Journal of Computational and Applied Mathematics 276 (2015) 129–142

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Reconstruction of conditional expectations from product moments with applications

Charles Hagwood

Statistical Engineering Division, National Institute of Standards and Technology, Gaithersburg, MD 20016, USA

ARTICLE INFO

Article history:
Received 29 May 2012
Received in revised form 12 August 2014

MSC:
44A60
46N30

Keywords:
Moment problem
Characterization
Conditional expectation
Vandermonde
Product moments

ABSTRACT

In this paper, it is shown under conditions associated with the moment problem that a sequence of product moments uniquely determines a conditional expectation of two random variables. Then, a numerical procedure is derived to reconstruct a conditional expectation in terms of a sequence of its product moments.

Published by Elsevier B.V.

1. Introduction

In this paper, a version of the moment problem is considered. It is shown, if U and W are jointly continuous random variables having a sequence of finite product moments $\alpha_j = E[UW^j]$, $j = 0, 1, 2, \ldots$ satisfying uniqueness conditions associated with a moment problem, then these moments uniquely determine $\psi(w) = E[U \mid W = w]$. Also, a numerical procedure is derived to recover $\psi(w)$ from the α_j. The classical moment problem asks, given a sequence of complex numbers $\alpha_0, \alpha_1, \alpha_2, \ldots,$ when does there exist a measure or a function of bounded variation with α_j its jth moment and when is it unique. There are at least three versions of the moment problem, the Hausdorff moment problem, the Stieltjes moment problem and the Hamburger moment problem, all differing by their supports. These classic problems gave rise to an abundance of mathematics, Stieltjes integrals, Pade approximations, orthogonal polynomials, Riesz–Markov theorem. A partial list of famous mathematicians involved in its solution are Chebyshev, Hausdorff, Riesz, Stieltjes, Krein, Markov, Hamburger, Nevanlinna, Akhiezer, Karlin. The texts and tracts [1–3] contain comprehensive coverage of the moment problem.

The Hamburger and Stieltjes moment problems are defined in terms of the Hankel matrices

$$H_n^{(1)} = \begin{bmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_{n+1} \\ \alpha_2 & \alpha_3 & \alpha_4 & \cdots & \alpha_{n+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_n & \alpha_{n+1} & \alpha_{n+2} & \cdots & \alpha_{2n} \end{bmatrix} \qquad H_n^{(2)} = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_{n+1} \\ \alpha_2 & \alpha_3 & \alpha_4 & \cdots & \alpha_{n+2} \\ \alpha_3 & \alpha_4 & \alpha_5 & \cdots & \alpha_{n+3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_{n+1} & \alpha_{n+2} & \alpha_{n+3} & \cdots & \alpha_{2n+2} \end{bmatrix}.$$

E-mail address: hagwood@nist.gov.

http://dx.doi.org/10.1016/j.cam.2014.08.018
0377-0427/Published by Elsevier B.V.

The Hamburger moment problem. Suppose an infinite sequence of complex numbers α_j $(j = 0, 1, 2, \ldots)$ is given. A necessary and sufficient condition for the existence of a function of bounded variation $\sigma(x)(-\infty < x < \infty)$ with moments α_j satisfying

$$\alpha_j = \int_{-\infty}^{\infty} x^j \sigma(dx) \quad j = 0, 1, 2, \ldots$$

is that the sequence of α_j is positive definite, i.e. $\sum_{j,k\geq 0} \alpha_{j+k} c_j c_k \geq 0$ for an arbitrary sequence $\{c_j,\ j \geq 0\}$, or equivalently $\det(H_n^{(1)}) > 0,\ n = 0, 1, 2, \ldots$. If in addition, there exist constants $C > 0$ and $R > 0$ such that

$$|\alpha_j| \leq CR^j j! \quad j = 1, 2, \ldots$$

then $\sigma(x)$ is unique.

Stieltjes moment problem. Suppose an infinite sequence of complex numbers α_j $(j = 0, 1, 2, \ldots)$ is given. Necessary and sufficient conditions for the existence of a function of bounded variation $\sigma(x)(0 < x < \infty)$ with moments α_j satisfying

$$\alpha_j = \int_{0}^{\infty} x^j \sigma(dx) \quad j = 0, 1, 2, \ldots$$

are that the sequences $\det(H_n^{(1)}) > 0$ and $\det(H_n^{(2)}) > 0,\ n = 0, 1, 2, \ldots$. The function $\sigma(x)$ is unique if there exist constants $C > 0$ and $R > 0$ such that

$$|\alpha_j| \leq CR^j (2j)! \quad j = 1, 2, \ldots.$$

Hausdorff moment problem. Suppose an infinite sequence of complex numbers α_j $(j = 0, 1, 2, \ldots)$ is given, then a necessary and sufficient condition for the existence of a function of bounded variation, $\sigma(x)$ with finite support $-\infty < a < x < b < \infty$ satisfying

$$\alpha_j = \int_{a}^{b} x^j \sigma(dx) \quad j = 0, 1, 2, \ldots$$

is that the sequence of moments $\alpha_j,\ j = 0, 1, 2, \ldots$ satisfy the inequalities

$$\Delta^k \alpha_j = \sum_{i=0}^{k} (-1)^i \binom{k}{i} \alpha_{i+j} \geq 0, \quad \text{for all } k, j = 0, 1, 2, \ldots. \tag{1}$$

If a solution exists, then it is unique.

There are distributions not uniquely determined by their moments. For example, Stieltjes [4,5] showed distributions with densities

$$a \exp(-t^{1/4}) \quad \text{and} \quad b_k t^{k-\log t} \quad t \in (0, \infty)$$

are not uniquely determined by their moments, where a, b_k are positive normalizing constants. Heyde [6] showed that the familiar lognormal distribution is not moment determinate. Other examples can be found in [7,8]. Common terminology calls a distribution uniquely determined by its moments M-determinate, if not, it is called M-indeterminate. The lognormal is an M-indeterminate distribution. If $F(w)$ is the distribution function of the random variable W that is M-indeterminate, then setting $U = F(W)$

$$E[U \mid W = w] = F(w)$$

makes $E[U \mid W = w]$ M-indeterminate in terms of $\sigma(x) = \int_0^x F(w)dF(w) = F^2(x)/2$. This follows from the Krein condition: $F(x)$ is M-indeterminate if $\int_{-\infty}^{\infty} -\ln f(x)/(1 + x^2)dx < \infty$, $f(x)$ the probability density of $F(x)$, $f(x) > 0$ for all x. Criteria other than those stated in the moment problems above can be used to determine if a distribution is M-determinate. There are the Carleman and Krein conditions reviewed in [9]. The criteria in the Krein condition is pleasing, because it depends entirely on the probability density function of $F(x)$.

Except for trivial text book examples, an exact expression for $E[U \mid W = w]$ is rare. The Gaussian case is an exception, i.e. when U and W are jointly normal, the conditional expectation is exactly known to be linear. For the Hausdorff case, it is shown in [10] how to recover a distribution function in terms of its moments. Recovery of a distribution function from moments for the Hausdorff moment problem can be found in [11–15]. We derive an expression for $E[U \mid W = w]$ in terms of $\alpha_j, j = 0, 1, 2, \ldots$. More precisely, we show in Theorem 2 under certain assumptions

$$E[U \mid W = w] = \lim_{n\to\infty} \frac{1}{p(w)} \sum_{j=0}^{2n} v_{[n(M_n+w)/M_n]j} \frac{n^{j+1}}{M_n^{j+1}} \alpha_j$$

where v_{ij} are entries in a Vandermonde matrix, $p(w)$ is the marginal density of W and $M_n/n \to 0$.

Numerous applications exist either in stochastic processes, economics, statistics, control theory and other branches of applied mathematics where an estimate for $E[U \mid W]$ is needed. Because conditional expectations are projections, they are important for deriving optimal estimators. This concept is particularly important in regression and filtering type problems. Applications in structural equations in economics, the Kalman–Bucy Filter in control theory and in regression in statistics can be found in [16–18]. Our solution, is applied to the errors in variables regression problem, see [19] for details on errors in variables.

2. Characterization

It will be assumed throughout that, (i) (U, W) is a continuous random vector with W having marginal distribution $F(w)$ with support either of the moment intervals $(-\infty, \infty)$ (Hamburger), $(0, \infty)$ (Stieltjes) or $[a, b]$ (Hausdorff), (ii) all the product moments $E[UW^j], j = 0, 1, 2, \ldots$ are finite and (iii) $F(w)$ is differentiable with derivative $p(w)$. The function, $E[U \mid W = w]$ is continuous for all w where $p(w) > 0$ and $E[U \mid W = w]$ is integrable with respect to $F(w)$ since $E[U] < \infty$.

Applying the law of total probability reveals the moment relationship between $E[U \mid W = w]$ and its product moments $E[UW^j]$

$$\alpha_j = E[UW^j] = \int_A^B w^j E[U \mid W = w] dF(w) \quad j = 0, 1, 2, \ldots \tag{2}$$

where the integration is over one of the moment intervals and whenever an integral sign of this form appears, it is so to be interpreted.

Under assumptions (i)–(iii), it is shown that $E[U \mid W = w]$ is the only function of w integrable with respect to $F(dw)$ with moments, $E[UW^j]$.

Theorem 1 (*Characterization Theorem*)**.** *If $\psi(w)$ is any continuous, Riemann–Stieltjes integrable function with respect to $F(w)$ over the moment interval defined by A and B,*

$$\int_A^B w^j \psi(w) dF(w) = \alpha_j \quad j = 0, 1, 2, \ldots \tag{3}$$

and the sequence $\alpha_j, j = 0, 1, 2, \ldots$ satisfies the uniqueness conditions associated with the moment problem for the limits of integration, then $\psi(w) = E[U \mid W = w]$. That is, the product moments $E[UW^j], j = 0, 1, 2, \ldots$ uniquely determine $E[U \mid W = w]$.

Proof. Since $\psi(w)$ is integrable with respect to $F(w)$, then

$$\sigma(x) = \int_A^x \psi(w) dF(w) \quad A < x < B \tag{4}$$

is of bounded variation, see [20]. From (2), the function of bounded variation $\int_A^x E[U \mid W = w]dF(w)$ has moments α_j. By the uniqueness property in the moment problem, $\sigma(x) \equiv \int_A^x E[U \mid W = w]dF(w)$. Thus, for each $A < x < B$ and $\delta > 0$

$$\frac{1}{\delta}\int_x^{x+\delta} \psi(w)dF(w) = \frac{1}{\delta}\int_x^{x+\delta} E[U|W = w]dF(w).$$

Using the continuity of both $\psi(w)$ and $E[U \mid W = w]$ and letting δ go to zero gives $\psi(w) = E[U \mid W = w]$.

3. Reconstruction

Next, it is shown how to recover $\psi(w) = E[U \mid W = w]$ from its product moments, i.e. $\psi(w)$ is expressed in terms of $\alpha_j = E[UW^j]$. Let M_n be a sequence of positive real numbers increasing to infinity, with M_n/n approaching zero (specified exactly in Theorem 2), then

$$\psi(w) = \lim_{n\to\infty} \frac{1}{p(w)} \sum_{j=0}^{2n} v_{[n(M_n+w)/M_n],j} \frac{n^{j+1}}{M_n^{j+1}} \alpha_j \tag{5}$$

where [ ] denotes the greatest integer function, $p(w) = F'(w)$ and v_{ij} is the ijth entry of the inverse of V_{2n} (set $v_{i,j} = 0$ if i or $j > 2n$), the Vandermonde matrix with nodes at $m_j = -n + j, j = 0, 1, \ldots, 2n$

$$V_{2n} = V(-n, -n+1, \ldots, n) = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ -n & -n+1 & \cdots & n \\ (-n)^2 & (-n+1)^2 & \cdots & n^2 \\ \vdots & \vdots & \vdots & \vdots \\ (-n)^{2n} & (-n+1)^{2n} & \cdots & n^{2n} \end{pmatrix}. \tag{6}$$

The convention in this paper is to allow the indices of matrices to start at zero rather than 1, i.e. the first row and column are denoted by $v_{0,j}, v_{i,0}$.

Explicit formulae and computational schemes for solving Vandermonde systems and finding the inverse of its matrix exist. The most celebrated is by Björck and Pereyra [21]. The high accuracy of its solutions has been justified theoretically in [22]. Other results on the inverse of Vandermonde matrices can be found in [23,24]. Ref. [25] contains a short algorithm for solving a moment system of equations.

We need the following lemmas to validate our reconstruction. The proof of Lemma 1 is a very close adaptation of the proof of Theorem 1 from [26]. Let σ_m be the mth elementary symmetric function in the variables $x_1, \ldots, x_N$, $N \geq 1$

$$\sigma_m = \sigma_m(x_1, \ldots, x_m) = \sum x_{\nu_1} x_{\nu_2} \cdots x_{\nu_m} \quad (1 \leq m \leq N) \tag{7}$$

and let $\sigma_m^\lambda = \sigma_m(x_1, \ldots, x_{\lambda-1}, x_{\lambda+1}, \ldots, x_m)$.

Lemma 1. *Let $V_N = V_N(x_1, \ldots, x_N)$ be the Vandermonde matrix with nodes at $x_1, \ldots, x_N$ and let $V_N^{-1} = (v_{\lambda\mu})$, $1 \leq \lambda, \mu \leq N$ denote its inverse. Assume $x_\nu \neq x_\mu$ for $\nu \neq \mu$. The following holds:*

$$\sum_{\mu=0}^{N} |v_{\lambda\mu}|\, |x|^\mu \leq \frac{\prod\limits_{\nu\neq\lambda}^{N} (|x| + |x_\nu|)}{\prod\limits_{\nu\neq\lambda}^{N} |x_\nu - x_\lambda|} 1 \leq \lambda \leq N, \quad \textit{for all } x. \tag{8}$$

Proof. First we show

$$\sum_{m=1}^{N} |\sigma_m|\, |x|^{N-m} \leq \prod_{\nu=1}^{N} (|x| + |x_\nu|). \tag{9}$$

Let $q(x) = \prod_{\nu=1}^{N} (x - x_\nu)$. Then

$$q(x) = \sum_{m=0}^{N} (-1)^m \sigma_m x^{N-m}.$$

In particular,

$$q(-x) = (-1)^N \sum_{m=0}^{N} \sigma_m x^{N-m}. \tag{10}$$

On the other hand, by definition

$$q(-x) = (-1)^N \prod_{\nu=1}^{N} (x + x_\nu). \tag{11}$$

If $x_\nu \geq 0$, $1 \leq \nu \leq N$, then all $\sigma_m \geq 0$ and form (10) and (11), we have

$$\sum_{m=0}^{N} |\sigma_m|\, |x|^{N-m} = \sum_{m=0}^{N} \sigma_m |x|^{N-m} = (-1)^N q(-|x|) = \prod_{\nu=1}^{N} (|x| + x_\mu) = \prod_{\nu=1}^{N} (|x| + |x_\nu|).$$

So, (9) holds, exactly, for nonnegative x_ν. Using the definition (7)

$$|\sigma_m(x_1, \ldots, x_m)| \leq \sigma_m(|x_{\nu_1}|, \ldots, |x_{\nu_m}|) \quad 1 \leq m \leq N.$$

We find using the case of positive x_ν that

$$\sum_{m=0}^{N} |\sigma_m(x_1, \ldots, x_m)|\, |x|^{n-m} \leq \prod_{\nu=1}^{N} (|x| + |x_\nu|) \tag{12}$$

proving (9) holds for all x_ν. It is shown in [27, p. 416] that

$$v_{\lambda\mu} = (-1)^{N-\mu} \frac{\sigma_{N-\mu}^\lambda}{\prod\limits_{\nu\neq\lambda} (x_\nu - x_\lambda)}. \tag{13}$$

Therefore,

$$\sum_{m=0}^{N} |v_{\lambda\mu}|\, |x|^\mu = \frac{\sum\limits_{\mu=1}^{N} |\sigma_{N-\mu}^\lambda|\, |x|^\mu}{\prod\limits_{\nu\neq\lambda} |x_\nu - x_\lambda|}.$$

Lemma 1 follows from (9).

Lemma 2. *For $1 \le i \le 2n$, the following holds:*

$$\frac{\prod_{j\neq i}^{2n}(n+|-n+j|)}{\prod_{j\neq i}^{2n}|i-j|} \le C_n \sim \left(4^n\sqrt{\frac{n}{\pi}}\right)^2. \tag{14}$$

Proof. Here we will use the fact that the largest binomial coefficient is the central binomial coefficient and its asymptotic expansion given in [28].

$$\frac{\prod_{j\neq i}^{2n}(n+|-n+j|)}{\prod_{j\neq i}^{2n}|i-j|} \le \frac{\prod_{j=1}^{n}(2n-j)\prod_{j=n+1}^{2n}j}{\prod_{j\neq i}^{2n}|i-j|} = \frac{2\left(\prod_{j=1}^{n}2n-j\right)^2}{\prod_{j\neq i}^{2n}|i-j|} \tag{15}$$

$$= \frac{(2n!/(n-1)!)^2}{i!(2n-i)!} = \frac{2n!}{((n-1)!)^2}\binom{2n}{i} = n^2\binom{2n}{n}\binom{2n}{i} \le \left(n\binom{2n}{n}\right)^2 \sim \left(4^n\sqrt{\frac{n}{\pi}}\right)^2. \tag{16}$$

Theorem 2. *Let $\psi(w)$ have support $(-\infty,\infty)$. If there exists a sequence of positive constants M_n satisfying $M_n/n \to 0$ and*

$$\left(\frac{4}{M_n}\right)^{2n+1}\int_{|u|>M_n} u^{2n}d\sigma(u) \to 0 \quad \text{as } n\to\infty \tag{17}$$

where $\sigma(u) = \int_{-\infty}^{u}\psi(x)dF(x)$ and a neighborhood $\mathcal{N}$ where $p(w)$ is nonzero for $w \in \mathcal{N}$, then the following holds:

$$\psi(w) = \lim_{n\to\infty}\frac{1}{p(w)}\sum_{j=0}^{2n} v_{[n(M_n+w)/M_n],j}\frac{n^{j+1}}{M_n^{j+1}}\alpha_j \quad w \in \mathcal{N}. \tag{18}$$

Proof. Set $\psi_n(w)$ equal to

$$\psi_n(w) = \frac{1}{p(w)}\sum_{j=0}^{2n} v_{[n(M_n+w)/M_n],j}\frac{n^{j+1}}{M^{j+1}}\alpha_j \quad w \in \mathcal{N}. \tag{19}$$

To show pointwise convergence at $w \in (-\infty,\infty)$, fix $\epsilon > 0$ and substitute the expression for α_j in (19),

$$\begin{aligned}\psi_n(w) &= \frac{1}{p(w)}\sum_{j=0}^{2n} v_{[n(M_n+w)/M_n],j}\frac{n^{j+1}}{M_n^{j+1}}\int_{-\infty}^{\infty}u^j d\sigma(u)\\ &= \frac{1}{p(w)}\sum_{j=0}^{2n} v_{[n(M_n+w)/M_n],j}\frac{n^{j+1}}{M_n^{j+1}}\int_{-M_n}^{M_n}u^j d\sigma(u)\\ &\quad + \frac{1}{p(w)}\sum_{j=0}^{2n} v_{[n(M_n+w)/M_n],j}\frac{n^{j+1}}{M_n^{j+1}}\int_{|u|>M_n}u^j d\sigma(u) = \psi_{1n}(w) + \psi_{2n}(w).\end{aligned} \tag{20}$$

For each $n > 0$, divide $[-M_n, M_n]$ into subintervals with endpoints contained in $\mathcal{E}_n = \{w_j = w_{jn} = \frac{M_n}{n}(-n+j), j = 0, 1, \dots, 2n\}$. For w fixed in $\mathcal{N}$ and $\epsilon, \zeta > 0$, we may choose N_1 large enough such that for $n > N_1$,

(i) There exists, $w_{i_n} \in \mathcal{E}_n \cap \mathcal{N}$, $w_{i_n} \le w$ with $w - w_{i_n} < M_n/2n$, which implies $[n(M_n+w)/M_n] = [n + (n/M_n)w_{i_n} + (n/M_n)(w - w_{i_n})] = [i_n + (n/M_n)(w - w_{i_n})] = i_n$.

(ii) $p(w_{i_n}) > 0$.

(iii) $|\frac{p(w_{i_n})}{p(w)} - 1| < \zeta$.

(iv) $|\psi(w) - \psi(w_{i_n})| < \epsilon/4$.

Therefore for $n > N_1$,

$$\begin{aligned}
|\psi_n(w) - \psi(w)| &= |\psi_n(w_{i_n}) - \psi(w) + \psi_n(w) - \psi_n(w_{i_n})| \\
&= \left|\psi_n(w_{i_n}) - \psi(w) + \left[\frac{1}{p(w)} - \frac{1}{p(w_{i_n})}\right] \sum_{j=1}^{2n} v_{i_n j}\alpha_j \frac{n^{j+1}}{M_n^{j+1}}\right| \\
&= \left|\psi_n(w_{i_n}) - \psi(w) + \psi_n(w_{i_n})\left(1 - \frac{p(w_{i_n})}{p(w)}\right)\right| \\
&= \left|\psi_n(w_{i_n}) + \psi_n(w_{i_n})\left(1 - \frac{p(w_{i_n})}{p(w)}\right) - [\psi(w) - \psi(w_{i_n}) + \psi(w_{i_n})]\right| \\
&= \left|\psi_n(w_{i_n}) - \psi(w_{i_n}) + \psi_n(w_{i_n})\left(1 - \frac{p(w_{i_n})}{p(w)}\right) - [\psi(w) - \psi(w_{i_n})]\right| \\
&\le \zeta|\psi_n(w_{i_n})| + \epsilon/4 + |\psi_n(w_{i_n}) - \psi(w_{i_n})|.
\end{aligned} \tag{21}$$

To show convergence, it suffices to consider, $|\psi_n(w_{i_n}) - \psi(w_{i_n})|$. Set $i = i_n$, then

$$\psi_{1n}(w_i) = \frac{1}{p(w_i)} \sum_{j=0}^{2n} v_{ij} \frac{n^{j+1}}{M_n^{j+1}} \sum_{k=0}^{2n-1} \int_{w_k}^{w_{k+1}} u^j d\sigma(u). \tag{22}$$

The Riemann–Stieltjes integral may be bounded by the maximum and the minimum of u^j in the interval $[w_k, w_{k+1}]$ giving

$$\frac{1}{p(w_i)} \sum_{j=0}^{2n} v_{ij} \frac{n^{j+1}}{M_n^{j+1}} \sum_{k=0}^{2n-1} w_k^j \Delta_+(w_k) \le \psi_{1n}(w_i) \le \frac{1}{p(w_i)} \sum_{j=0}^{2n} v_{ij} \frac{n^{j+1}}{M_n^{j+1}} \sum_{k=0}^{2n-1} w_{k+1}^j \Delta_-(w_{k+1}) \tag{23}$$

or equivalently

$$\frac{1}{p(w_i)} \sum_{k=0}^{2n-1} (n/M_n)\Delta_+(w_k) \sum_{j=0}^{2n} v_{ij} \frac{n^j}{M_n^j} w_k^j \le \psi_{1n}(w_i) \le \frac{1}{p(w_i)} \sum_{k=0}^{2n-1} (n/M_n)\Delta_-(w_{k+1}) \sum_{j=0}^{2n} v_{ij} \frac{n^j}{M_n^j} w_{k+1}^j \tag{24}$$

where $\Delta_+(w_k) = \sigma(w_{k+1}) - \sigma(w_k) = \sigma(w_k + M/n) - \sigma(w_k)$ and $\Delta_-(w_{k+1}) = \sigma(w_{k+1}) - \sigma(w_k) = \sigma(w_{k+1}) - \sigma(w_{k+1} - M/n)$.

It was stated in the proof of Lemma 1, that the Vandermonde matrix is related to the Lagrange interpolation polynomial basis function by

$$q_i(m) = \prod_{\substack{j=0 \\ j \ne i}}^{2n} \frac{m - m_j}{m_i - m_j} = \sum_{j=0}^{2n} v_{ij} m^j \quad i = 0, 1, \dots, 2n \; m \in (-\infty, \infty). \tag{25}$$

Obviously, $q_i(m_k) = \sum_{j=0}^{2n} v_{ij} m_k^j = \delta_{ik}$, $0 \le k \le 2n$. Therefore, Eq. (24) becomes

$$\begin{aligned}
&\frac{1}{p(w_i)} \sum_{k=0}^{2n-1} (n/M_n)\Delta_+(w_k)\delta_{ik} \le \psi_{1n}(w_i) \le \frac{1}{p(w_i)} \sum_{k=0}^{2n-1} (n/M_n)\Delta_-(w_k)\delta_{ik+1} \\
&(n/M_n)\frac{\sigma(w_i + M_n/n) - \sigma(w_i)}{p(w_i)} \le \psi_{1n}(w_i) \le (n/M_n)\frac{\sigma(w_i) - \sigma(w_i - M_n/n)}{p(w_i)}.
\end{aligned} \tag{26}$$

This inequality implies

$$\begin{aligned}
-\psi(w_i) + \frac{n/M_n}{p(w_i)} \int_{w_i}^{w_i + M_n/n} \psi(u)p(u)du &\le \psi_{1n}(w_i) - \psi(w_i) \\
&\le -\psi(w_i) + \frac{n/M_n}{p(w_i)} \int_{w_i - M_n/n}^{w_i} \psi(u)p(u)du.
\end{aligned}$$

Now choose $N_2 > N_1$, such that for all $n > N_2$

$$\max\left\{\left|\psi(w_i) - \frac{n/M_n}{p(w_i)} \int_{w_i - M_n/n}^{w_i} \psi(u)p(u)du\right|, \left|\psi(w_i) - \frac{n/M_n}{p(w_i)} \int_{w_i}^{w_i + M_n/n} \psi(u)p(u)du\right|\right\} < \epsilon/4.$$

Therefore for all $n > N_2$

$$|\psi_{1n}(w_i) - \psi(w_i)| < \epsilon/4.$$

Applying Lemmas 1 and 2,

$$\begin{aligned}
|\psi_{2n}(w_i)| &\le \frac{1}{p(w_i)} \sum_{j=0}^{2n} |v_{ij}| \frac{n^{j+1}}{M_n^{j+1}} M_n^{j-2n} \int_{|u|>M_n} |u|^{2n} d\sigma(u) \\
&\le \frac{1}{p(w_i)} \frac{n}{M_n} M_n^{-2n} \int_{|u|>M_n} u^{2n} d\sigma(u) \sum_{j=0}^{2n} |v_{ij}| n^j \\
&\le \frac{1}{p(w_i)} \frac{n^2}{M_n} M_n^{-2n} \frac{4^{2n}}{\pi} \int_{|u|>M_n} u^{2n} d\sigma(u) \\
&\le C_0 \frac{4^{2n+1}}{M_n^{2n+1}} \int_{|u|>M_n} u^{2n} d\sigma(u) \\
&\le \epsilon/4
\end{aligned} \tag{27}$$

for all $n > N_3$ for some $N_3 > N_2$ by hypothesis. Therefore, or $n > N_3$,

$$|\psi_n(w_{i_n}) - \psi(w_{i_n})| \le \epsilon/2 \tag{28}$$

and by (28) $\psi_n(w_{i_n})$ is bounded. Therefore in (21), ζ may be chosen such that $\zeta|\psi_n(w_{i_n})| < \epsilon/4$, giving

$$|\psi_n(w) - \psi(w)| < \epsilon$$

as desired to show pointwise convergence.

Theorem 3. *Let $\psi(w)$ have support $(0, \infty)$. If there exists a sequence of positive constants M_n satisfying $M_n/n \to 0$ and*

$$\left(\frac{4}{M_n}\right)^{2n+1} \int_{u>M_n} u^{2n} d\sigma(u) \to 0 \quad \text{as } n \to \infty \tag{29}$$

where $\sigma(u) = \int_0^u \psi(x) dF(x)$, $u > 0$ and a neighborhood $\mathcal{N}$ where $p(w)$ is nonzero for $w \in \mathcal{N}$, then the following holds:

$$\psi(w) = \lim_{n\to\infty} \frac{1}{p(w)} \sum_{j=0}^{2n} v_{[nw/M_n],j} \frac{n^{j+1}}{M_n^{j+1}} \alpha_j \quad w \in \mathcal{N} \tag{30}$$

where v_{ij} is the i, j entry of the inverse of the Vandermonde matrix $V(0, 1, 2, \ldots, n)$.

Proof. The proof is almost identical to the proof of Theorem 2.

Theorem 4. *Let $\psi(w)$ have support (A, B), A and B finite. Let $p, \psi \in C^{2n+2}[A, B]$, i.e. p, ψ have $2n + 2$ continuous derivatives in $[A, B]$ and let*

$$\psi_n(w_i) = \frac{2n}{(B-A)\pi} \sum_{j=1}^{n} v_{ij}\alpha_j \Bigg/ \left(p(w_i) \sqrt{1 - \left(\frac{2w_i - B - A}{B - A}\right)^2} \right)$$

where w_i are the Chebyshev nodes in $[A, B]$

$$w_i = \frac{B-A}{2} \cos\left(\frac{2i-1}{2n}\pi\right) + \frac{B+A}{2}$$

and let v_{ij} denote the ijth entry in the inverse of $V(w_1, w_2, \ldots, w_n)$. For $1 \le i \le n$ and n sufficiently large

$$|\psi_n(w_i) - \psi(w_i)| \le \frac{R_n}{(2n)!2^{2n+1}} 3^{3/4} (1 + \sqrt{2})^n \left(\frac{B-A}{2}\right)^{2n+1} \Bigg/ \left(p(w_i) \sqrt{1 - \left(\frac{2w_i - B - A}{B - A}\right)^2} \right) \tag{31}$$

$R_n = 2\pi \sup_{1\le j\le n} |f_j^{(2n)}(\xi_j)|$, $A < \xi_j < B$ *and* $f_j(w) = w^j \psi(w) p(w) \sqrt{1 - (\frac{2w-B-A}{B-A})^2}$.

Proof. The Gauss–Chebyshev quadrature approximation, [29] gives

$$\alpha_j = \int_A^B w^j \psi(w) p(w) dw = \frac{B-A}{2n} \pi \sum_{i=1}^{n} w_i^j \psi(w_i) p(w_i) \sqrt{1 - \left(\frac{2w_i - B - A}{B - A}\right)^2} + R_{j,n}$$

for $j = 1, \dots, n$ where $R_{j,n} = \frac{2\pi f_j^{(2n)}(\xi_j)\left(\frac{B-A}{2}\right)^{2n+1}}{2^{2n+1}(2n)!}$, $A < \xi_j < B$ and $f_j(w) = w^j\psi(w)p(w)\sqrt{1-(\frac{2w-B-A}{B-A})^2}$. It suffices to prove the result for $[A, B] = [-1, 1]$.

$$\begin{aligned}
\psi_n(w_i)p(w_i)\sqrt{1-w_i^2} &= \frac{n}{\pi}\sum_{j=1}^{n} v_{ij}\alpha_j = \frac{n}{\pi}\sum_{j=1}^{n} v_{ij}\int_{-1}^{1} w^j\psi(w)p(w)dw \\
&= \frac{n}{\pi}\sum_{j=1}^{n} v_{ij}\left[\frac{\pi}{n}\sum_{k=1}^{n} w_k^j\psi(w_k)p(w_k)\sqrt{1-w_k^2} + R_{j,n}\right] \\
&= \sum_{k=1}^{n} w_k^j\psi(w_k)p(w_k)\sqrt{1-w_k^2}\sum_{j=1}^{n} v_{ij}w_k^j + \frac{n}{\pi}\sum_{j=1}^{n} v_{ij}R_{j,n}.
\end{aligned}$$

Now we may apply the result following (25) in Theorem 2, to get for n sufficiently large

$$\begin{aligned}
|\psi_n(w_i) - \psi(w_i)| &\le \sup_j R_{j,n}\frac{n}{\pi}\|V^{-1}\|_\infty(p(w_i)\sqrt{1-w_i^2})^{-1} \\
&\le \frac{R_n}{(2n)!2^{2n+1}}3^{3/4}(1+\sqrt{2})^n(p(w_i)\sqrt{1-w_i^2})^{-1} \qquad (32)
\end{aligned}$$

$R_n = \sup_j |f_j^{(2n)}(\xi_j)|$ and the estimate in (32) for the infinity norm for the inverse of a Vandermonde matrix can be found in [26].

Illustration: Assume U, W have a bivariate distribution with

$$\psi(w)p(w) \sim w^s e^{-\beta w^t} \quad s, \beta > 0, t > 1\ w \in (0, \infty).$$

This incorporates the symmetric standard normal distribution. Here M_n is found so that the condition (17)

$$\left(\frac{4}{M_n}\right)^{2n+1}\int_{w>M_n} w^{2n}d\sigma(w) = \left(\frac{4}{M_n}\right)^{2n+1}\int_{w>M_n} w^{2n}\psi(w)p(w)dw = \left(\frac{4}{M_n}\right)^{2n+1} I_n \to 0$$

is satisfied. Using formula 9 from Section 3.381 of Gradshteyn and Ryzhik [30]

$$I_n = \int_{M_n}^{\infty} w^{2n+s}e^{-\beta w^t}dw = \frac{\Gamma\left(\frac{2n+s+1}{t}, \beta M_n^t\right)}{t\beta^{(2n+s+1)/t}} = \frac{\Gamma(k_n+1, r_n)}{t\beta^{(2n+s+1)/t}} \qquad (33)$$

$$k_n = \frac{2n+s+1}{t} - 1 \qquad (34)$$

$$r_n = \beta M_n^t \qquad (35)$$

where $\Gamma(a, x) = \int_x^\infty e^{-t}t^{a-1}dt$ denotes the incomplete gamma function. By a derivation similar to the derivation of the asymptotic expansion 8.11.12 in the NIST Handbook of Mathematical Functions [31]

$$\Gamma(k_n+1, r_n) = r_n^{k_n}e^{-r_n}\sqrt{2\pi r_n}\Big(1 + \frac{1}{12r_n} + \frac{1}{288r_n^2} + \cdots + \cdots\Big). \qquad (36)$$

Let

$$\tilde{I}_n = \left(\frac{4}{M_n}\right)^{2n+1}\frac{r_n^{k_n}e^{-r_n}\sqrt{2\pi r_n}}{t\beta^{(2n+s+1)/t}} = C4^{2n}M_n^{s-t/2}e^{-\beta M_n^t}.$$

Giving

$$\log(\tilde{I}_n) = \log(C) + 2n\log(4) + (s - t/2)\log(M_n) - \beta M_n^t.$$

Take $M_n = n^\epsilon$ where $1/t < \epsilon < 1$, then $M_n/n \to 0$ and $\log(\tilde{I}_n) \to -\infty$.

4. Implementation, examples and caveats

In this section, the reconstruction of $\psi(w) = E[X \mid W = w]$ using product moments is implemented. This is performed with a known finite sequence of cross product moments $\alpha_j = E[XW^j]$, $j = 1, 2, \dots, n$ and when the α_j's are estimated using sample cross product moments. Also, in this section, the reconstruction is applied to the errors in variables problem.

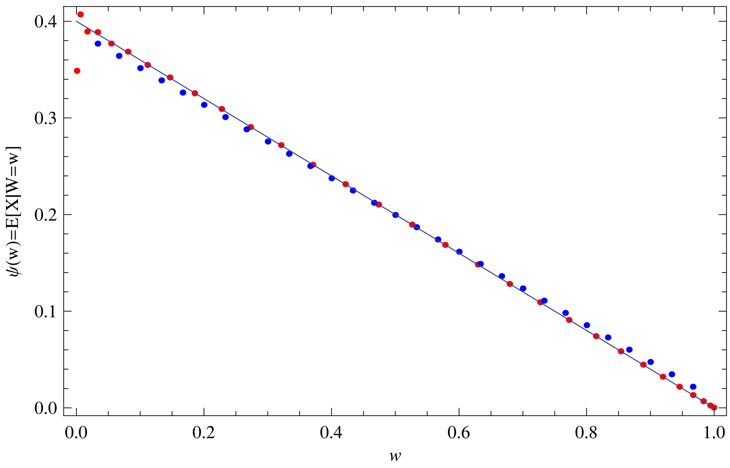

Fig. 1. Comparison with Mnatsakanov's method. The red plotmarkers represent $\psi_n(w)$ from this paper and the blue plotmarkers represent $\bar{\psi}_n(w)$ from Mnatsakanov's paper [15]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Implementation: The product moments can be obtained several ways. If the joint distribution, $f(x, w)$, of X and W is known, then by integration the moments $\alpha_j, j = 1, 2, \ldots$ may be obtained. If one observes $(X_i, W_i), i = 1, 2, \ldots, N$ from the distribution $f(x, w)$, then the sample product moments may be estimated by

$$\hat{\alpha}_j = \frac{1}{N}\sum_{i=1}^{N} X_i W_i^j.$$

In all the examples, the Vandermonde matrices were inverted using the Björck and Pereyra algorithm [21], see p. 187 in the book by Golub and Van Loan [25] for the algorithm used.

Example 1. First the product moments reconstruction is compared to some well known work by Mnatsakanov, in [15,32, 33]. Mnatsakanov used a moment procedure based on a Bernstein polynomial basis to recover probability densities and distribution functions. See Feller, [10] for a discussion of the Bernstein polynomials. Mnatsakanov's results may be used to obtain an estimate of ψ. Consider the case where ψ is nonnegative and (X, W) has a bivariate Dirichlet joint density with parameters $a_1 = 1/3, a_2 = 1$ and $a_3 = 1/2$

$$f(x, w) = \frac{\Gamma(a_1 + a_2 + a_3)}{\Gamma(a_1)\Gamma(a_2)\Gamma(a_3)} x^{a_1 - 1} w^{a_2 - 1}(1 - x - w)^{a_3 - 1}$$

where $x > 0, w > 0$ and $x + w < 1$. The marginal densities of X and W have beta distributions, Beta$(a_1, a_2 + a_3)$, Beta$(a_2, a_1 + a_3)$, respectively and $\psi(w) = E[X \mid W = w] = a_1(1 - w)/(a_1 + a_3)$. The product moments are given by $\alpha_j = E[XW^j] = (a_1)_1(a_2)_j/(a_1 + a_2 + a_3)_{1+j}$, where $(a)_k$ represents the Pochhammer symbol. The approximation for $E[X \mid W = w]$ from Eq. (5) of Mnatsakanov [15] is

$$\bar{\psi}_n(w) = \frac{\Gamma(n+2)}{\Gamma([nw]+1)} \sum_{m=0}^{n-[nw]} \frac{(-1)^m \alpha_{m+[nw]}}{m!(n - [nw] - m)!} \quad w \in [0, 1].$$

A graphical comparison of $\bar{\psi}_n(w)$ with the product moments estimate $\psi_n(w)$ gotten using Theorem 4 is shown in Fig. 1. In this example, $A = 0$ and $B = 1$ and we set $n = 35$.

The plots of $\psi_n(w)$ and $\bar{\psi}_n(w)$ are similar to each other. Further investigation is required to make this comparison more accurate.

Example 2. In the Hamburger case, let us consider the regression problem with the data $(W_i, Y_i), i = 1, 2, \ldots, N$,

$$Y_i = f(W_i) + \epsilon_i \quad i = 1, 2, \ldots, N.$$

Assume $f(x) = (x - 3.5)^2, x \in (2, 5)$, W_i's i.i.d. random variables distributed according to Normal (3.5, 1) and the ϵ_i's distributed according to Normal $(0, 0.5^2)$, with the W_i's and ϵ_i's independent. Suppose we seek to estimate the regression function, $\psi(w) = E[Y \mid W = w] = f(w)$. See Efron, [34] for an interesting application.

Each $\alpha_j = E[YW^j]$ is estimated using the sample moment

$$\hat{\alpha}_j = \frac{1}{N}\sum_{i=1}^{N} Y_i W_i^j.$$

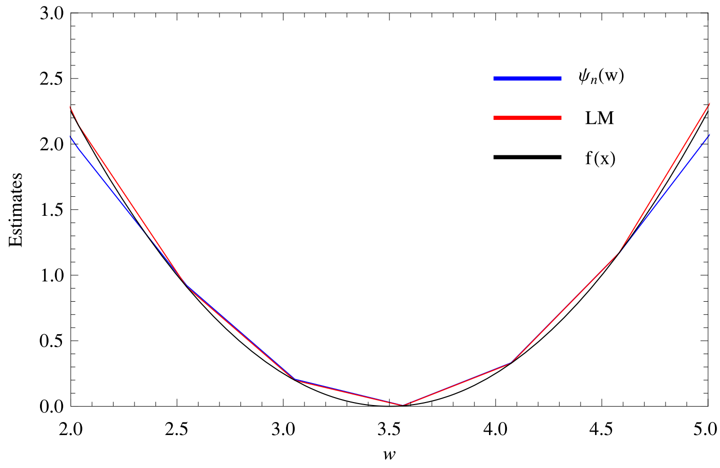

Fig. 2. Plot of ψ_n versus a Linear Models (LM) estimate.

We simulated (W_i, Y_i), $i = 1, 2, \ldots, N = 100$, formed $\hat{\alpha}_j$, $j = 1, \ldots, n$ and created $\psi_n(w)$ using Eq. (19), 1000 times. Then, the mean of the $\psi_n(w)$'s over all 1000 simulations was computed as our estimate of $f(w)$. When using (19), we took $n = 55$ and $M = 28$ and used the kernel density, "KernelMixtureDistribution" from Mathematica. Also, an estimate of $f(x)$ was obtained using least squares assuming the linear model $Y = \beta_0 + \beta_1 W + \beta_2 W^2 + error$ and using the same simulation procedure. Both estimates of $f(x)$ are plotted in Fig. 2, along with the true function $f(x) = (x-3.5)^2$. These plots are in close agreement. Also, the percent error, $N^{-1}\sum_{i=1}^{N} |\psi_n(w_i) - f(w_i)|/f(w_i)$ for each simulation was calculated and averaged over all 1000 simulation replicates giving a value of 1.999. A similar average percent error was performed for the least squares estimate giving a value 2.06. The estimates are very similar.

Example 3. Assume again, one wants to estimate the curve $f(x)$ from data (X_i, Y_i), $Y_i = f(X_i) + \epsilon_i$. In this example, assume the true X_i cannot be observed directly, but with noise, say, $W_i = X_i + U_i$ is observed rather than X_i, where the noise U_i satisfies $E[U_i] = 0$. These errors occur in timing instruments, dosimeters or other instruments where the independent variable cannot be measured exactly. For example, sampled waveforms are known to suffer from timing jitter errors, see [35] and in medicine, dosimeters for measuring dosage exposure to radiation are subject to errors, e.g., in [36]. This problem is common in statistics and is called the errors in variables problem by the authors in [19] and in [17].

The errors in variables problem is not uniquely solvable, i.e. identifiable, without additional information about the distribution of the noise, U. For normal data with mean zero and variance σ_U^2, knowledge of σ_U^2 is sufficient to determine U via its moment. Indeed, when r is odd $E[U^r] = 0$ and when r is even $E[U^r] = \sigma_U^r r!/[2^{r/2}(r/2)!]$. If replicates of U are available, σ_U may be estimated from the replicates. For an arbitrary U that is moment determinate, the moments $\mu_U^{(j)} = E[U^j]$, $j = 1, 2, \ldots$ determine the distribution of U.

It is well known, if a regression model is used to fit $f(x)$, then ignoring the measurement error in X_i leads to biased estimates of the regression parameters. A relatively recent approach to provide better estimates of the regression parameters is to regress Y on $E[X \mid W]$ rather than the noisy W, since X is more likely to be closer to its expectation than to W. It is popularity is due to its transparency and intuitive appeal. Indeed, for the simple model $Y = a + bX + \epsilon$, $E[Y \mid X] = a + bX$ and thus $E[Y \mid W] = E[E[Y \mid X] \mid W] = a + bE[X \mid W]$. Thus, using $E[X \mid W]$ occurs naturally in this case. The approach can be applied to more complicated models. This procedure is called regression calibration, see [17]. Regression calibration is usually applied in the multiple linear regression problem. For the nonlinear case, that cannot be transformed to a linear model, regression calibration can be applied when the degree of nonlinearity is small and when a Taylor series provides a good linear approximation to the model as discussed in [17]. Even in the nonlinear case, when measurement error is small, X is more likely to be closer to its expectation, $E[X \mid W, Z]$ than to W. One caveat of this approach is, an exact expression for $E[X \mid W]$ exists only in rare cases, e.g. for Gaussian variates. For X and W jointly normal random variates with correlation coefficient, ρ

$$E[X \mid W] = \mu_X + \rho\frac{\sigma_X}{\sigma_W}(W - \mu_W) \tag{37}$$

in the univariate case and in the multivariate case a similar expression holds. This caveat can be dealt with by using the approximation, $\psi_n(w)$ for $E[X \mid W]$ provided in Section 3. Estimates of $\alpha_j = E[UW^j]$, $j = 0, 1, \ldots$ are only required. Below, it is shown how to calculate these moments from data (W_j, U_j), $j = 1, 2, \ldots, N$.

Since $E[X \mid W] = W - E[U \mid W]$, all computations can be done using $E[U \mid W]$. If one observes (W_i, U_i, Y_i), $i = 1, 2, \ldots, N$, then one can estimate α_j by

$$\hat{\alpha}_j = \frac{1}{N}\sum_{i=1}^{N} U_i W_i^j.$$

Often though, the noise U is unobservable, but assumptions about the moments of the noise can be made from experience. In this case, assuming $\mu_W^{(j)} = E[W^j]$ and $\mu_U^{(j)} = E[U^j]$, $j = 1, 2, \ldots$ are available (or estimates of), α_j may be found as follows. Using the binomial expansion

$$E[UW^j] = \sum_{i=0}^{j} \binom{j}{i} E[U^{i+1}]E[X^{j-i}] = \sum_{i=0}^{j} \binom{j}{i} \mu_U^{(i+1)} \mu_X^{(j-i)}.$$

The moments, $\mu_X^{(j)} = E[X^j]$ may be obtained from the binomial expansion of $E[W^k] = E[(X+U)^k]$ by recursively solving

$$\mu_W = \mu_X \tag{38}$$

$$\mu_W^{(2)} = \mu_X^{(2)} + \mu_U^{(2)} \tag{39}$$

$$\mu_W^{(3)} = \mu_X^{(3)} + 3\mu_X \mu_U^{(2)} + \mu_U^{(3)} \tag{40}$$

$$\vdots \tag{41}$$

$$\mu_W^{(r)} = \sum_{j=0}^{r} \binom{r}{j} \mu_U^{(j)} \mu_X^{(r-j)}. \tag{42}$$

In regression calibration, the normal approximation (37) is often used for $E[U \mid W = w]$ when the data are not normal. The following simulation experiment illustrates the improvement of $\psi_n(w)$ over the normal approximation when the data are not normal. In the simulation, the noise U is distributed as a symmetric normal, $N(0, \sigma_U^2)$, X is assumed to be independent of U and from a contaminated normal distribution,

$$P[X \le x] = (1-\epsilon)\Phi(x/\sigma_1) + \epsilon\Phi(x/\sigma_2) \quad 0 \le \epsilon \le 1 \tag{43}$$

where $\Phi(x)$ is the standard normal distribution function and the following parameters are used:

σ_U	σ_1	σ_2	ϵ
0.2	0.1	0.2	0.1

First we find the distribution of $W = X + U$ and the exact expression for $E[U \mid W]$ when X has a contaminated normal distribution. One can easily show that

$$F(w) = (1-\epsilon)\Phi\left(w/\sqrt{\sigma_1^2 + \sigma_U^2}\right) + \epsilon\Phi(w/\sqrt{\sigma_2^2 + \sigma_U^2}) \tag{44}$$

$$E[U \mid W = w] = (1-\epsilon)\frac{f_{X_1+U}(w)}{f_W(w)} E[U \mid X_1 + U] + \frac{f_{X_2+U}(w)}{f_W(w)} E[U \mid X_2 + U] \tag{45}$$

where $f_Z(x)$ denotes the density of the random variable Z and X_i are standard normal random variables with variances σ_i^2, i.e. the contamination components of the variate X. Justification of (44) is shown below.

$$\begin{aligned} P[W \le w] &= P[X + U \le w] = \int_{-\infty}^{\infty} P[U \le w - x] dP[X \le x] \\ &= (1-\epsilon)\int_{-\infty}^{\infty} P[U \le w - x] dP[X_1 \le x] + \epsilon \int_{-\infty}^{\infty} P[U \le w - x] dP[X_2 \le x] \\ &= (1-\epsilon)P[U + X_1 \le w] + \epsilon P[U + X_2 \le w]. \end{aligned}$$

Expression (45) can be shown similarly.

In this simulation, the interval $[-M_n, M_n] = [-3, 3]$ with $n = 35$ is used, which is sufficiently large to approximate $E[U \mid W = w]$. Indeed, when $w = -1.8$, $p(w) = F'(w) = 2.264(10^{-10})$. The results are shown in Table 1. The moment approximations are close to the exact values, $E[U \mid W = w]$, even up through $p(w)$ of the order 10^{-10}, whereas the normal approximate is reasonably close to the exact approximation for values of $p(w)$ below 0.14, but performs poorly for values of w with $p(w) < 0.14$.

Another issue with using the normal approximation (37) in the errors in variables problem is that it requires ρ, the correlation coefficient between X and W. This correlation coefficient is usually unknown, since X is not observed. If it is assumed the true value, X, and the measurement error, U, are independent, then $\rho = \sigma_U/\sigma_W$, which is usually estimable. This dependence is briefly discussed in [19].

The moment algorithm described above requires only $E[UW^j]$, $j = 1, 2, \ldots$, not the independence of X and U. To show $\psi_n(w)$ provides a better approximation than the normal approximation when independence between X and U fails, the following simulation is run. Let X and U have Gaussian marginal distributions with means and variances $(0, \sigma_X^2)$ and $(0, \sigma_U^2)$, respectively, but rather than being uncorrelated, their joint distribution is determined by the Farlie–Gumbel–Morgenstern

Table 1
Contaminated normal data: Comparison of $\psi_n(w)$ with its true value $E[U \mid W = w]$ and with the normal approximation given in (37). (Results rounded to 4 digit precision.)

w	$\psi_n(w)$	$E[U \mid W = w]$	Normal aprox	$p(w)$
−1.71429	−0.8134	−0.8572	−1.294	1.488 (10^{-9})
−1.62857	−0.8376	−0.8146	−1.229	8.923 (10^{-9})
−1.54286	−0.7599	−0.7721	−1.164	4.884 (10^{-8})
−1.45714	−0.7369	−0.7303	−1.100	2.442 (10^{-7})
−1.37143	−0.6861	−0.6897	−1.035	1.118 (10^{-6})
−1.28571	−0.6536	−0.6516	−0.9704	4.702 (10^{-6})
−1.2	−0.6165	−0.6176	−0.9057	0.00001830
−1.11429	−0.5904	−0.5898	−0.8410	0.00006664
−1.02857	−0.5686	−0.5690	−0.7763	0.0002304
−0.942857	−0.5533	−0.5531	−0.7116	0.0007662
−0.857143	−0.5365	−0.5366	−0.6469	0.002464
−0.771429	−0.5130	−0.5130	−0.5822	0.007600
−0.685714	−0.4789	−0.4789	−0.5175	0.02204
−0.6	−0.4345	−0.4344	−0.4528	0.05874
−0.514286	−0.3819	−0.3819	−0.3881	0.1410
−0.428571	−0.3237	−0.3237	−0.3235	0.3006
−0.342857	−0.2619	−0.2619	−0.2588	0.5633
−0.257143	−0.1979	−0.1979	−0.1941	0.9222
−0.171429	−0.1325	−0.1325	−0.1294	1.314
−0.0857143	−0.06644	−0.06644	−0.06469	1.627
0	0	0	0	1.747
0.0857143	0.06644	0.06644	0.06469	1.627
0.171429	0.1325	0.1325	0.1294	1.314
0.257143	0.1979	0.1979	0.1941	0.9222
0.342857	0.2619	0.2619	0.2588	0.5633
0.428571	0.3237	0.3237	0.3235	0.3006
0.514286	0.3819	0.3819	0.3881	0.1410
0.6	0.4345	0.4344	0.4528	0.05874
0.685714	0.4789	0.4789	0.5175	0.02204
0.771429	0.5130	0.5130	0.5822	0.007600
0.857143	0.5365	0.5366	0.6469	0.002464
0.942857	0.5533	0.5531	0.7116	0.0007662
1.02857	0.5686	0.5690	0.7763	0.0002304
1.11429	0.5904	0.5898	0.8410	0.00006664
1.2	0.6165	0.6176	0.9057	0.00001830
1.28571	0.6536	0.6516	0.9704	4.702 (10^{-6})
1.37143	0.6861	0.6897	1.035	1.118 (10^{-6})
1.45714	0.7369	0.7303	1.100	2.442 (10^{-7})
1.54286	0.7599	0.7721	1.164	4.884 (10^{-8})
1.62857	0.8376	0.8146	1.229	8.923 (10^{-9})
1.71429	0.8134	0.8572	1.294	1.488 (10^{-9})

copula $C(u, v) = uv + \theta u(1-u)v(1-v)$, $0 \le u, v \le 1$, see [37,38]. Let the normal approximation be computed assuming independence of X and U. Table 2 contains the results of this simulation. Shown are $\psi_n(w)$, $E[U \mid W = w]$, and the normal approximation. The joint distribution of (X, W), as well as, the marginal distribution of $W = X + U$ is required to compute $E[U \mid W = w]$ and α_j. These distributions can be found in [39]. In the computations, $\sigma_X = 0.1$, $\sigma_U = 0.2$, and the Hamburger approximation with $(-M_n, M_n)$, $M_n = 4$, $n = 35$ were used. The results of Table 2 clearly show $\psi_n(w)$ is closer to the true conditional expectation, even for the Farlie–Gumbel–Morgenstern copula, where the dependence between the marginals is modest. Thus, using $\psi_n(w)$ should lead to less error, regardless of whether dependence exists between X and U.

Comments, caveats and comparisons: The condition numbers of the inverse of $n \times n$ Vandermonde matrices on $[-1, 1]$ with equispaced and Chebyshev nodes are approximately $\pi^{-1}e^{-\pi/4}8^n$ and $4^{-1}3^{3/4}(1+\sqrt{2})^n$, see [25]. Thus, inaccuracies in the higher moments caused by roundoff or measurement errors result in errors in the solution of a Vandermonde system and can make the moment algorithm unstable. Some software packages allow one to increase the precision of calculations and thus mitigate some of these errors. In our calculations, Mathematica was used, which allowed calculations with 100 digits of precision. For instances when the algorithm is not stable (i.e. precision and roundoff errors produce large values of ψ_n), in addition to increasing precision, I found by centering $p(w)$, i.e. solving the moment problem for $W' = W - \mu_W$ instead of W helps stabilize the algorithm. Condition number calculations shown in (p. 418) [27] suggest that the condition number of the inverse of a Vandermonde matrix decreases for intervals symmetric about zero. There is no loss in generality by centering, because $\psi_W(w) = \psi_{W'}(w - \mu_W)$. For symmetric distributions W about the origin whose tails decay like the normal distribution or faster, I found the algorithm had no problem with stability. If the algorithm fails to stabilize, one may have to perform some form of regularization. I should point out that stability of almost all moment reconstructions will be affected if empirical moments are used to estimate α_j, $j = 0, 1, 2, \ldots$. Usually, $p(w)$ is not known; it can be estimated with a nonparametric kernel density. Since computation time is not a problem with this algorithm, n the number of summands

Table 2
Comparison of $\psi_n(w)$ and the normal approximation with the true value $E[U \mid W = w]$ in the errors in variables problem, when the true value X and the measurement error U are dependent having joint distribution defined by the Farlie–Gumbel–Morgenstern copula. (Results rounded to 4 digit precision.)

w	$\psi_n(w)$	$E[U \mid W = w]$	Normal approx	$p(w)$
−1.48571	−0.08613	−0.2978	−0.2971	$5.524\,(10^{-10})$
−1.37143	−0.2929	−0.2753	−0.2743	$1.443\,(10^{-8})$
−1.25714	−0.2510	−0.2528	−0.2514	$2.901\,(10^{-7})$
−1.14286	−0.2307	−0.2305	−0.2286	$4.485\,(10^{-6})$
−1.02857	−0.2082	−0.2082	−0.2057	0.00005331
−0.914286	−0.1861	−0.1861	−0.1829	0.0004870
−0.8	−0.1640	−0.1640	−0.16	0.003417
−0.685714	−0.1419	−0.1419	−0.1371	0.01840
−0.571429	−0.1196	−0.1196	−0.1143	0.07594
−0.457143	−0.09699	−0.09699	−0.09143	0.2396
−0.342857	−0.07377	−0.07377	−0.06857	0.5777
−0.228571	−0.04981	−0.04981	−0.04571	1.068
−0.114286	−0.02514	−0.02514	−0.02286	1.530
0	0	$-1.4619\,(10^{-17})$	0	1.722
0.114286	0.02514	0.02514	0.02286	1.530
0.228571	0.04981	0.04981	0.04571	1.068
0.342857	0.07377	0.07377	0.06857	0.5777
0.457143	0.09699	0.09699	0.09143	0.2396
0.571429	0.1196	0.1196	0.1143	0.07594
0.685714	0.1419	0.1419	0.1371	0.01840
0.8	0.1640	0.1640	0.16	0.003417
0.914286	0.1861	0.1861	0.1829	0.0004870
1.02857	0.2082	0.2082	0.2057	0.00005331
1.14286	0.2307	0.2305	0.2286	$4.485\,(10^{-6})$
1.25714	0.2510	0.2528	0.2514	$2.901\,(10^{-7})$
1.37143	0.2929	0.2753	0.2743	$1.443\,(10^{-8})$
1.48571	0.08613	0.2978	0.2971	$5.524\,(10^{-10})$

was chosen as large as possible, greater than M_n, i.e. until the algorithm appears to become stable. There are probably more rigorous methods to choose n, e.g. by simulation if possible.

There are several different approaches to the moment problem, see for example, the work by Mnatsakanov, Tagliani among others in [15,32,33,40–43]. Most of these approaches deal with reconstruction of a density or its distribution function and mostly for the Hausdorff moment problem. The reconstruction of the conditional moment function requires less constraints, and thus is simpler to implement. The most common numerical moment problem is the determination of a probability density, $f(x)$ from it moment sequence. This problem requires three constraints (1) $\int f(x)dx = 1$, (2) $\int x^k f(x)dx = \mu_k$, $k = 0, 1, 2, \ldots$ and (3) $f(x) > 0$. Because of these constraints, this problem is usually solved by finding the maximum entropy density, i.e., by maximizing the entropy functional $E(f) = \int f(x)\ln(f(x))dx$ subject to the constraints (1) and (2). By using Lagrange multipliers and taking the Frechet derivative of $E(f)$, this problem has a closed form solution $f(x) = \exp[-\lambda_0 - \lambda_1 x - \cdots - \lambda_n x^n]$ where $\int x^k \exp[-\lambda_0 - \lambda_1 x - \cdots - \lambda_n x^n]dx = \mu_k$, $k = 0, 1, \ldots, n$. Resulting in the latter system of nonlinear equations for λ_i. Tagliani has done quite a bit of work on this problem, see [11–13]. The Jacobian matrix for this system of nonlinear equations is an ill-conditioned Hankel matrix with condition number of the order $(1+\sqrt{2})^{4n}/\sqrt{n}$. Thus, it suffers the same stability issue the Vandermonde system suffers. Tagliani and Gzyl [40] have found a way to mitigate this issue using fractional moments. For U nonnegative, $\psi(w) = E[U \mid W = w] > 0$, a maximum entropy reconstruction of $\psi(w)$ can be determined by removing constraint (1), $\hat{\psi}(w) = \exp[-\lambda_1 x - \cdots - \lambda_n x^n]$, $\int x^k \exp[-\lambda_1 x - \cdots - \lambda_n x^n]dx = \mu_k$, $k = 0, 1, \ldots, n$ can be applied only for the Hausdorff moments. I was not able to get this approximation for $\psi(w)$ with infinite support to stabilize.

5. Conclusions

It has been shown that for a jointly continuous random vector (U, W) with conditional expectation function $\psi(w) = E[U \mid W = w]$, $\psi(w)$ is determined by the moments $\alpha_j = E[UW^j]$, $j = 0, 1, \ldots$. Uniqueness holds provided the sequence $\alpha_j, j = 0, 1, \ldots$ is either a Hamburger, Stieltjes, or Hausdorff moment sequence, depending on the support of the marginal distribution, $F(w)$ of W. Furthermore, a procedure to reconstruct $\psi(w)$ from its moments was derived and it was shown how this reconstruction provides an improvement over normal approximations in the regression calibration problem.

Disclaimer: The National Institute of Standards and Technology does not endorse any software product mentioned in this article.

Acknowledgments

Thanks to two reviewers, both who made invaluable suggestions for the paper.

References

[1] J. Shohat, J.D. Tamarkin, The Problem of Moments, American Mathematical Society, Providence, RI, 1963.
[2] D. Widder, The Laplace Transform, Princeton University Press, 1941.
[3] N. Akhiezer, The Classical Moment Problem and Some Related Questions in Analysis, Oliver & Boyd, Edinburgh, 1965.
[4] T.J. Stieltjes, Recherches sur les fractions continues, Ann. Fac. Sci. Toulouse Math. 8 (1894) J. 1–122.
[5] T.J. Stieltjes, Recherches sur les fractions continues, Ann. Fac. Sci. Toulouse Math. 9 (1895) A. 1–47.
[6] C.C. Heyde, On a property of the lognormal distribution, J. R. Stat. Soc. Ser. B Stat. Methodol. 25 (1963) 392–393.
[7] C. Berg, The cube of a normal distribution is indeterminate, Ann. Probab. 16 (1988) 910–913.
[8] M.L. Targhetta, On a family of indeterminate distributions, J. Math. Anal. Appl. 147 (1990) 477–479.
[9] J. Stoyanov, Stieltjes classes for moment-indeterminate probability distributions, J. Appl. Probab. 41 (2004) 281–294.
[10] W. Feller, An Introduction to Probability Theory and its Applications II, Wiley, New York, 1966.
[11] A. Tagliani, Hausdorff moment problem and maximum entropy: a unified approach, Appl. Math. Comput. 105 (1999) 291–305.
[12] A. Tagliani, Numerical aspects of finite Hausdorff moment problem by maximum entropy approach, Appl. Math. Comput. 118 (2001) 133–149.
[13] A. Tagliani, Recovering a function from a finite number of moments, Inverse Problems 3 (1987) 501–517.
[14] V. John, I. Angelov, A. Öncül, D. Thevenin, Techniques for the reconstruction of a distribution from a finite number of its moments, Chem. Eng. Sci. 62 (2007) 2890–2904.
[15] R. Mnatsakanov, Hausdorff moment problem: reconstruction of probability density functions, Statist. Probab. Lett. 78 (2008) 1869–1877.
[16] J. Wooldridge, Economic Analysis of Cross Section and Panel Data, MIT Press, Cambridge, MA, 2008.
[17] R. Carroll, D. Rupert, L. Stefanski, C. Crainiceanu, Measurement Error in Nonlinear Models. A Modern Perspective, second ed., Chapman & Hall/CRC, 2006.
[18] K. Astriim, Stochastic Control Theory, Academic Press, 1970.
[19] W. Fuller, Measure Error Models, Wiley, New York, 1987.
[20] T. Apostol, Mathematical Analysis, second ed., Addison-Wesley Publishing, Reading, MA, 1974.
[21] A. Björck, V. Pereyra, Solution of Vandermonde systems of equations, Math. Comp. 65 (1970) 893–903.
[22] N.J. Higham, Error analysis of the Björck–Pereyra algorithm for solving Vandermonde systems, Numer. Math. 50 (1987) 613–632.
[23] D. Knuth, The Art of Computer Programming, second ed., Addison-Wesley, Reading, MA, 1981.
[24] N.N. Macon, A. Spitzbart, Inverses of Vandermonde matrices, Amer. Math. Monthly 65 (1958) 95–100.
[25] G.H. Golub, C.F. van Loan, Matrix Computations, third ed., The Johns Hopkins University Press, London, 1996.
[26] W. Gautschi, On inverses of Vandermonde and confluent Vandermonde matrices, Numer. Math. 4 (1962) 117–123.
[27] N. Higham, Accuracy and Stability of Numerical Algorithms, SIAM, Philadelphia, PA, 2002.
[28] W. Feller, An Introduction to Probability Theory and its Applications I, Wiley, New York, 1966.
[29] P. Davis, P. Rabinowitz, Methods of Numerical Integration, second ed., Academic Press, Inc., 1984.
[30] I. Gradshteyn, I. Ryzhik, Table of Integral Series and Products, seventh ed., Academic Press, 2007.
[31] F. Olver, D. Lozier, R. Boivert, C. Clark, NIST Handbook of Mathematical Functions, Cambridge University Press, 2010.
[32] R. Mnatsakanov, Moment-recovered approximations of multivariate distributions; Laplace transform inversion, Statist. Probab. Lett. 81 (2011) 1–7.
[33] R. Mnatsakanov, K. Sarkisian, A note on recovering the distributions from exponential moments, Appl. Math. Comput. 219 (2013) 8730–8737.
[34] B. Efron, Estimation and accuracy after model selection, J. Amer. Statist. Assoc. (2014) (in press).
[35] T. Souders, D.R. Flach, C. Hagwood, G.L. Yang, The effects of timing jitter in sampling systems, IEEE Trans. Instrum. Meas. 39 (1989) 80–85.
[36] D. Pierce, D. Stram, M. Vaeth, Structure of dietary measurement error: results of the open biomarker study, J. Amer. Statist. Assoc. 87 (1992) 351–359.
[37] R. Nelson, Introduction to Copulas, Springer Verlag, Heidelberg, 2006.
[38] W.V.R. Cherubini, E. Luciano, Copula Methods in Finance, John Wiley & Sons, Ltd., Chichester, 2004.
[39] R. Cherubini, S. Mulinacci, S. Romagnoli, On the distribution of the (un)bounded sum of random variables, Insurance Math. Econom. 48 (2011) 56–63.
[40] H. Gzyl, A. Tagliani, Hausdorff moment problem and and fractional moments: a simplified procedure, Appl. Math. Comput. 216 (2010) 3319–3328.
[41] N. Inverardi, P. Petri, G. Pontuala, A. Tagliani, Hausdorff moment problem via fractional moments, Appl. Math. Comput. 144 (2003) 61–74.
[42] J. Laserre, A semidefinite programming approach to the generalized problem of moments, Math. Program. Ser. B 112 (2008) 65–92.
[43] H. Gzyl, A. Tagliani, Stieltjes moment problem and fractional moments, Appl. Math. Comput. 216 (2010) 3307–3318.