Supplementary Material

APPENDIX

Proof of Theorem 1 (comparison between likelihood and path-sampling estimators).
Assume that \((\theta_1, \ldots, \theta_m)\) are equidistant; \(m^4/N \to \infty\) as \(N \to \infty\); \(q(x; t)\) is three-
times continuously differentiable in \(t \in [\theta_1, \theta_m]\); and there exists a nonnegative function
\(K(x)\) such that \(\sup_{\theta_j \leq t \leq \theta_{j+1}} \{|\partial^2 q(x; t)/\partial t^2|/q(x; \theta_j)| \leq K(x)\) and \(E_{\theta_j}(K) < \infty\) for
\(j = 1, \ldots, m - 1\). We show that the difference, \(\delta\), between estimators (5) and (9) is of
order \(o_p(N^{-1/2})\). In fact, by the error bound for the trapezoidal rule,

\[
|\delta| \leq \frac{(\theta_m - \theta_1)^3}{12(m-1)^2} \sup_{j=1,\ldots,m-1} \sup_{\theta_j \leq t \leq \theta_{j+1}} \left| \frac{\partial^2}{\partial t^2} \left( \frac{\theta_{j+1} - t}{\theta_{j+1} - \theta_j} h_j(t) + \frac{t - \theta_j}{\theta_{j+1} - \theta_j} h_{j+1}(t) \right) \right|
\]

\[
= \frac{(\theta_m - \theta_1)^3}{12(m-1)^2} \sup_{j=1,\ldots,m-1} \sup_{\theta_j \leq t \leq \theta_{j+1}} \left| -2 \frac{\partial}{\partial t} \left( h_j(t) - h_{j+1}(t) \right) + \frac{\theta_{j+1} - t}{\theta_{j+1} - \theta_j} \frac{\partial^2}{\partial t^2} h_j(t) + \frac{t - \theta_j}{\theta_{j+1} - \theta_j} \frac{\partial^2}{\partial t^2} h_{j+1}(t) \right|
\]

where \(h_j(t) = \tilde{E}_{\theta_j} \left\{ \frac{x \theta_j}{q(x; \theta_j)} \right\} / \tilde{E}_{\theta_j} \left\{ \frac{q(x; \theta_j)}{q(x; \theta_j)} \right\}\). By the uniform (strong) law of large numbers (e.g., Ferguson 1996, Chapter 16), \(\sup_{\theta_j \leq t \leq \theta_{j+1}} \left| \frac{\partial}{\partial t} h_j(t) - \frac{d^3}{dt^3} \log \left( \frac{Z_t}{Z_{\theta_j}} \right) \right| = o_p(1)\) and
\(\sup_{\theta_j \leq t \leq \theta_{j+1}} \left| \frac{\partial^2}{\partial t^2} h_j(t) - \frac{d^3}{dt^3} \log \left( \frac{Z_t}{Z_{\theta_j}} \right) \right| = o_p(1)\). Therefore,

\[
|\delta| \leq \frac{(\theta_m - \theta_1)^3}{12(m-1)^2} \max_{j=1,\ldots,m-1} \sup_{\theta_j \leq t \leq \theta_{j+1}} \frac{d^3}{dt^3} \log \left( \frac{Z_t}{Z_{\theta_j}} \right) + o_p(1)
\]

and hence \(|\delta| = o_p(N^{-1/2})\) because \(m^4/N \to \infty\).

Next, we show that estimator (1) has no greater asymptotic variance than estimator
(9). By numerical integration, estimator (9) can be approximated by a convex combi-
nation of a finite number of estimators in the form

\[
\sum_{j=1}^{m-1} \log \left[ \frac{\tilde{E}_{\theta_j} \left\{ \frac{q(x; t_j)}{q(x; \theta_j)} \right\}}{\tilde{E}_{\theta_{j+1}} \left\{ \frac{q(x; t_j)}{q(x; \theta_{j+1})} \right\}} \right]
\]

such that the error is of order \(o_p(N^{-1/2})\), where \(t_j \in [\theta_j, \theta_{j+1}]\) is fixed for \(j = 1, \ldots, m - 1\). Each such estimator belongs to Meng and Wong’s (1996) class of extended bridge
sampling estimators and hence, by Tan (2004, Theorem 1), has no smaller asymptotic
variance than estimator (1). The claim follows because a convex combination of such
estimators still has no smaller asymptotic variance than estimator (1).
Proof of Theorem 2. To maximize the log-likelihood, the Lagrange function is

\[
\sum_{h=1}^{3} \left\{ \sum_{i=1}^{N_h} \log \mu_h(\{x_{hi}\}) - \sum_{j=1}^{m_h} n_{hj} \log \int q(x; \theta_{hj}) \, d\mu_h \right\} - \sum_{l=1}^{r_{12}} \lambda_{12,l} \left\{ \int q(x; \theta_{12,l}) \, d\mu_1 - \int q(x; \theta_{12,l}) \, d\mu_2 \right\}
- \sum_{l=1}^{r_{23}} \lambda_{23,l} \left\{ \int q(x; \theta_{23,l}) \, d\mu_2 - \int q(x; \theta_{23,l}) \, d\mu_3 \right\}.
\]

The Lagrange conditions are

\[
\mu_1^{-1}(\{x_{1i}\}) = \sum_{j=1}^{m_1} n_{1j} Z_{\theta_{1j}}^{-1} q(x_{1i}; \theta_{1j}) + \sum_{l=1}^{r_{12}} \lambda_{12,l} q(x_{1i}; \theta_{12,l}), \quad i = 1, \ldots, N_1,
\]

\[
\mu_2^{-1}(\{x_{2i}\}) = \sum_{j=1}^{m_2} n_{2j} Z_{\theta_{2j}}^{-1} q(x_{2i}; \theta_{2j}) - \sum_{l=1}^{r_{12}} \lambda_{12,l} q(x_{2i}; \theta_{12,l}) + \sum_{l=1}^{r_{23}} \lambda_{23,l} q(x_{2i}; \theta_{23,l}), \quad i = 1, \ldots, N_2,
\]

\[
\mu_3^{-1}(\{x_{3i}\}) = \sum_{j=1}^{m_3} n_{3j} Z_{\theta_{3j}}^{-1} q(x_{3i}; \theta_{3j}) + \sum_{l=1}^{r_{23}} \lambda_{23,l} q(x_{3i}; \theta_{23,l}), \quad i = 1, \ldots, N_3,
\]

which lead to the expressions for the MLEs \((\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3)\). Moreover, multiplying both sides of these conditions by \(\mu_1(\{x_{1i}\}), \mu_2(\{x_{2i}\}), \text{ or } \mu_3(\{x_{3i}\})\) shows that \(\sum_{l=1}^{r_{12}} \hat{\lambda}_{12,l} \hat{Z}_{\theta_{12,l}} = 0\) and \(\sum_{l=1}^{r_{23}} \hat{\lambda}_{23,l} \hat{Z}_{\theta_{23,l}} = 0\).

Asymptotic variance matrix of \(\log \hat{Z}_{(1),\text{lik}}\) under IID sampling (unknown normalizing constants). We derive a sandwich formula for estimating the asymptotic variance matrix of \(\log \hat{Z}_{(1),\text{lik}}\) under IID sampling. For \(h = 1, 2, 3\), let \(\Pi_h\) be the \(M_h \times M_h\) diagonal matrix with \((\theta, \theta)\)th element \(n_{hj}/N\) if \(\theta = \theta_{hj}\) and 0 otherwise, \(R_h\) be the \(N_h \times m_h\) matrix with \((i, j)\)th element \((n_{hj}/N)^{-1}\) if \(x_{hi}\) is simulated from \(P_{\theta_{hj}}\) and 0 otherwise, and \(W_h\) be the \(N_h \times M_h\) matrix with \((i, \theta)\)th element \(N \hat{Z}_{\theta_{1i}}^{-1} q(x_{hi}; \theta) \hat{\mu}_h(\{x_{hi}\})\). Let \(U_{12}\) be the \(N_1 \times r_{12}\) matrix with \((i, l)\) element \(N \hat{Z}_{\theta_{1i}}^{-1} q(x_{1i}; \theta_{12,l}) \hat{\mu}_1(\{x_{1i}\})\) and \(V_{12}\) be the \(N_2 \times r_{12}\) matrix with \((i, l)\) element \(N \hat{Z}_{\theta_{1i}}^{-1} q(x_{2i}; \theta_{12,l}) \hat{\mu}_2(\{x_{2i}\})\). Let \(U_{23}\) be the \(N_2 \times r_{23}\) matrix with \((i, l)\) element \(N \hat{Z}_{\theta_{1i}}^{-1} q(x_{2i}; \theta_{23,l}) \hat{\mu}_2(\{x_{2i}\})\) and \(V_{23}\) be the \(N_3 \times r_{23}\) matrix with \((i, l)\) element \(N \hat{Z}_{\theta_{1i}}^{-1} q(x_{3i}; \theta_{23,l}) \hat{\mu}_3(\{x_{3i}\})\). Let \(I\) be the \(M^\dagger \times M^\dagger\) identity matrix, \(\Pi\) be the \(M^\dagger \times M^\dagger\) matrix with \((\Pi_1, \Pi_2, \Pi_3)\) on the diagonal, \(\Pi_s\) be the \(m \times m\) principal submatrix of \(\Pi\) with nonzero diagonal elements, and
\[
R = \begin{pmatrix} R_1 \\ R_2 \\ R_3 \end{pmatrix}, \quad W = \begin{pmatrix} W_1 \\ W_2 \\ W_3 \end{pmatrix}, \quad U = \begin{pmatrix} U_{12} \\ -V_{12} \\ U_{23} \\ -V_{23} \end{pmatrix}.
\]

The sum of each column in \(W\) is \(N\) and that of each column in \(U\) is 0. The sum of each row of the combined matrix \(\{W\Pi, U\text{diag}(\hat{\lambda}_{12}, \hat{\lambda}_{23})\}\) is 1. Let

\[
O = \frac{1}{N} \begin{pmatrix} W^T \\ U^T \end{pmatrix} (W, U), \quad H = \frac{1}{N} \begin{pmatrix} W^T \\ U^T \end{pmatrix} (W\Pi, U) + \begin{pmatrix} I \\ 0 \end{pmatrix}
\]

\[
Q = \frac{1}{N} \begin{pmatrix} W^T \\ U^T \end{pmatrix} R, \quad G = O - Q\Pi_sQ^T
\]

The asymptotic variance matrix of \(\log \hat{Z}_{(1),\text{lik}}\) can be consistently estimated by the leading \((M^\dagger - 1) \times (M^\dagger - 1)\) principal submatrix of (20) below.

Consider \((\log \hat{Z}_{(1),\text{lik}}; \hat{Z}_{\theta_1}, \hat{\lambda}_{12}, \hat{\lambda}_{23})\) jointly as a solution to \(\int \hat{Z}_\theta^{-1} q(x; \theta) = 1 (\theta \in \Theta_1, \theta \neq \theta_{11}, h = 1 \text{ or } \theta \in \Theta_h, h = 2, 3)\) and equations (14)–(15). By the asymptotic theory for general M-estimators (e.g., van der Vaart 1998), the asymptotic variance matrix can be consistently estimated by

\[
\frac{1}{N} H_{(1)}^{-1} G_{(1)} H_{(1)}^{-1},
\]

where \(H_{(1)}\) and \(G_{(1)}\) are formed by deleting \(\theta_{11}\)th row and column from \(H\) and \(G\). The probability limits of \(G_{(1)}\) and \(H_{(1)}^{-1} G_{(1)} H_{(1)}^{-T}\) are of full rank minus 2. Alternatively, the formula is valid with \(G\) replaced by \(O - O\Pi_sO_s^T\), where \(O_s = N^{-1}(W, U)^T W_s\) and \(W_s\) consists of the columns in \(W\) indexed by \(\theta_{hj}\) \((j = 1, \ldots, m_h, h = 1, 2, 3)\). In fact, \(Q\) and \(O_s\) are asymptotically equal to each other.

We provide a simplification of the sandwich formula. Redefine \(O\) and \(Q\) as their probability limits. Then

\[
G = O - O\Pi_sO_s^T,
\]

which is of full rank minus 3. Write \(G_{(1)}\) and \(H_{(1)}\) in the following partitions:

\[
G_{(1)} = \begin{pmatrix} G_{11} & G_{12} \\ G_{12}^T & G_{22} \end{pmatrix} = \begin{pmatrix} O_{w(1)w(1)} - O_{w(1)w} \Pi O_{w(1)} & O_{w(1)u} - O_{w(1)w} \Pi O_{wu} \\ O_{u(1)w} - O_{uw} \Pi O_{w(1)} & O_{uu} - O_{uw} \Pi O_{wu} \end{pmatrix},
\]

\[
H_{(1)} = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} = \begin{pmatrix} -O_{w(1)w} \Pi(1) + I_{(1)} & -O_{w(1)u} \\ -O_{u(1)w} \Pi(1) & -O_{uu} \end{pmatrix},
\]

3
where $I_{(1)}$ and $\Pi_{(1)}$ are formed by deleting $\theta_{11}$th row and column from $I$ and $\Pi$. To simplify $H_{(1)}^{-1} G_{(1)} H_{(1)}^{-1\top}$, we make use of the following formula:

$$H_{(1)}^{-1} = \begin{pmatrix} H_{11:2}^{-1} & -H_{11:2}^{-1} H_{12:2}^{-1} \\ -H_{22:2}^{-1} H_{12:2}^{-1} H_{11:2}^{-1} & H_{22:2}^{-1} + H_{22:2}^{-1} H_{11:2}^{-1} H_{12:2}^{-1} \end{pmatrix},$$

where $H_{11:2} = H_{11} - H_{12:2}^{-1} H_{21}$ is $I_{(1)} - (O_{w(1)w(1)} - O_{w(1)w(1)} O_{uu}^{-1} O_{uw(1)}) \Pi_{(1)}$. Let $\pi_1 = n_{11}/N$. The $(1, 1)$ block of $H_{(1)}^{-1} G_{(1)} H_{(1)}^{-1\top}$ is

$$H_{11:2}^{-1} \{(G_{11} - H_{12:2}^{-1} G_{12:2}) - (G_{12:2} - H_{12:2}^{-1} G_{22:2}) H_{22:2}^{-1} H_{12:2}^{-1}\} H_{11:2}^{-1\top} \Pi_{(1)}$$

because $O_{w(1)w(1)} \Pi_{(1)} 1_{(1)} = 1_{(1)}$ and $O_{uu} \Pi_{(1)} 1_{(1)} = 0$ and hence

$$H_{11:2}^{-1} \{O_{w(1)w(1)} - O_{w(1)w(1)} O_{uu}^{-1} O_{uw(1)}\} \Pi_{(1)} 1_{(1)} = 1_{(1)}.$$ The $(1, 2)$ block of $H_{(1)}^{-1} G_{(1)} H_{(1)}^{-1\top}$ is

$$H_{11:2}^{-1} \{(G_{11} - H_{12:2}^{-1} G_{12:2}) - (G_{12:2} - H_{12:2}^{-1} G_{22:2}) H_{22:2}^{-1} H_{12:2}^{-1}\} H_{11:2}^{-1\top} \Pi_{(1)}$$

Therefore, $\log \hat{Z}_{(1),\lambda} 1_{(1)}$ and $(\hat{Z}_{\theta_1, \lambda_{12}}, \hat{Z}_{\theta_1, \lambda_{23}})$ are asymptotically uncorrelated.
Asymptotic variance matrix of \( (\hat{Z}_{(1),\text{LIK}}^\top, \hat{Z}_{(1),\text{LIK}}^{\phi \top})^\top \) under IID sampling (unknown normalizing constants). The asymptotic variance matrix of \( (\hat{Z}_{(1),\text{LIK}}^\top, \hat{Z}_{(1),\text{LIK}}^{\phi \top})^\top \) can be estimated by the leading \( (2M^\dagger - 1) \times (2M^\dagger - 1) \) principal submatrix of (20) with the following modification throughout. For \( h = 1, 2, 3 \), replace the \( (\theta_{hj}, \theta_{hj}) \)th element in \( \Pi_h \) by \( (\hat{Z}_{\theta_{hj}}/\hat{Z}_{\theta_{11}})^{-2}(n_{hj}/N) \), the \((i,j)\)th element in \( R_h \) by \( (\hat{Z}_{\theta_{hj}}/\hat{Z}_{\theta_{11}})(n_{hj}/N)^{-1} \) if \( x_{hi} \) is simulated from \( P_{\theta_{hj}} \), and the \((i, \theta)\)th element in \( W_h \) by \( N\hat{Z}_{\theta_{11}}^{-1}q(x_{hi}; \theta)\hat{\mu}_h(\{x_{hi}\}) \). Replace \( W \) by \( (W, W^\phi) \) and \( U \) by \( (U, U^\phi) \) with

\[
W^\phi = \begin{pmatrix}
W_1^\phi \\
W_2^\phi \\
W_3^\phi
\end{pmatrix}, \quad U^\phi = \begin{pmatrix}
U_{12}^\phi \\
-V_{12}^\phi \\
U_{23}^\phi \\
-V_{23}^\phi
\end{pmatrix},
\]

where \( W_h^\phi \) is the \( N_h \times M_h \) matrix with \((i, \theta)\)th element \( N\hat{Z}_{\theta_{11}}^{-1}\phi(x_{hi})q(x_{hi}; \theta)\hat{\mu}_h(\{x_{hi}\}) \), \( U_{12}^\phi \) is the \( N_1 \times r_{12} \) matrix with \((i, l)\) element \( N\hat{Z}_{\theta_{11}}^{-1}\phi(x_{1i})q(x_{1i}; \theta_{12,l})\hat{\mu}_1(\{x_{1i}\}) \), \( V_{12}^\phi \) is the \( N_2 \times r_{12} \) matrix with \((i, l)\) element \( N\hat{Z}_{\theta_{11}}^{-1}\phi(x_{2i})q(x_{2i}; \theta_{21,l})\hat{\mu}_2(\{x_{2i}\}) \), and \( U_{23}^\phi \) and \( V_{23}^\phi \) are \( N_2 \times r_{23} \) and \( N_3 \times r_{23} \) matrices similar to \( U_{23} \) and \( V_{23} \).

Relationship between the likelihood, GMM, and regression estimators (unknown normalizing constants). Consider the estimating equations for \( \log \hat{Z}_{(1),\text{LIK}}, \hat{Z}_{\theta_{11}}, \lambda_{12}, \hat{Z}_{\theta_{11}}, \lambda_{23} \) but set \( (\lambda_{12}, \lambda_{23}) \) to 0. Let \( \hat{Z}_{(1),\text{GMM}} \) be the resulting GMM estimator by Hansen (1982). The asymptotic variance matrix of \( \log \hat{Z}_{(1),\text{GMM}} \) is

\[
\frac{1}{N} \left\{ (H_{11}, H_{21}^\top) G^{-1}_{(1)} \begin{pmatrix} H_{11} \\ H_{21} \end{pmatrix} \right\}^{-1},
\]

where \( G^{-1}_{(1)} \) is the Moore-Penrose generalized inverse of \( G_{(1)} \). The asymptotic variance matrix of \( \log \hat{Z}_{(1),\text{LIK}} \) or the \((1, 1)\) block of \( N^{-1}H_{(1)}^{-1} G_{(1)} H_{(1)}^{-1} \) is

\[
\frac{1}{N} \left\{ (I_{(1)}, 0) \begin{pmatrix} H_{(1)}^{-1} G_{(1)} H_{(1)}^{-1} \end{pmatrix}^{-1} \begin{pmatrix} I_{(1)} \\ 0 \end{pmatrix} \right\}^{-1},
\]

because the \((1, 2)\) block of \( H_{(1)}^{-1} G_{(1)} H_{(1)}^{-1} \) is 0. Therefore, the asymptotic variance matrices of \( \log \hat{Z}_{(1),\text{LIK}} \) and \( \log \hat{Z}_{(1),\text{GMM}} \) are identical, i.e., the two estimators are asymptotically equivalent to the first order to each other.
We now discuss the relationship between the GMM estimator and the regression estimator. Consider the estimating equations

\[ 1 = \int Z_\theta^{-1} q(x; \theta) \, d\mu_h, \quad \theta \in \Theta_h, \theta \neq \theta_h, \ h = 1, 2, 3, \]

\[ \int q(x; \theta_{12,l}) \, d\mu_1 = T_{12,l} \int q(x; \theta_{12,l}) \, d\mu_2, \quad l = 1, \ldots, r_{12}, \]

\[ \int q(x; \theta_{23,l}) \, d\mu_2 = T_{23,l} \int q(x; \theta_{23,l}) \, d\mu_3, \quad l = 1, \ldots, r_{23}, \]

where \( \mu_h = \left\{ \sum_{j=1}^{n_h} n_{hj} Z_{\theta_{hj}}^{-1} q(x_{hj}; \theta_{hj}) \right\}^{-1} \), and \( T_{12} = (T_{12,1}, \ldots, T_{12,r_{12}})^\top \) and \( T_{23} = (T_{23,1}, \ldots, T_{23,r_{23}})^\top \) with \( T_{12,1} = T_{23,1} \equiv 1 \) are unknowns in addition to \( Z(1) \). Let \( T(1) \) be the vector formed by deleting the two elements of 1 from \( (T_{12}^\top, T_{23}^\top)^\top \). This system of estimating equations with \( \log T \) set to 0 is equivalent to that used to define \( \hat{Z}_{(1),GMM} \).

By the discussion in Section 2.3.1, the solutions for \( \log Z(1) \) and \( \log T(1) \) are \( D \log \hat{Y}(1) \) and \( C^\top \log \hat{Y}(1) \) respectively, where \( D \) is an \( (M^\uparrow - 1) \times (M^\uparrow - 3) \) matrix and \( C^\top \) is a \( (r_{12} + r_{23} - 2) \times (M^\uparrow - 3) \) matrix. On one hand, the asymptotic variance matrix of \( (D^\top, C^\top)^\top \log \hat{Y}(1) \) is \( U = (D^\top, C^\top^\top) V (D^\top, C^\top^\top)^\top \), where \( V \) is the asymptotic variance matrix of \( \log \hat{Y}(1) \). Then the asymptotic variance matrix of \( \log \hat{Z}_{(1),REG} \) using control variates \( C^\top \log \hat{Y}(1) \) is \( \{(I_{(3)}, 0) U^{-1} (I_{(3)}, 0)^\top\}^{-1} \), where \( I_{(3)} \) is the identity matrix of size \( M^\uparrow - 3 \). On the other hand, by the sandwich formula for general M-estimators, the asymptotic variance matrix \( U \) is

\[ \frac{1}{N} \begin{pmatrix} H_{11,(2)} & \ast \\ H_{21} & \ast \end{pmatrix}^{-1} G_{(3)} \begin{pmatrix} H_{11,(2)}^\top & H_{21}^\top \\ \ast & \ast \end{pmatrix}^{-1}, \]

where \( G_{(3)} \) is formed by deleting \( \theta_{hi} \) \( (h = 1, 2, 3) \) row and column from \( G \) and \( H_{11,(2)} \) is formed by deleting \( \theta_{hi} \) \( (h = 2, 3) \) row from \( H_{11} \). Then the asymptotic variance matrix of \( \log \hat{Z}_{(1),REG} \) is

\[ \frac{1}{N} \left\{ (H_{11,(2)}^\top, H_{21}^\top) G_{(3)}^{-1} \begin{pmatrix} H_{11,(2)} & H_{21}^\top \\ H_{21} & \ast \end{pmatrix} \right\}^{-1}, \]

which is also the asymptotic variance matrix of \( \log \hat{Z}_{(1),GMM} \). Therefore, \( \hat{Z}_{(1),REG} \) and \( \hat{Z}_{(1),GMM} \) are asymptotically equivalent to the first order.
Proof of Theorem 3. To maximize the log-likelihood, the Lagrange function is

\[
\sum_{h=1}^{3} \sum_{i=1}^{N_h} \log \mu_h(x_{hi}) - N \sum_{h=1}^{3} \sum_{j=1}^{m_h} \nu_{hj} \left\{ \int f(x; \theta_{hj}) \, d\mu_h - 1 \right\} \\
- N \sum_{l=1}^{r_{12}} \lambda_{12,l} \left\{ \int q(x; \theta_{12,l}) \, d\mu_1 - \int q(x; \theta_{12,l}) \, d\mu_2 \right\} \\
- N \sum_{l=1}^{r_{23}} \lambda_{23,l} \left\{ \int q(x; \theta_{23,l}) \, d\mu_2 - \int q(x; \theta_{23,l}) \, d\mu_3 \right\}.
\]

The Lagrange conditions are

\[
\mu_1^{-1}(\{x_{i1}\}) = N \sum_{j=1}^{m_1} \nu_{1j} f(x_{i1}; \theta_{1j}) + N \sum_{l=1}^{r_{12}} \lambda_{12,l} q(x_{i1}; \theta_{12,l}), \quad i = 1, \ldots, N_1,
\]

\[
\mu_2^{-1}(\{x_{i2}\}) = N \sum_{j=1}^{m_2} \nu_{2j} f(x_{i2}; \theta_{2j}) - N \sum_{l=1}^{r_{12}} \lambda_{12,l} q(x_{i2}; \theta_{12,l}) + N \sum_{l=1}^{r_{23}} \lambda_{23,l} q(x_{i2}; \theta_{23,l}),
\]

\[
\mu_3^{-1}(\{x_{i3}\}) = N \sum_{j=1}^{m_3} \nu_{3j} f(x_{i3}; \theta_{3j}) + N \sum_{l=1}^{r_{23}} \lambda_{23,l} q(x_{i3}; \theta_{23,l}), \quad i = 1, \ldots, N_3,
\]

which lead to the expressions for the MLEs \((\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3)\). Multiplying both sides of these conditions by \(\mu_1(\{x_{i1}\}), \mu_2(\{x_{i2}\}), \text{ or } \mu_3(\{x_{i3}\})\) shows that \(\sum_{j=1}^{m_1} \hat{\nu}_{1j} + \sum_{l=1}^{r_{12}} \hat{\lambda}_{12,l} \hat{Z}_{\theta_{12,l}} = N_1/N\), \(\sum_{j=1}^{m_2} \hat{\nu}_{2j} + \sum_{l=1}^{r_{23}} \hat{\lambda}_{23,l} \hat{Z}_{\theta_{23,l}} = N_3/N\), and \(\sum_{h=1}^{3} \sum_{j=1}^{m_h} \hat{\nu}_{hj} = 1\).

Asymptotic variance matrix of \(\log \hat{Z}_{\text{lik}}\) under IID sampling (known normalizing constants). We derive a sandwich formula for estimating the asymptotic variance matrix of \(\log \hat{Z}_{\text{lik}}\) under IID sampling. For \(h = 1, 2, 3\), let \(W_h\) be the \(N_h \times M_h\) matrix with \((i, \theta)\)th element \(N \hat{Z}_\theta^{-1} q(x_{hi}; \theta) \hat{\mu}_h(\{x_{hi}\})\) and \(\Gamma_h\) be the \(N_h \times m_h\) matrix with \((i, j)\)th element \(N f(x_{hi}; \theta_{hj}) \hat{\mu}_h(\{x_{hi}\})\). Let \(U_{12}\) be the \(N_1 \times r_{12}\) matrix with \((i, l)\)th element \(N q(x_{hi}; \theta_{12,l}) \hat{\mu}_1(\{x_{hi}\})\) and \(V_{12}\) be the \(N_2 \times r_{12}\) matrix with \((i, l)\)th element \(N q(x_{hi}; \theta_{23,l}) \hat{\mu}_2(\{x_{hi}\})\). Let \(U_{23}\) be the \(N_2 \times r_{23}\) matrix with \((i, l)\)th element \(N q(x_{hi}; \theta_{23,l}) \hat{\mu}_2(\{x_{hi}\})\) and \(V_{23}\) be the \(N_3 \times r_{23}\) matrix with \((i, l)\)th element \(N q(x_{hi}; \theta_{23,l}) \hat{\mu}_3(\{x_{hi}\})\). Let

\[
W = \begin{pmatrix} W_1 & W_2 & W_3 \end{pmatrix}, \Gamma = \begin{pmatrix} \Gamma_1 \\ \Gamma_2 \\ \Gamma_3 \end{pmatrix}, U = \begin{pmatrix} U_{12} & -V_{12} & U_{23} \end{pmatrix},
\]

\huip
The sum of each column in $W$ or $\Gamma$ is 1 and that of each column in $U$ is 0. The sum of each row of the combined matrix \{$\Gamma \text{diag}(\tilde{\nu})$, $U \text{diag}(\tilde{\lambda}_{12}, \tilde{\lambda}_{23})$\} is 1. Let

$$O = \frac{1}{N} W^\top W, \quad A = \frac{1}{N} \begin{pmatrix} \Gamma^\top \\ U^\top \end{pmatrix} (\Gamma, U), \quad B = \frac{1}{N} W^\top (\Gamma, U).$$

The asymptotic variance matrix of $\log \hat{Z}_{\text{LIK}}$ can be consistently estimated by

$$\frac{1}{N} \left( O - I I^\top - B(1) A^{-1}(1) B^\top(1) \right),$$

where $I$ is the $M^\top \times 1$ vector of ones, $B(1) = BJ$, $A(1) = J^\top AJ$, and $J$ is the matrix formed by stacking the row vector $(-1, -1, \ldots, -1, 0, \ldots, 0)$ with $m-1$ zeros and $r_{12} + r_{23}$ ones and the identity matrix of size $m + r_{12} + r_{23} - 1$.

Let $\tilde{\nu}(1)$ be the vector formed by deleting $\tilde{\nu}_{11}$ in $\tilde{\nu}$. Consider ($\log \hat{Z}_{\text{LIK}}, \tilde{\nu}(1), \tilde{\lambda}_{12}, \tilde{\lambda}_{23}$) jointly as a solution to $\int \hat{Z}_{\text{d}}^{-1} q(x; \theta) \, d\mu_h = 1$ ($\theta \in \Theta_h, h = 1,2,3$), $\int f(x; \theta_h) \, d\mu_h - \int f(x; \theta_{11}) \, d\mu_1 = 0$ ($j = 2, \ldots, m_h, h = 1$ or $j = 1, \ldots, m_h, h = 2,3$), and equations (18)–(19). The asymptotic variance matrix of $\log \hat{Z}_{\text{LIK}}$ can be consistently estimated by the leading $M^\top \times M^\top$ principal submatrix of

$$\frac{1}{N} \left( \begin{pmatrix} I & B(1) \\ 0 & A(1) \end{pmatrix} \right)^{-1} \Omega(1) \left( \begin{pmatrix} I & 0 \\ B(1) & A(1) \end{pmatrix} \right)^{-1},$$

with

$$\Omega(1) = \left( \begin{pmatrix} O - W^\top R \Pi_s R^\top W_N \\ B(1) - \frac{W^\top R \Pi_s R^\top (\Gamma(1),U)}{N} \end{pmatrix} A(1) - \frac{W^\top R \Pi_s R^\top (\Gamma(1),U)}{N} \right),$$

where $(\Gamma(1), U) = (\Gamma, U)J$, and $I$, $R$ and $\Pi_s$ are defined as in Section 2.3.2. We need only to show that

$$\begin{pmatrix} I & -B(1)A^{-1}(1) \\ -A_{(1)}^{-1} B^\top(1) \end{pmatrix} \left( \begin{pmatrix} W^\top R \Pi_s R^\top W_N \\ (\Gamma(1),U)^\top R \Pi_s R^\top W_N \end{pmatrix} \frac{W^\top R \Pi_s R^\top (\Gamma(1),U)}{N} \right) \left( \begin{pmatrix} I \\ -A_{(1)}^{-1} B^\top(1) \end{pmatrix} \right)$$

is asymptotically equal to $1^\top 1$. Because $W^\top R / N$ is asymptotically equal to $W^\top \Gamma / N$ and $(\Gamma(1), U)^\top R / N$ to $(\Gamma(1), U)^\top \Gamma / N$, the claim is equivalent to saying that $\Delta$ is asymptotically equal to $1^\top 1$, where

$$\Delta = \frac{1}{N} \left\{ W^\top - B(1)A_{(1)}^{-1}(\Gamma(1), U)^\top \right\} \Pi_s \Pi^\top \left\{ W - (\Gamma(1), U)A_{(1)}^{-1} B^\top(1) \right\} \frac{1}{N}. $$

8
By using the fact that $N^{-1}\{W^T - B(1)A^{-1}(\Gamma_1), U\} = 0$, we find that $N^{-1}\{W^T - B(1)A^{-1}(\Gamma_1), U\}\Gamma_1 = 0$, where $1_s$ is the $m \times 1$ vector of 1s and $\Gamma_1$ is the first column of $\Gamma$. By using the fact that $N^{-1}\{W^T - B(1)A^{-1}(\Gamma_1), U\} = 0$, we find that $\Delta$ is asymptotically equal to $N^{-1}\{W^T - B(1)A^{-1}(\Gamma_1), U\}\Gamma_1$. The claim follows because $N^{-1}\{W^T - B(1)A^{-1}(\Gamma_1), U\}\Gamma_1 = N^{-1}\{W^T - B(1)A^{-1}(\Gamma_1), U\}\Gamma_1(1_s^2\Gamma_11_s) = N^{-1}\{W^T - B(1)A^{-1}(\Gamma_1), U\}\Delta$ is asymptotically equal to 0. As a byproduct, we see that $N^{-1}\{W^T - B(1)A^{-1}(\Gamma_1), U\}\Delta = N^{-1}\{W^T - B(1)A^{-1}(\Gamma_1), U\}\Delta = N^{-1}\{W^T - B(1)A^{-1}(\Gamma_1), U\}\Delta$ is asymptotically equal to 0. Therefore, log $\hat{Z}_{\text{L melodies}}$ and $(\hat{\theta}_1, \hat{\lambda}_{12}, \hat{\lambda}_{23})$ are asymptotically uncorrelated.

Asymptotic variance matrix of $(\hat{Z}_{\text{L melodies}}^T, \hat{Z}_{\text{L melodies}}^\phi_T)^T$ under IID sampling (known normalizing constants). The asymptotic variance matrix of $(\hat{Z}_{\text{L melodies}}^T, \hat{Z}_{\text{L melodies}}^\phi_T)^T$ can be estimated by (21) with the following modification. Replace 1 in (21) by $(\hat{Z}_{\text{L melodies}}^T, \hat{Z}_{\text{L melodies}}^\phi_T)^T$. For $h = 1, 2, 3$, replace the $(i, \theta)$th element in $W_h$ by $Nq(x_{hi}; \theta)\hat{\mu}_h(x_{hi})$. Replace $W$ by $(W, W^\phi)$ and $U$ by $(U, U^\phi)$ with

$$W^\phi = \begin{pmatrix} W_1^\phi & W_2^\phi & W_3^\phi \\ U_1^\phi & U_2^\phi & U_3^\phi \\ -V_1^\phi & V_2^\phi & V_3^\phi \end{pmatrix}, \quad U^\phi = \begin{pmatrix} U_1^\phi \\ U_2^\phi \\ -V_3^\phi \end{pmatrix},$$

where $W_h^\phi$ is the $N_h \times M_h$ matrix with $(i, \theta)$th element $N\phi(x_{hi})q(x_{hi}; \theta)^\phi_h(x_{hi})$, $U_1^\phi$ is the $N_1 \times r_{12}$ matrix with $(i, l)$ element $N\phi(x_{hi})q(x_{1i}; \theta_{12}, l)^\phi_1(x_{1i})$, $V_1^\phi$ is the $N_2 \times r_{12}$ matrix with $(i, l)$ element $N\phi(x_{2i})q(x_{2i}; \theta_{12}, l)^\phi_2(x_{2i})$, and $U_2^\phi$ and $V_2^\phi$ are $N_2 \times r_{23}$ and $N_3 \times r_{23}$ matrices similar to $U_2$ and $V_2$. 

Relationship between the likelihood, GMM, and regression estimators (known normalizing constants). Consider the estimating equations for $(\hat{Z}_{\text{L melodies}}, \hat{\theta}_1, \hat{\lambda}_{12}, \hat{\lambda}_{23})$ but set $\hat{\nu}$ to diag($\Pi_s$) and $(\hat{\lambda}_{12}, \hat{\lambda}_{23})$ to 0. Let $\hat{Z}_{\text{GMM}}$ be the resulting GMM estimator by Hansen (1982). The asymptotic variance matrix of log $\hat{Z}_{\text{GMM}}$ is

$$\frac{1}{N} \left\{ (I, 0) \Omega_{(1)}^{-1} \begin{pmatrix} I \\ 0 \end{pmatrix} \right\}^{-1}.$$
The asymptotic variance matrix of \( \log \hat{Z}_{\text{LIK}} \) is

\[
\frac{1}{N} \left[ \begin{pmatrix} I & 0 \end{pmatrix} \left\{ \begin{pmatrix} I & B_{(1)} \end{pmatrix} \begin{pmatrix} I \ 0 \end{pmatrix} \begin{pmatrix} I & 0 \end{pmatrix} \begin{pmatrix} I & B_{(1)} \end{pmatrix} \right\}^{-1} \begin{pmatrix} I & 0 \end{pmatrix} \right]^{-1},
\]

because the (1, 2) block of the matrix in the curly bracket is 0. Therefore, the asymptotic variance matrices of \( \log \hat{Z}_{\text{LIK}} \) and \( \log \hat{Z}_{\text{GMM}} \) are identical, i.e., the two estimators are asymptotically equivalent to the first order to each other.

We now discuss the relationship between the GMM estimator and the regression estimator. Consider the estimating equations \( \int Z_{\theta}^{-1} q(x; \theta) \, d\mu_h = 1 \) (\( \theta \in \Theta_h, h = 1, 2, 3 \)) and

\[
\int f(x; \theta_{hj}) \, d\mu_h = \int f(x; \theta_{11}) \, d\mu_1, \quad j = 2, \ldots, m_h, \ h = 1, 2, 3,
\]

\[
\int q(x; \theta_{12,l}) \, d\mu_1 = T_{12,l} \int q(x; \theta_{12,l}) \, d\mu_2, \quad l = 1, \ldots, r_{12},
\]

\[
\int q(x; \theta_{23,l}) \, d\mu_2 = T_{23,l} \int q(x; \theta_{23,l}) \, d\mu_3, \quad l = 1, \ldots, r_{23},
\]

where \( \mu_h(\{x_{hj}\}) = \left\{ N \sum_{j=1}^{m_h} \nu_{hj} f(x_{hi}; \theta_{hj}) \right\}^{-1} \), and \( T_{12} = (T_{12,1}, \ldots, T_{12,r_{12}})^\top \) and \( T_{23} = (T_{23,1}, \ldots, T_{23,r_{23}})^\top \) are unknowns in addition to \( Z \) and \( \nu \) with \( \sum_{j=1}^{m_h} \nu_{hj} = N_h/N \) (\( h = 1, 2, 3 \)). Let \( T = (T_{12}^\top, T_{23}^\top)^\top \). If \( \nu \) is set to \( \text{diag}(\Pi_s) \) and \( T \) to 1, then this system of estimating equations is equivalent to that used to define \( \hat{Z}_{\text{GMM}} \). By the discussion in Section 3.2.1, the solutions for \( \log Z \) and \( \log T \) are \( \log \hat{Z}_{\text{RAW}} \) and \( C^\top \log \hat{Z}_{\text{RAW}} \) respectively, where \( C^\top \) is a \( (r_{12} + r_{23}) \times M^\top \) matrix. On one hand, the asymptotic variance matrix of \( (I, C^\top)^\top \log \hat{Z}_{\text{RAW}} \) is \( U = (I, C^\top) V (I, C^\top)^\top \), where \( V \) is the asymptotic variance matrix of \( \log \hat{Z}_{\text{RAW}} \). Then the asymptotic variance matrix of \( \log \hat{Z}_{\text{REC}} \) using control variates \( C^\top \log \hat{Z}_{\text{RAW}} \) is \( \left\{ (I, 0) U^{-1} (I, 0)^\top \right\}^{-1} \). On the other hand, by the sandwich formula for general M-estimators, the asymptotic variance matrix \( U \) is

\[
\frac{1}{N} \left( \begin{pmatrix} I & * \ 0 & * \end{pmatrix} \begin{pmatrix} \Omega(3) & \begin{pmatrix} I & 0 \end{pmatrix} \end{pmatrix} \right)^{-1},
\]

where \( \Omega(3) \) is formed by deleting from \( \Omega(1) \) the row and column aligned to the \( \theta_{hi} \)th (\( h = 2, 3 \)) row and column in the (2, 2) block of \( \Omega(1) \). Then the asymptotic variance
matrix of $\log \hat{Z}_{\text{REG}}$ is

$$
\frac{1}{N} \left\{ (I, \ 0) \Omega_{(3)}^{-1} \begin{pmatrix} I \\ 0 \end{pmatrix} \right\}^{-1}.
$$

which is also the asymptotic variance matrix of $\log \hat{Z}_{\text{GMM}}$. Therefore, $\hat{Z}_{\text{REG}}$ and $\hat{Z}_{\text{GMM}}$ are asymptotically equivalent to the first order.

REFERENCES
