0.288	45.96	0.047	0.439	0.298	0.323	43.00
Sieve($\widetilde{J}$)	0.035	0.256	0.243	0.079	8.00	0.038	0.285	0.250	0.137	8.00	0.038	**0.293**	0.245	0.160	8.00
IV spline($\widetilde{\lambda}$)	0.042	0.329	0.246	0.218	-	0.066	0.553	0.414	0.367	-	0.046	0.387	0.251	0.295	-
LLS	0.057	0.478	0.367	0.307	-	0.078	0.806	0.772	0.231	-	0.059	0.516	0.368	0.361	-

Table F.7: Simulation results of the RND estimates based on the M-LWM model and five competing estimators with $N = 100$ and **large error variance** (4×) under **alternative option error designs**. The specification and choice of tuning parameters of the competing estimators are detailed in Online Appendix D. For the LWM method, EW (VW) refers to the equal (Vega) weighting scheme. MTVD, RMISE, RISB and RIV respectively stands for mean total variation distance, root mean integrated squared error, root integrated squared bias, root integrated variance, and are defined in Eq. (25)-(27). M_1 reports (i) for LWM method, the average number of log normal densities chosen in each M-LWM model, (ii) for PCA method, the average number of normal mixtures, and (iii) for Sieve method, the average order of Hermite expansion (i.e. J^* or $\widetilde{J}$). The statistics in the table are computed based on 1,000 simulated paths. The best and the top three statistics among all estimators are highlighted in **bold** and underlined, respectively.

	$N = 25$ & Small error variance (1×)														
	Alternative Error Design I					Alternative Error Design II					Alternative Error Design III				
Estimator	MTVD	RMISE	RISB	RIV	M_1	MTVD	RMISE	RISB	RIV	M_1	MTVD	RMISE	RISB	RIV	M_1
Panel 1: DGP I															
2-LWM (EW)	0.189	0.948	0.805	0.502	0.00	0.186	0.892	0.866	0.210	0.00	0.187	0.916	0.849	0.344	0.00
2-LWM (VW)	0.607	2.844	2.791	0.544	0.03	0.814	4.468	3.138	3.181	0.39	0.606	2.842	2.788	0.549	0.02
3-LWM (EW)	0.314	1.156	0.514	1.035	0.29	0.136	**0.652**	0.601	0.253	0.05	0.145	0.702	0.574	0.405	0.00
3-LWM (VW)	0.349	1.575	1.461	0.588	0.07	0.587	3.134	1.996	2.416	1.07	0.346	1.569	1.457	0.581	0.06
4-LWM (EW)	1.113	4.081	1.058	3.941	1.09	0.351	1.790	**0.382**	1.749	0.63	0.506	2.338	0.711	2.228	0.85
4-LWM (VW)	0.606	2.515	1.151	2.236	0.77	0.531	2.862	1.811	2.216	1.66	0.566	2.324	1.100	2.047	0.77
5-LWM (EW)	0.914	3.543	0.757	3.462	1.74	0.329	1.663	0.475	1.593	1.23	0.438	2.087	0.802	1.927	1.51
5-LWM (VW)	0.421	1.796	1.025	1.475	1.38	0.507	2.758	1.788	2.100	2.21	0.366	1.540	0.967	1.198	1.40
PCA(h^*)	1.584	7.642	7.642	**0.001**	23.00	1.509	7.290	7.218	1.020	23.00	1.584	7.642	7.642	**0.002**	23.00
Sieve(J^*)	0.735	3.493	3.489	0.168	4.00	0.709	3.511	3.510	**0.073**	4.00	0.735	3.493	3.487	0.194	4.00
IV spline(λ^*)	0.216	1.102	**0.393**	1.030	-	0.462	6.379	1.384	6.227	-	0.144	0.634	0.392	0.498	-
HE	0.860	4.963	4.961	0.149	-	0.861	5.002	4.999	0.159	-	0.860	4.964	4.959	0.222	-
PCA($\widetilde{h}$)	0.211	1.121	0.781	0.805	78.00	**0.121**	0.690	0.622	0.298	76.36	0.138	0.763	**0.093**	0.757	125.00
Sieve($\widetilde{J}$)	0.260	1.485	0.822	1.237	15.00	0.147	0.946	0.824	0.464	15.00	0.175	1.059	0.853	0.627	15.00
IV spline($\widetilde{\lambda}$)	**0.162**	**0.863**	0.594	0.626	-	0.467	2.123	2.029	0.626	-	**0.129**	**0.584**	0.402	0.423	-
LLS	0.601	3.283	3.225	0.614	-	1.900	11.739	8.178	8.422	-	0.610	3.287	3.273	0.304	-
Panel 2: DGP II															
2-LWM (EW)	0.038	0.244	0.079	0.231	0.05	0.034	**0.211**	0.102	0.184	0.16	0.037	0.221	0.098	0.198	0.01
2-LWM (VW)	0.039	0.238	0.215	0.102	0.00	0.115	0.756	0.551	0.517	0.46	0.041	0.248	0.206	0.138	0.00
3-LWM (EW)	0.507	1.217	0.105	1.213	0.67	0.067	0.294	0.110	0.273	0.70	0.204	0.603	0.072	0.598	0.52
3-LWM (VW)	0.036	0.204	0.171	0.111	0.47	0.095	0.610	0.297	0.532	1.29	0.037	0.218	0.167	0.139	0.47
4-LWM (EW)	1.115	2.528	0.163	2.523	1.28	0.126	0.434	0.069	0.429	1.38	0.443	1.150	0.104	1.145	1.11
4-LWM (VW)	0.097	0.462	0.080	0.455	1.19	0.104	0.631	0.232	0.587	1.96	0.066	0.340	0.073	0.332	1.21
5-LWM (EW)	1.918	4.188	0.243	4.181	1.94	0.190	0.618	0.064	0.615	2.03	0.620	1.613	0.117	1.609	1.67
5-LWM (VW)	0.180	0.752	0.093	0.746	1.86	0.115	0.675	0.111	0.666	2.62	0.085	0.422	0.060	0.418	1.92
PCA(h^*)	0.105	0.683	0.676	0.101	23.00	0.078	0.553	0.460	0.306	23.00	0.106	0.688	0.675	0.134	23.00
Sieve(J^*)	0.150	1.002	0.998	0.089	4.00	0.163	1.051	0.984	0.370	4.00	0.151	1.007	1.000	0.119	4.00
IV spline(λ^*)	0.149	0.777	**0.032**	0.776	-	0.170	1.655	0.196	1.643	-	0.126	0.576	**0.026**	0.575	-
HE	0.196	1.536	1.534	**0.064**	-	0.193	1.512	1.507	**0.127**	-	0.199	1.537	1.534	**0.103**	-
PCA($\widetilde{h}$)	0.048	0.324	0.176	0.272	27.00	**0.033**	0.229	0.100	0.206	27.49	0.047	0.300	0.143	0.263	28.00
Sieve($\widetilde{J}$)	0.075	0.497	0.457	0.195	7.00	0.050	0.381	0.311	0.220	10.00	0.056	0.400	0.351	0.191	10.00
IV spline($\widetilde{\lambda}$)	**0.026**	**0.168**	0.125	0.112	-	0.039	0.252	**0.042**	0.249	-	**0.027**	**0.174**	0.059	0.164	-
LLS	0.133	0.716	0.551	0.457	-	0.247	1.058	0.428	0.968	-	0.094	0.521	0.336	0.398	-
Panel 3: DGP III															
2-LWM (EW)	0.008	0.078	0.054	0.056	2.00	0.008	0.077	0.058	0.050	1.96	0.009	0.080	0.055	0.058	2.00
2-LWM (VW)	0.018	0.155	0.154	**0.019**	2.00	0.032	0.320	0.245	0.206	1.59	0.018	0.157	0.153	**0.035**	2.00
3-LWM (EW)	0.055	0.310	0.050	0.306	2.83	0.012	0.094	0.014	0.093	2.63	0.026	0.173	0.029	0.170	2.77
3-LWM (VW)	0.013	0.121	0.049	0.110	2.97	0.029	0.274	0.123	0.245	2.21	0.011	0.093	0.045	0.082	2.99
4-LWM (EW)	0.123	0.709	0.038	0.708	3.52	0.017	0.128	**0.010**	0.128	3.32	0.042	0.272	0.012	0.272	3.43
4-LWM (VW)	0.012	0.111	0.022	0.108	3.72	0.036	0.343	0.047	0.340	2.92	0.009	0.083	0.016	0.082	3.70
5-LWM (EW)	0.192	1.110	0.046	1.109	4.06	0.025	0.177	0.013	0.176	3.96	0.052	0.343	0.020	0.343	3.98
5-LWM (VW)	0.022	0.186	0.017	0.185	4.52	0.041	0.372	0.055	0.368	3.55	0.012	0.107	0.009	0.106	4.45
PCA(h^*)	0.026	0.253	0.252	0.025	23.00	0.017	0.189	0.142	0.126	23.00	0.026	0.255	0.252	0.040	23.00
Sieve(J^*)	0.010	0.108	0.086	0.064	4.00	0.009	0.098	0.085	0.049	4.00	0.010	0.106	0.084	0.065	4.00
IV spline(λ^*)	0.030	0.284	**0.007**	0.284	-	0.321	8.265	1.467	8.134	-	0.034	0.232	**0.007**	0.232	-
HE	0.010	0.099	0.089	0.044	-	0.010	0.097	0.088	**0.041**	-	0.011	0.104	0.088	0.055	-
PCA($\widetilde{h}$)	0.008	0.083	0.047	0.068	25.00	**0.005**	**0.054**	0.021	0.049	24.41	0.008	0.078	0.027	0.073	26.00
Sieve($\widetilde{J}$)	0.010	0.108	0.086	0.064	4.00	0.008	0.092	0.067	0.063	5.00	0.010	0.103	0.066	0.079	5.00
IV spline($\widetilde{\lambda}$)	**0.005**	**0.045**	0.021	0.039	-	0.011	0.108	0.037	0.102	-	**0.006**	**0.054**	0.022	0.050	-
LLS	0.045	0.350	0.235	0.259	-	0.089	0.740	0.663	0.329	-	0.027	0.255	0.217	0.134	-
Panel 4: DGP IV															
2-LWM (EW)	0.065	0.690	0.546	0.422	1.00	0.066	0.710	0.556	0.442	1.00	0.063	0.664	0.540	0.386	1.00
2-LWM (VW)	0.155	1.582	1.564	0.234	1.00	0.153	1.495	1.395	0.538	0.82	0.155	1.584	1.567	0.231	1.00
3-LWM (EW)	**0.029**	**0.236**	0.163	0.172	1.17	0.030	0.231	0.176	0.149	1.21	0.027	0.209	0.156	0.138	1.10
3-LWM (VW)	0.125	1.158	1.096	0.374	1.79	0.137	1.378	1.239	0.603	1.55	0.125	1.156	1.095	0.370	1.79
4-LWM (EW)	0.071	0.454	**0.100**	0.443	1.51	0.024	**0.177**	0.107	0.141	1.44	0.028	**0.208**	0.096	0.184	1.44
4-LWM (VW)	0.078	0.617	0.500	0.361	2.66	0.104	0.992	0.812	0.570	2.23	0.073	0.587	0.485	0.331	2.67
5-LWM (EW)	0.179	1.024	0.126	1.016	1.96	0.031	0.196	**0.045**	0.191	1.90	0.038	0.279	**0.053**	0.273	1.83
5-LWM (VW)	0.056	0.420	0.342	0.245	3.08	0.088	0.801	0.583	0.550	2.92	0.051	0.387	0.340	0.185	2.99
PCA(h^*)	0.121	1.239	1.239	**0.012**	23.00	0.113	1.123	1.090	0.269	23.00	0.121	1.239	1.239	**0.020**	23.00
Sieve(J^*)	0.118	0.947	0.947	0.021	4.00	0.118	0.946	0.946	0.024	4.00	0.118	0.947	0.947	0.031	4.00
IV spline(λ^*)	0.081	0.569	0.124	0.556	-	0.243	3.898	0.831	3.808	-	0.041	0.284	0.113	0.260	-
HE	0.184	1.896	1.896	0.014	-	0.184	1.896	1.896	**0.021**	-	0.184	1.896	1.896	0.022	-
PCA($\widetilde{h}$)	0.041	0.366	0.277	0.239	43.00	**0.022**	0.200	0.135	0.148	49.24	0.028	0.251	0.143	0.207	52.00
Sieve($\widetilde{J}$)	0.031	0.250	0.223	0.113	9.00	0.025	0.214	0.166	0.135	14.00	**0.025**	0.211	0.165	0.132	14.00
IV spline($\widetilde{\lambda}$)	0.046	0.336	0.241	0.234	-	0.072	0.479	0.432	0.207	-	0.036	0.233	0.166	0.163	-
LLS	0.057	0.513	0.458	0.232	-	0.086	0.793	0.694	0.384	-	0.054	0.467	0.454	0.111	-

Table F.8: Simulation results of the RND estimates based on the M-LWM model and five competing estimators with $N = 25$ and **small error variance** (1×) under **alternative option error designs**. The specification and choice of tuning parameters of the competing estimators are detailed in Online Appendix D. For the LWM method, EW (VW) refers to the equal (Vega) weighting scheme. MTVD, RMISE, RISB and RIV respectively stands for mean total variation distance, root mean integrated squared error, root integrated squared bias, root integrated variance, and are defined in Eq. (25)-(27). M_1 reports (i) for LWM method, the average number of log normal densities chosen in each M-LWM model, (ii) for PCA method, the average number of normal mixtures, and (iii) for Sieve method, the average order of Hermite expansion (i.e. J^* or $\widetilde{J}$). The statistics in the table are computed based on 1,000 simulated paths. The best and the top three statistics among all estimators are highlighted in **bold** and underlined, respectively.

	$N = 25$ & Medium error variance (2×)														
	Alternative Error Design I					Alternative Error Design II					Alternative Error Design III				
Estimator	MTVD	RMISE	RISB	RIV	M_1	MTVD	RMISE	RISB	RIV	M_1	MTVD	RMISE	RISB	RIV	M_1
Panel 1: DGP I															
2-LWM (EW)	0.213	1.085	0.760	0.774	0.02	0.180	0.892	0.837	0.306	0.00	0.193	0.950	0.820	0.479	0.00
2-LWM (VW)	0.608	2.858	2.798	0.584	0.03	0.942	5.042	3.306	3.808	0.51	0.607	2.848	2.796	0.542	0.02
3-LWM (EW)	0.518	1.985	**0.358**	1.952	0.47	0.161	0.741	0.572	0.472	0.18	0.173	0.816	0.544	0.608	0.05
3-LWM (VW)	0.353	1.613	1.468	0.670	0.06	0.656	3.433	2.127	2.694	1.20	0.352	1.646	1.461	0.758	0.04
4-LWM (EW)	1.285	4.562	0.891	4.474	1.24	0.301	1.443	0.311	1.409	0.75	0.602	2.662	0.739	2.557	0.89
4-LWM (VW)	0.607	2.557	1.010	2.349	0.74	0.611	3.231	1.884	2.624	1.73	0.551	2.296	0.981	2.076	0.71
5-LWM (EW)	1.198	4.370	0.657	4.320	1.85	0.335	1.574	**0.242**	1.555	1.32	0.526	2.437	0.766	2.313	1.56
5-LWM (VW)	0.489	2.110	0.987	1.865	1.37	0.572	3.065	1.856	2.439	2.23	0.378	1.618	0.798	1.407	1.29
PCA(h^*)	1.584	7.642	7.642	**0.001**	23.00	1.501	7.256	7.173	1.093	23.00	1.584	7.642	7.642	**0.002**	23.00
Sieve(J^*)	0.736	3.500	3.491	0.240	4.00	0.710	3.512	3.511	**0.103**	4.00	0.735	3.498	3.487	0.279	4.00
IV spline(λ^*)	0.256	1.328	0.439	1.253	-	0.457	5.400	1.294	5.243	-	0.170	0.801	**0.400**	0.694	-
HE	0.859	4.964	4.959	0.212	-	0.860	5.010	5.006	0.207	-	0.860	4.968	4.958	0.319	-
PCA($\widetilde{h}$)	0.245	1.311	0.914	0.940	76.00	**0.131**	**0.729**	0.612	0.397	75.92	0.178	0.914	0.594	0.694	83.00
Sieve($\widetilde{J}$)	0.317	1.770	0.836	1.560	15.00	0.160	1.004	0.806	0.599	15.00	0.206	1.180	0.825	0.843	15.00
IV spline($\widetilde{\lambda}$)	**0.194**	**1.039**	0.712	0.756	-	0.459	2.161	1.982	0.862	-	**0.144**	**0.711**	0.442	0.557	-
LLS	0.598	3.300	3.189	0.849	-	2.086	14.161	8.436	11.374	-	0.606	3.280	3.253	0.420	-
Panel 2: DGP II															
2-LWM (EW)	0.054	0.352	0.052	0.349	0.14	**0.040**	**0.252**	0.101	0.231	0.20	0.051	0.322	0.072	0.314	0.07
2-LWM (VW)	0.039	0.242	0.207	0.125	0.00	0.126	0.838	0.532	0.647	0.46	0.044	0.275	0.197	0.191	0.01
3-LWM (EW)	0.902	2.013	0.156	2.007	0.79	0.114	0.431	0.112	0.416	0.74	0.330	0.939	0.072	0.936	0.60
3-LWM (VW)	0.047	0.250	0.167	0.187	0.47	0.114	0.770	0.295	0.711	1.29	0.043	0.256	0.153	0.206	0.45
4-LWM (EW)	1.742	3.783	0.238	3.775	1.43	0.202	0.647	0.085	0.641	1.42	0.684	1.707	0.103	1.704	1.21
4-LWM (VW)	0.163	0.695	0.087	0.689	1.17	0.141	0.849	0.248	0.812	1.99	0.092	0.461	0.069	0.456	1.17
5-LWM (EW)	2.745	5.807	0.274	5.801	2.08	0.317	0.954	0.085	0.950	2.07	0.941	2.360	0.125	2.357	1.85
5-LWM (VW)	0.315	1.239	0.113	1.234	1.82	0.162	0.875	0.168	0.859	2.65	0.142	0.648	0.077	0.643	1.84
PCA(h^*)	0.107	0.692	0.676	0.147	23.00	0.080	0.559	0.452	0.329	23.00	0.110	0.706	0.680	0.191	23.00
Sieve(J^*)	0.151	1.008	1.000	0.127	4.00	0.168	1.091	0.996	0.447	4.00	0.152	1.018	1.004	0.169	4.00
IV spline(λ^*)	0.158	0.891	**0.028**	0.890	-	0.138	1.098	0.143	1.088	-	0.168	0.781	**0.018**	0.781	-
HE	0.198	1.538	1.535	**0.092**	-	0.193	1.510	1.500	**0.177**	-	0.202	1.543	1.536	**0.145**	-
PCA($\widetilde{h}$)	0.060	0.406	0.256	0.315	26.00	0.043	0.305	0.118	0.281	27.28	0.061	0.392	0.187	0.345	27.00
Sieve($\widetilde{J}$)	0.080	0.540	0.458	0.287	7.00	0.073	0.518	0.452	0.253	7.00	0.062	0.446	0.349	0.278	10.00
IV spline($\widetilde{\lambda}$)	**0.032**	**0.202**	0.125	0.158	-	0.049	0.316	**0.056**	0.311	-	**0.035**	**0.236**	0.127	0.198	-
LLS	0.149	0.826	0.638	0.524	-	0.351	1.441	0.798	1.200	-	0.116	0.640	0.433	0.471	-
Panel 1: DGP III															
2-LWM (EW)	0.010	0.101	0.052	0.086	1.98	0.010	0.094	0.064	0.069	1.88	0.011	0.099	0.056	0.082	1.99
2-LWM (VW)	0.018	0.157	0.154	**0.028**	2.00	0.033	0.331	0.234	0.234	1.57	0.018	0.161	0.153	**0.051**	2.00
3-LWM (EW)	0.087	0.493	0.048	0.491	2.82	0.018	0.136	0.021	0.134	2.58	0.039	0.246	0.022	0.245	2.77
3-LWM (VW)	0.021	0.195	0.058	0.186	2.94	0.034	0.322	0.145	0.287	2.18	0.014	0.126	0.044	0.118	2.99
4-LWM (EW)	0.201	1.083	0.046	1.082	3.49	0.030	0.200	**0.019**	0.199	3.27	0.066	0.412	0.022	0.411	3.44
4-LWM (VW)	0.020	0.186	0.034	0.183	3.70	0.045	0.422	0.071	0.416	2.92	0.014	0.124	0.017	0.123	3.73
5-LWM (EW)	0.313	1.698	0.078	1.696	4.02	0.044	0.286	0.023	0.285	3.95	0.086	0.551	0.022	0.550	3.98
5-LWM (VW)	0.036	0.305	0.016	0.305	4.48	0.053	0.484	0.068	0.479	3.51	0.019	0.166	0.008	0.166	4.49
PCA(h^*)	0.026	0.253	0.251	0.035	23.00	0.019	0.199	0.148	0.133	23.00	0.026	0.257	0.251	0.056	23.00
Sieve(J^*)	0.012	0.123	0.085	0.089	4.00	0.010	0.109	0.084	0.071	4.00	0.012	0.123	0.082	0.091	4.00
IV spline(λ^*)	0.039	0.399	**0.013**	0.399	-	0.312	7.505	1.621	7.328	-	0.043	0.311	**0.007**	0.311	-
HE	0.011	0.107	0.087	0.062	-	0.011	0.105	0.087	**0.059**	-	0.012	0.116	0.087	0.076	-
PCA($\widetilde{h}$)	0.010	0.104	0.060	0.084	25.00	**0.007**	**0.072**	0.029	0.066	24.40	0.011	0.102	0.044	0.093	25.00
Sieve($\widetilde{J}$)	0.012	0.123	0.085	0.089	4.00	0.010	0.109	0.084	0.071	4.00	0.012	0.123	0.082	0.091	4.00
IV spline($\widetilde{\lambda}$)	**0.007**	**0.060**	0.021	0.056	-	0.015	0.141	0.045	0.133	-	**0.008**	**0.074**	0.022	0.070	-
LLS	0.059	0.431	0.304	0.305	-	0.112	0.984	0.704	0.688	-	0.034	0.294	0.221	0.194	-
Panel 2: DGP IV															
2-LWM (EW)	0.068	0.726	0.557	0.466	1.00	0.066	0.707	0.549	0.446	0.99	0.066	0.715	0.553	0.453	0.99
2-LWM (VW)	0.156	1.586	1.570	0.226	1.00	0.158	1.526	1.416	0.569	0.77	0.156	1.591	1.576	0.215	1.00
3-LWM (EW)	**0.034**	0.283	0.174	0.224	1.27	0.031	0.248	0.179	0.171	1.30	0.029	**0.230**	0.160	0.166	1.11
3-LWM (VW)	0.125	1.158	1.100	0.365	1.81	0.144	1.441	1.303	0.616	1.54	0.126	1.159	1.102	0.357	1.81
4-LWM (EW)	0.132	0.755	**0.116**	0.746	1.65	0.030	**0.209**	0.106	0.181	1.52	0.041	0.296	0.101	0.278	1.43
4-LWM (VW)	0.080	0.620	0.498	0.369	2.68	0.111	1.050	0.861	0.601	2.23	0.073	0.583	0.485	0.324	2.67
5-LWM (EW)	0.302	1.473	0.162	1.464	2.13	0.044	0.274	**0.048**	0.270	2.00	0.072	0.497	**0.088**	0.489	1.86
5-LWM (VW)	0.061	0.457	0.344	0.301	3.13	0.090	0.832	0.632	0.542	2.94	0.053	0.403	0.342	0.212	3.02
PCA(h^*)	0.121	1.239	1.239	**0.017**	23.00	0.112	1.113	1.071	0.304	23.00	0.121	1.239	1.239	**0.027**	23.00
Sieve(J^*)	0.118	0.947	0.947	0.030	4.00	0.118	0.945	0.945	0.034	4.00	0.118	0.948	0.947	0.043	4.00
IV spline(λ^*)	0.095	0.679	0.137	0.665	-	0.263	5.053	1.033	4.947	-	0.055	0.380	0.115	0.362	-
HE	0.184	1.897	1.896	0.019	-	0.184	1.896	1.895	**0.032**	-	0.184	1.897	1.896	0.031	-
PCA($\widetilde{h}$)	0.046	0.421	0.300	0.296	43.00	**0.026**	0.242	0.150	0.191	48.61	0.033	0.303	0.192	0.234	48.00
Sieve($\widetilde{J}$)	0.036	**0.266**	0.236	0.123	8.00	0.030	0.241	0.223	0.090	9.00	**0.028**	0.243	0.165	0.178	14.00
IV spline($\widetilde{\lambda}$)	0.052	0.387	0.284	0.263	-	0.083	0.599	0.563	0.206	-	0.041	0.276	0.197	0.194	-
LLS	0.061	0.555	0.459	0.312	-	0.133	1.397	0.918	1.053	-	0.054	0.478	0.452	0.155	-

Table F.9: Simulation results of the RND estimates based on the M-LWM model and five competing estimators with $N = 25$ and **medium error variance** (2×) under **alternative option error designs**. The specification and choice of tuning parameters of the competing estimators are detailed in Online Appendix D. For the LWM method, EW (VW) refers to the equal (Vega) weighting scheme. MTVD, RMISE, RISB and RIV respectively stands for mean total variation distance, root mean integrated squared error, root integrated squared bias, root integrated variance, and are defined in Eq. (25)-(27). M_1 reports (i) for LWM method, the average number of log normal densities chosen in each M-LWM model, (ii) for PCA method, the average number of normal mixtures, and (iii) for Sieve method, the average order of Hermite expansion (i.e. J^* or $\widetilde{J}$). The statistics in the table are computed based on 1,000 simulated paths. The best and the top three statistics among all estimators are highlighted in **bold** and underlined, respectively.

	$N = 25$ & Large error variance (4×)														
	Alternative Error Design I					Alternative Error Design II					Alternative Error Design III				
Estimator	MTVD	RMISE	RISB	RIV	M_1	MTVD	RMISE	RISB	RIV	M_1	MTVD	RMISE	RISB	RIV	M_1
Panel 1: DGP I															
2-LWM (EW)	0.273	1.404	0.737	1.195	0.09	0.179	0.905	0.805	0.413	0.00	0.213	1.057	0.783	0.710	0.00
2-LWM (VW)	0.612	2.879	2.814	0.608	0.03	1.153	5.839	3.712	4.507	0.54	0.610	2.889	2.800	0.713	0.03
3-LWM (EW)	1.022	3.830	**0.504**	3.797	0.72	0.195	**0.875**	0.520	0.703	0.31	0.263	1.136	0.512	1.015	0.20
3-LWM (VW)	0.357	1.669	1.468	0.795	0.06	0.818	4.307	2.547	3.473	1.13	0.356	1.662	1.471	0.774	0.07
4-LWM (EW)	1.644	5.512	0.851	5.446	1.48	0.327	1.488	0.323	1.453	0.82	0.764	3.146	0.710	3.065	0.97
4-LWM (VW)	0.629	2.690	0.941	2.520	0.70	0.727	3.807	2.035	3.218	1.65	0.552	2.345	0.870	2.178	0.70
5-LWM (EW)	1.753	6.024	0.826	5.967	2.01	0.419	1.970	**0.246**	1.954	1.43	0.685	2.918	0.691	2.835	1.61
5-LWM (VW)	0.584	2.574	0.972	2.383	1.37	0.691	3.616	1.899	3.077	2.15	0.419	1.836	0.754	1.675	1.27
PCA(h^*)	1.584	7.642	7.642	**0.002**	23.00	1.498	7.247	7.155	1.153	23.00	1.584	7.642	7.642	**0.003**	23.00
Sieve(J^*)	0.736	3.503	3.486	0.336	4.00	0.710	3.515	3.512	**0.145**	4.00	0.737	3.512	3.490	0.388	4.00
IV spline(λ^*)	0.308	1.605	0.564	1.503	-	0.490	5.231	1.320	5.062	-	0.216	1.065	**0.401**	0.987	-
HE	0.862	4.969	4.960	0.296	-	0.858	5.004	4.989	0.377	-	0.859	4.968	4.949	0.443	-
PCA($\widetilde{h}$)	0.285	1.562	1.016	1.186	76.00	**0.164**	0.936	0.662	0.662	74.24	0.214	1.119	0.682	0.887	80.00
Sieve($\widetilde{J}$)	0.370	2.016	1.502	1.344	11.00	0.186	1.131	0.809	0.790	15.00	0.255	1.428	0.810	1.176	15.00
IV spline($\widetilde{\lambda}$)	**0.234**	**1.249**	0.872	0.895	-	0.385	2.001	1.519	1.303	-	**0.175**	**0.907**	0.519	0.744	-
LLS	0.618	3.353	3.149	1.151	-	1.853	9.526	9.058	2.951	-	0.603	3.290	3.234	0.604	-
Panel 2: DGP II															
2-LWM (EW)	0.078	0.520	0.088	0.512	0.25	**0.051**	**0.330**	**0.105**	0.313	0.25	0.071	0.459	0.051	0.456	0.17
2-LWM (VW)	0.042	0.274	0.204	0.183	0.01	0.160	1.205	0.575	1.058	0.61	0.051	0.338	0.192	0.278	0.02
3-LWM (EW)	1.366	3.063	0.280	3.050	0.91	0.147	0.555	0.110	0.544	0.80	0.491	1.370	0.073	1.368	0.71
3-LWM (VW)	0.086	0.389	0.149	0.359	0.50	0.160	1.071	0.331	1.018	1.42	0.060	0.349	0.138	0.321	0.48
4-LWM (EW)	2.543	5.417	0.362	5.404	1.55	0.284	0.867	0.121	0.858	1.48	0.911	2.386	0.130	2.383	1.29
4-LWM (VW)	0.265	1.012	0.090	1.008	1.18	0.184	1.133	0.352	1.076	2.10	0.147	0.697	0.081	0.692	1.15
5-LWM (EW)	3.770	7.761	0.345	7.754	2.21	0.486	1.347	0.130	1.341	2.10	1.324	3.304	0.196	3.298	1.87
5-LWM (VW)	0.513	1.820	0.115	1.817	1.83	0.219	1.215	0.284	1.181	2.73	0.216	0.945	0.080	0.942	1.80
PCA(h^*)	0.111	0.717	0.685	0.213	23.00	0.088	0.602	0.466	0.381	23.00	0.115	0.733	0.680	0.273	23.00
Sieve(J^*)	0.152	1.022	1.006	0.180	4.00	0.171	1.124	1.006	0.501	4.00	0.155	1.037	1.009	0.241	4.00
IV spline(λ^*)	0.176	1.095	**0.057**	1.093	-	0.242	3.664	0.378	3.644	-	0.208	1.023	**0.026**	1.022	-
HE	0.201	1.545	1.539	**0.134**	-	0.196	1.515	1.497	**0.234**	-	0.206	1.550	1.537	**0.205**	-
PCA($\widetilde{h}$)	0.074	0.499	0.354	0.352	25.00	0.056	0.388	0.203	0.330	25.74	0.076	0.502	0.268	0.424	26.00
Sieve($\widetilde{J}$)	0.092	0.622	0.476	0.401	7.00	0.086	0.590	0.465	0.363	7.00	0.092	0.607	0.463	0.392	7.00
IV spline($\widetilde{\lambda}$)	**0.041**	**0.260**	0.161	0.204	-	0.061	0.393	0.112	0.377	-	**0.047**	**0.308**	0.143	0.273	-
LLS	0.169	0.953	0.733	0.608	-	0.367	1.569	0.897	1.287	-	0.134	0.763	0.532	0.548	-
Panel 3: DGP III															
2-LWM (EW)	0.015	0.159	0.046	0.153	1.91	0.012	0.120	0.073	0.095	1.76	0.014	0.130	0.051	0.119	1.97
2-LWM (VW)	0.018	0.159	0.154	**0.039**	2.00	0.033	0.369	0.218	0.297	1.51	0.018	0.169	0.153	**0.071**	2.00
3-LWM (EW)	0.128	0.729	0.059	0.727	2.70	0.025	0.183	0.029	0.181	2.44	0.059	0.360	0.035	0.358	2.75
3-LWM (VW)	0.032	0.277	0.083	0.264	2.93	0.041	0.402	0.162	0.368	2.14	0.021	0.188	0.041	0.184	2.95
4-LWM (EW)	0.301	1.573	0.083	1.571	3.29	0.043	0.280	**0.028**	0.279	3.16	0.099	0.600	0.024	0.600	3.37
4-LWM (VW)	0.035	0.299	0.063	0.292	3.70	0.063	0.586	0.090	0.579	2.89	0.023	0.204	0.029	0.202	3.73
5-LWM (EW)	0.497	2.536	0.142	2.532	3.78	0.073	0.453	0.040	0.452	3.81	0.140	0.854	0.036	0.853	3.90
5-LWM (VW)	0.061	0.480	0.049	0.478	4.49	0.075	0.665	0.081	0.660	3.49	0.033	0.269	0.023	0.268	4.50
PCA(h^*)	0.026	0.255	0.250	0.049	23.00	0.020	0.208	0.144	0.150	23.00	0.027	0.261	0.249	0.078	23.00
Sieve(J^*)	0.014	0.149	0.096	0.115	4.00	0.012	0.131	0.084	0.100	4.00	0.015	0.156	0.091	0.127	4.00
IV spline(λ^*)	0.047	0.522	0.038	0.520	-	0.310	7.853	1.520	7.704	-	0.052	0.392	**0.015**	0.392	-
HE	0.013	0.124	0.089	0.086	-	0.013	0.123	0.089	**0.084**	-	0.014	0.137	0.087	0.106	-
PCA($\widetilde{h}$)	0.013	0.131	0.081	0.103	25.00	**0.010**	**0.100**	0.047	0.088	24.25	0.014	0.138	0.063	0.123	25.00
Sieve($\widetilde{J}$)	0.014	0.149	0.096	0.115	4.00	0.012	0.131	0.084	0.100	4.00	0.015	0.156	0.091	0.127	4.00
IV spline($\widetilde{\lambda}$)	**0.009**	**0.081**	**0.022**	0.078	-	0.020	0.183	0.060	0.173	-	**0.011**	**0.101**	0.020	0.099	-
LLS	0.067	0.497	0.383	0.317	-	0.155	1.551	0.941	1.233	-	0.043	0.349	0.229	0.263	-
Panel 4: DGP IV															
2-LWM (EW)	0.069	0.728	0.531	0.498	0.98	0.067	0.707	0.541	0.455	0.99	0.066	0.701	0.543	0.443	1.00
2-LWM (VW)	0.155	1.583	1.566	0.234	1.00	0.166	1.636	1.476	0.705	0.83	0.155	1.581	1.563	0.243	1.00
3-LWM (EW)	0.043	0.378	0.185	0.330	1.33	0.033	0.264	0.166	0.205	1.36	**0.031**	**0.255**	0.164	0.195	1.16
3-LWM (VW)	0.125	1.162	1.097	0.384	1.80	0.146	1.468	1.299	0.685	1.66	0.126	1.167	1.098	0.395	1.77
4-LWM (EW)	0.213	1.123	**0.146**	1.114	1.73	0.036	**0.256**	0.107	0.233	1.60	0.059	0.409	0.099	0.397	1.50
4-LWM (VW)	0.089	0.670	0.516	0.427	2.64	0.114	1.092	0.872	0.657	2.33	0.076	0.622	0.499	0.371	2.61
5-LWM (EW)	0.481	2.239	0.263	2.223	2.19	0.061	0.373	**0.054**	0.369	2.09	0.114	0.717	**0.083**	0.713	1.92
5-LWM (VW)	0.072	0.528	0.360	0.386	3.15	0.098	0.886	0.666	0.584	2.98	0.055	0.424	0.338	0.256	3.08
PCA(h^*)	0.121	1.239	1.239	**0.024**	23.00	0.111	1.107	1.055	0.333	23.00	0.121	1.240	1.239	**0.041**	23.00
Sieve(J^*)	0.118	0.947	0.946	0.042	4.00	0.118	0.945	0.944	0.048	4.00	0.118	0.949	0.947	0.062	4.00
IV spline(λ^*)	0.104	0.790	0.168	0.772	-	0.312	6.891	1.415	6.744	-	0.069	0.483	0.115	0.469	-
HE	0.184	1.897	1.896	0.027	-	0.184	1.896	1.895	**0.047**	-	0.184	1.897	1.896	0.045	-
PCA($\widetilde{h}$)	0.054	0.482	0.368	0.311	39.00	**0.032**	0.301	0.171	0.248	47.72	0.039	0.355	0.224	0.276	47.00
Sieve($\widetilde{J}$)	**0.039**	**0.294**	0.237	0.174	8.00	0.035	0.257	0.234	0.108	8.00	0.036	0.268	0.235	0.129	8.00
IV spline($\widetilde{\lambda}$)	0.062	0.462	0.387	0.253	-	0.099	0.775	0.723	0.279	-	0.045	0.322	0.215	0.241	-
LLS	0.070	0.635	0.491	0.403	-	0.213	2.099	1.775	1.120	-	0.056	0.499	0.454	0.206	-

Table F.10: Simulation results of the RND estimates based on the M-LWM model and five competing estimators with $N = 25$ and **large error variance** (4×) under **alternative option error designs**. The specification and choice of tuning parameters of the competing estimators are detailed in Online Appendix D. For the LWM method, EW (VW) refers to the equal (Vega) weighting scheme. MTVD, RMISE, RISB and RIV respectively stands for mean total variation distance, root mean integrated squared error, root integrated squared bias, root integrated variance, and are defined in Eq. (25)-(27). M_1 reports (i) for LWM method, the average number of log normal densities chosen in each M-LWM model, (ii) for PCA method, the average number of normal mixtures, and (iii) for Sieve method, the average order of Hermite expansion (i.e. J^* or $\widetilde{J}$). The statistics in the table are computed based on 1,000 simulated paths. The best and the top three statistics among all estimators are highlighted in **bold** and underlined, respectively.

F.4 Coverage Properties of LWM Estimators

To examine finite-sample properties of the M-LWM estimator, in this section we study the coverage properties of confidence intervals (CIs) provided by this estimator. For each simulation scenario presented in Table 4.2, we compute the 95% estimated CIs over 10,000 equidistant grid points in the relevant strike range per simulated path. On each grid point, we count the number of times the true RND coordinate is contained in the corresponding estimated 95% CI and convert it into proportions (out of 1000 simulated paths). Plots of the 95% coverage probabilities, computed in a pointwise fashion, are depicted in Figure F.1.

Figure F.1 demonstrates that although the M-LWM estimators generally have smaller coverage than the nominal 95 when the sample size is small ($N = 25$). Also, for $M = 2$ when the parametric model is less flexible, we also expect low coverage rate as the estimated RND can be very different from the true one. As the sample size increases to $N = 100$ and with $M \geq 3$, the coverage probabilities are at least 95% (and even close to 100%) for the major part of the strike range. Therefore, despite that the finite sample confidence bands under possible model misspecification are not valid asymptotically, they still provide a conservative estimate for the range of the true RND when the model is mildly misspecified, which is of reference value to the users.

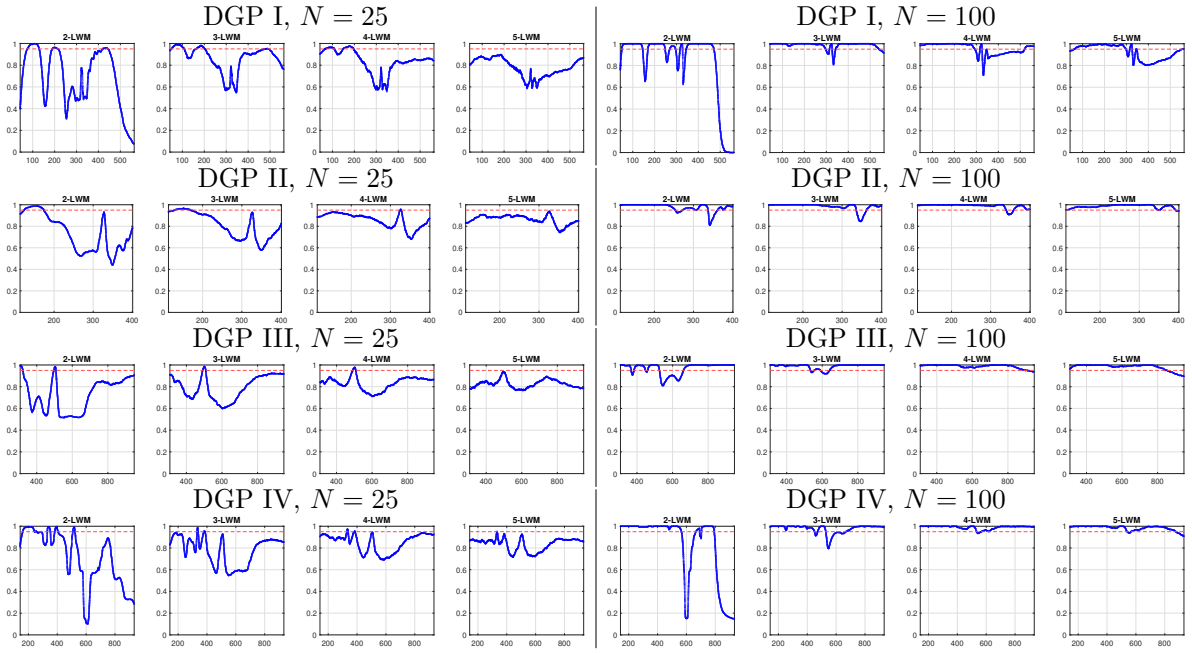

Figure F.1: 95% probability coverage plots for the M-LWM estimator with $M = 2$ to 5 for different DGPs under the baseline i.i.d option errors with a **large error variance (4× the estimated error variance of SPX options)**. The plots are computed pointwise based on 1,000 simulated paths. Red dashed horizontal line indicates the 95% threshold.

F.5 Additional Empirical Results

We present the RND estimates of the FTSE option prices before and after the Brexit using the PCA and the Sieve estimators with either recommended tuning parameters (PCA(h^*) and Sieve(J^*)) or tuning parameters adapted from the LWM-based RND estimates (PCA($\widetilde{h}$) and Sieve($\widetilde{J}$)) in Fig. F.2. The implementation details of these estimators are identical to our first example in Section 5, with the choices of the tuning parameters shown in the description of Fig. F.2.

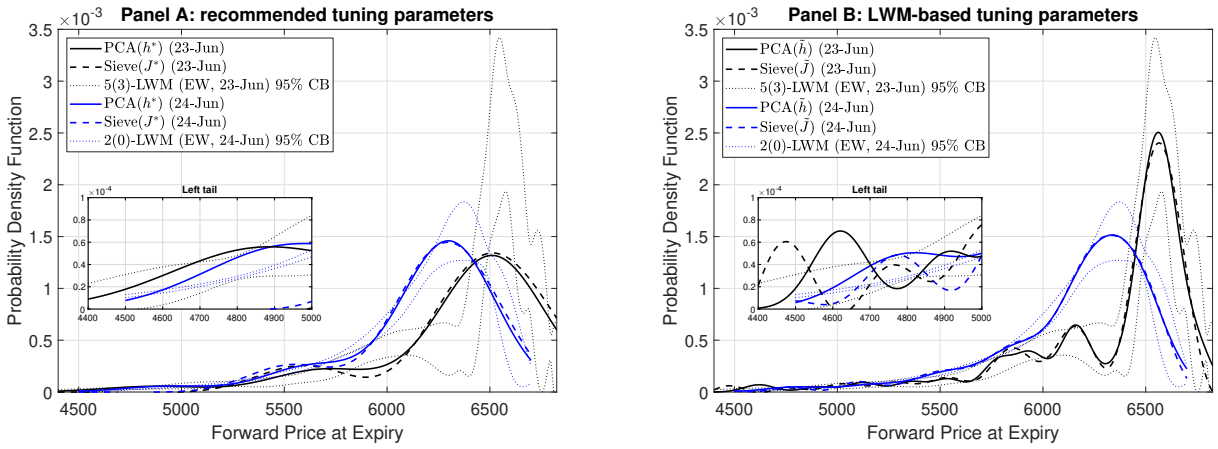

Figure F.2: RND estimates before and after the Brexit referendum based on the option prices in Fig. 5.5. In each figure, the black and blue dotted lines are the 95% confidence bands of the equally weighted LWM-based RND estimates. The left (resp. right) panel presents the RND estimates based on the PCA and the Sieve estimator based on recommended tuning parameters (resp. LWM-implied tuning parameters). On 23-Jun, the choices of the parameters are: $h^* = 243$ (23 parameters), $\widetilde{h} = 74$ (69 parameters), $J^* = 5$ with $\xi = 0$, $\widetilde{J} = 34$ with $\xi = 2.41 \times 10^{-5}$. On 24-Jun, the choices of the parameters are: $h^* = 220$ (23 parameters), $\widetilde{h} = 142$ (32 parameters), $J^* = 5$ with $\xi = 0$, $\widetilde{J} = 28$ with $\xi = 2.46 \times 10^{-5}$.

The findings in Fig. F.2 are largely consistent with our first example in Section 5. First, with the recommended tuning parameters as shown in Panel A, both the PCA and the sieve estimators produce over-smoothed RND estimates that are very different from the LWM-based estimates, which can be observed by comparing the estimates with the confidence bands of the LWM-based estimates. The difference is particularly striking on 23-Jun when the LWM-based RND estimate shows a concentrated dominating mode at around 6600, while both the PCA and the sieve estimator suggest much flatter RNDs. Interestingly, the Sieve-based RND estimates on both dates are bimodal, which is similar to the results in Fig. 5.2.

When switching to the tuning parameters implied by the LWM-fitted RND estimates as presented in Panel B, the resulting RND estimates are much closer to the LWM-based estimates. Both PCA($\widetilde{h}$) and Sieve($\widetilde{J}$) capture the dominating mode at around 6600 of the RND on 23-Jun, but the identification of the minor mode from the two estimators is less clear due to the noisy estimates of the left tail. Nevertheless, they are still contained within the confidence bands of the LWM-based RND estimates, which is in stark contrast to the result in Panel A. The unimodal RND estimates after the Brexit is also clearly recovered from Panel B, and the estimates from PCA($\widetilde{h}$) and Sieve($\widetilde{J}$) are again more in

line with the LWM-based confidence bounds compared to those estimates in Panel A. The noisy left tails from the two SNP methods are signs of potential overfitting, which calls for more refined methods to the choice of the tuning parameters.

To evaluate the goodness-of-fit of the RND estimates in Fig. F.2, we plot the corresponding model residuals in Fig. F.3. Similar to the findings in Section 5, the figure clearly shows that PCA(h^*) and Sieve(J^*) are not flexible enough to fully explain the observed option prices on both dates, as there are clear sinusoidal patterns in the residual series with very large SSE. Switching from recommended tuning parameters to the LWM-implied tuning parameters substantially improves the flexibility of the two methods, which drastically reduces the SSEs and improves the goodness-of-fit of the two models. The magnitude of the SSEs for PCA($\widetilde{h}$) and Sieve($\widetilde{J}$) are comparable with those in Table 5.1, confirming the improved quality of the RND estimates relative to those in Panel A. Overall, the conclusion is similar to our discussion in the second empirical example of Section 5 that, when compared with the PCA and the Sieve estimators with different choices of the tuning parameters, the LWM approach still produces a smooth RND estimates with a reasonable fit to the observed data.

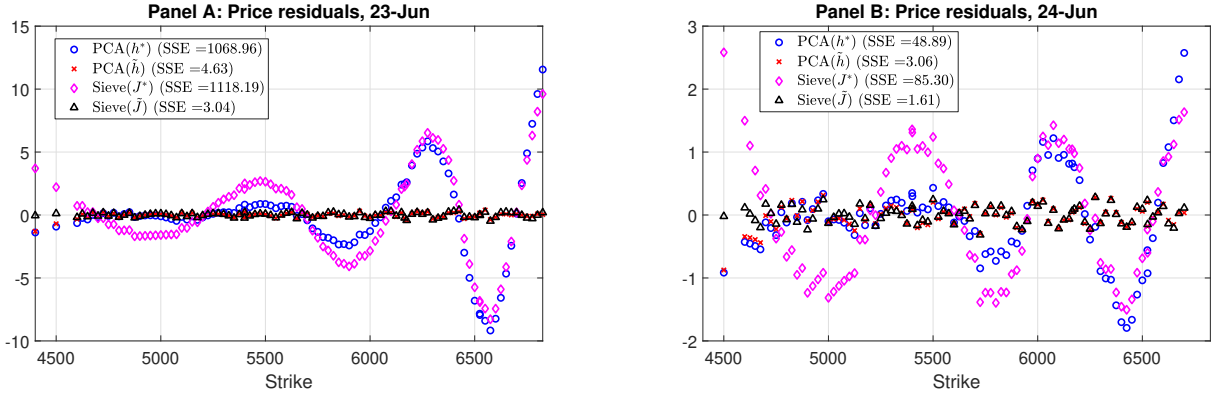

Figure F.3: Option price and IV residuals associated with the RND estimates in Figure F.2. For each estimator, the price residualsare defined as $O_n - \widehat{O}_n$, where O_n and $\widehat{O}_n$ denote the observed and the model-implied option prices at strike K_n, respectively. SSE in the legend represents the sum of squared error defined as $\sum_{n=1}^{N}(O_n - \widehat{O}_n)^2$.

References

Aït-Sahalia, Y. and Duarte, J. (2003). Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, 116(1-2):9–47.

Andersen, T. G., Fusari, N., and Todorov, V. (2015a). Parametric inference and dynamic state recovery from option panels. *Econometrica*, 83(3):1081–1145.

Andersen, T. G., Fusari, N., and Todorov, V. (2015b). The risk premia embedded in index options. *Journal of Financial Economics*, 117(3):558–584.

Andersen, T. G., Fusari, N., Todorov, V., and Varneskov, R. T. (2021). Spatial dependence in option observation errors. *Econometric Theory*, 37(2):205–247.

Andrews, D. W. K. (1988). Laws of Large Numbers for Dependent Non-Identically Distributed Random Variables. *Econometric Theory*, 4(3):458–467.

Black, F. (1976). The pricing of commodity contracts. *Journal of Financial Economics*, 3(1):167–179.

Bliss, R. R. and Panigirtzoglou, N. (2004). Option-Implied Risk Aversion Estimates. *The Journal of Finance*, 59(1):407–446.

Bondarenko, O. (2000). Recovering Risk-Neutral Densities: A New Nonparametric Approach. *Working Paper*.

Bondarenko, O. (2003). Estimation of risk-neutral densities using positive convolution approximation. *Journal of Econometrics*, 116(1-2):85–112.

Christoffersen, P., Fournier, M., and Jacobs, K. (2018). The Factor Structure in Equity Options. *The Review of Financial Studies*, 31(2):595–637.

Christoffersen, P., Jacobs, K., and Ornthanalai, C. (2012). Dynamic jump intensities and risk premiums: Evidence from s&p500 returns and options. *Journal of Financial Economics*, 106(3):447–472.

Domowitz, I. and White, H. (1982). Misspecified models with dependent observations. *Journal of Econometrics*, 20(1):35–58.

Fang, F. and Oosterlee, C. W. (2008). A Novel Pricing Method for European Options Based on Fourier-Cosine Series Expansions. *SIAM Journal on Scientific Computing*, 31(2):826–848.

Jackwerth, J. C. (2000). Recovering Risk Aversion from Option Prices and Realized Returns. *The Review of Financial Studies*, 13(2):433–451.

Jarrow, R. and Rudd, A. (1982). Approximate option valuation for arbitrary stochastic processes. *Journal of Financial Economics*, 10(3):347–369.

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

Kostakis, A., Mu, L., and Otsubo, Y. (2023). Detecting political event risk in the option market. *Journal of Banking & Finance*, 146:106624.

Lu, J. and Qu, Z. (2021). Sieve estimation of option-implied state price density. *Journal of Econometrics*, 224(1):88–112.

Orlando, G. and Taglialatela, G. (2017). A review on implied volatility calculation. *Journal of Computational and Applied Mathematics*, 320:202–220.

Pötscher, B. M. and Prucha, I. R. (1991). Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part II: asymptotic normality. *Econometric Reviews*, 10(3):253–325.

Singh, T. B. (2019). *Introduction to Topology*. Springer, Singapore.

White, H. (1980). Nonlinear Regression on Cross-Section Data. *Econometrica*, 48(3):721–746.

White, H. and Domowitz, I. (1984). Nonlinear Regression with Dependent Observations. *Econometrica*, 52(1):143–162.

Wooldridge, J. M. and White, H. (1988). Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes. *Econometric Theory*, 4(2):210–230.