cemmap
centre for microdata methods and practice

Posterior inference in curved exponential families under increasing dimensions

Alexandre Belloni
Victor Chernozhukov

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP68/13

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

POSTERIOR INFERENCE IN CURVED EXPONENTIAL FAMILIES UNDER INCREASING DIMENSIONS

*

By Alexandre Belloni and Victor Chernozhukov

Duke University and Massachusetts Institute of Technology

This work studies the large sample properties of the posterior-based inference in the curved exponential family under increasing dimension. The curved structure arises from the imposition of various restrictions on the model, such as moment restrictions, and plays a fundamental role in econometrics and others branches of data analysis. We establish conditions under which the posterior distribution is approximately normal, which in turn implies various good properties of estimation and inference procedures based on the posterior. In the process we also revisit and improve upon previous results for the exponential family under increasing dimension by making use of concentration of measure. We also discuss a variety of applications to high-dimensional versions of the classical econometric models including the multinomial model with moment restrictions, seemingly unrelated regression equations, and single structural equation models. In our analysis, both the parameter dimension and the number of moments are increasing with the sample size.

Key words: curved exponential family; Bernstein-Von Mises theorems; increasing dimension; single-equation structural equations; seemingly unrelated regression; multivariate linear models; multinomial model with moment restrictions

JEL codes: C1, C3; AMS codes: 62F15

1. Introduction. The main motivation for this paper is to obtain large sample results for posterior inference in the curved exponential family under increasing dimension. In the exponential family, the log of a density is linear in parameters $\theta \in \Theta$; in the curved exponential family, these parameters θ are restricted to lie on a curve $\eta \mapsto \theta(\eta)$ parameterized by a lower dimensional parameter $\eta \in \Psi$. There are many classical examples of densities that fall in the curved exponential family; see for example Efron [13], Lehmann and Casella [21], and Bandorff-Nielsen [1]. Curved exponential densities have also been extensively used in applications [13, 18, 20, 19]. An example of

*This is revision of a 2009 ArXiv paper arXiv:0904.3132. This version was compiled on October 24, 2013.

the condition that puts a curved structure onto an exponential family is a moment restriction of the type:

$$\int m(x,\alpha)f(x,\theta)dx = 0,$$

that restricts θ to lie on a curve that can be parameterized as $\{\theta(\eta), \eta \in \Psi\}$, where component $\eta = (\alpha, \beta)$ contains α and other parameters β that are sufficient to parameterize all parameters $\theta \in \Theta$ that solve the above equation for some α. In econometric applications, often moment restrictions represent Euler equations that result from the data being an outcome of an optimization by rational decision-makers; see e.g. Hansen and Singleton [14], Chamberlain [7], Imbens [16], and Donald, Imbens and Newey [10]. In the last section of the paper we discuss in more details other econometric models that fit this framework, such as multivariate linear models, seemingly unrelated regressions, single equation structural models, as in Zellner [27] and [28]; we also discuss multinomial model with moment restrictions. Thus, the curved exponential framework is a fundamental complement to the exponential framework.

Under high-dimensionality, despite of its applicability, theoretical properties of the curved exponential family are not as well understood as the corresponding properties of the exponential family. In this paper, we contribute to the theoretical analysis of the posterior inference in curved exponential families under high dimensionality. We provide sufficient conditions under which consistency and asymptotic normality of the posterior is achieved when both the dimension of the parameter space and the sample size are large. Our framework only requires weak conditions on the prior distribution, which allows for improper priors. In particular, the uninformative prior always satisfies our assumptions. We also study the convergence of moments and the rates with which we can estimate them. We then apply these results to a variety of models where both the parameter dimension and the number of moments are increasing with the sample size.

The present analysis of the posterior inference in the curved exponential family builds upon the work of Ghosal [17] who studied posterior inference in the exponential family under increasing dimension. Under sufficient growth restrictions on the dimension of the model, it was shown that the posterior distributions concentrate in neighborhoods of the true parameter and can be approximated by an appropriate normal distribution. Such analysis extended in a fundamental way the classical results of Portnoy [24] for maximum likelihood methods for the exponential family with increasing dimensions.

In addition to a detailed treatment of the curved exponential family, we also establish some useful results for the classical exponential families. In fact, we begin our analysis revisiting Ghosal's increasing dimension setup for the exponential family. We present several results that complement the results in Ghosal [17]. First, we amend the conditions on priors to allow for a larger set of priors, for example, improper priors; second, we use concentration inequalities for log-concave densities to sharpen the conditions under which the normal approximations apply; and third, we show that the approximation of α-th order moments of the posterior by the corresponding moments of the normal density becomes exponentially difficult in the moment order α.

We also note that by establishing the asymptotic normality of the posterior distribution we can invoke results in Belloni and Chernozhukov[2] that guarantees good computational properties for MCMC methods. Moreover, new results on sampling from manifolds (see Diaconis et al [9]) permits the implementation of different random walk schemes that complement the schemes analysed in [2]. The results derived in this work can be used to weaken the conditions required in C.1-C.3 in [2] to hold in curved exponential families which should also be beneficial to the new schemes proposed in [9].

This work allows for increasing dimension so it can be thought as a sieve technique. However, this paper does not formally account for the approximation errors resulting from using approximate functional forms as opposed to exact functional forms. Approximation errors can be introduced into the model and our results can also be shown to hold under more stringent conditions (approximations errors need to vanish at rates comparable to the sampling errors), a sharp analysis of the impact of the approximation error can be delicate and is outside of the scope of the present paper. An example where approximation errors are controlled is the work Bontemps [5] where a non-parametric regression setting in the classical Gaussian mode. We view the extension of the current (non-Gaussian) setting to non-parametric cases under sharp conditions as a major venue for future work.

The rest of the paper is organized as follows. In Section 2 we formally define the framework, assumptions, and develop results for the exponential family. In Section 3, the main section, we develop the results for the curved exponential family. In Section 4 we apply our results on a variety of applications. Appendices collect proofs of the main results and technical lemmas.

Notation. For $a, b \in \mathbb{R}^d$, their (Euclidean) inner product is denoted by $\langle a, b\rangle$, and $\|a\| = \sqrt{\langle a, a\rangle}$. The unit sphere in $\mathbb{R}^d$ is denoted by $S^{d-1} = \{v \in \mathbb{R}^d : \|v\| = 1\}$. For a linear operator A, the operator norm is denoted by $\|A\| = \sup\{\|Aa\| : \|a\| = 1\}$. Let $\phi_d(\cdot; \mu; V)$ denote the d-dimensional

Gaussian density function with mean μ and covariance matrix V.

2. Exponential Family Revisited.

Assume that we are have a triangular array of random samples $\{X_1^{(n)}\ \ X_2^{(n)} \cdots\ X_n^{(n)}, n \geq 1\}$. Assume further that the data $X_i^{(n)}$, $i = 1, \ldots, n$, are independent $d^{(n)}$-dimensional vectors draws from a $d^{(n)}$-dimensional exponential family whose density is defined by

$$f\left(x;\theta^{(n)}\right) = h(x)\exp\Big(\left\langle x, \theta^{(n)}\right\rangle - \psi^{(n)}\left(\theta^{(n)}\right)\Big), \tag{2.1}$$

where $\theta^{(n)} \in \Theta^{(n)}$ an open convex set of $\mathbb{R}^{d^{(n)}}$, $\psi^{(n)}$ is the associate normalizing function and $h^{(n)}$ depends only on the data. Let $\theta_0^{(n)} \in \Theta^{(n)}$ denote the (sequence of) true parameter which is assumed to be bounded away from the boundary of $\Theta^{(n)}$ (uniformly in n). For notational convenience we will suppress the superscript $^{(n)}$ but it is understood that the associate objects are changing with n.

Under this framework, the posterior density of θ given the observed data $\{X_i\}_{i=1}^n$ is defined as

$$\pi_n(\theta) = \frac{\pi(\theta)\prod_{i=1}^n f(X_i;\theta)}{\int_\Theta \pi(\xi)\prod_{i=1}^n f(X_i;\xi)d\xi} = \frac{\pi(\theta)\exp\left(\langle\sum_{i=1}^n X_i, \theta\rangle - n\psi(\theta)\right)}{\int_\Theta \pi(\xi)\exp\left(\langle\sum_{i=1}^n X_i, \xi\rangle - n\psi(\xi)\right)d\xi}, \tag{2.2}$$

where $\pi(\cdot)$ denotes a prior distribution on Θ.

Our results are stated in terms of a re-centered Gaussian distribution in the local parameter space. Let $\mu = \psi'(\theta_0)$ and $F = \psi''(\theta_0)$ be the mean and covariance matrix of X_i, and let $J = F^{1/2}$ be its square root (i.e., $JJ^T = F$). The re-centering is defined as $\Delta_n := \sqrt{n}J^{-1}\left(\frac{1}{n}\sum_{i=1}^n X_i - \mu\right)$; it follows that $E[\Delta_n] = 0$, and $E[\Delta_n\Delta_n^T] = I_d$ where I_d denotes the d-dimensional identity matrix. Moreover, the posterior in the local parameter space is defined for $u \in \mathcal{U} = \sqrt{n}J(\Theta - \theta_0)$ as

$$\pi^*(u) = \frac{\pi(\theta_0 + n^{-1/2}J^{-1}u)\prod_{i=1}^n f(X_i;\theta_0 + n^{-1/2}J^{-1}u)}{\int_\mathcal{U} \pi(\theta_0 + n^{-1/2}J^{-1}u)\prod_{i=1}^n f(X_i;\theta_0 + n^{-1/2}J^{-1}u)du}. \tag{2.3}$$

In the same lines of Portnoy [24] and Ghosal [17], conditions on the growth rates of the third and fourth moments are imposed. Following these references the following quantities play an important role in the analysis:

$$
\begin{aligned}
(2.4)\ B_{1n}(c) &= \sup_{\theta,a}\left\{ \left|E_\theta\left[\langle a,V\rangle^3\right]\right| : a \in S^{d-1}, \|J(\theta-\theta_0)\|^2 \le \frac{cd}{n}\right\},\\
(2.5)\ B_{2n}(c) &= \sup_{\theta,a}\left\{ E_\theta\left[\langle a,V\rangle^4\right] : a \in S^{d-1}, \|J(\theta-\theta_0)\|^2 \le \frac{cd}{n}\right\},\\
(2.6)\quad \lambda_n(c) &= \frac{1}{6}\left(\sqrt{\frac{cd}{n}}B_{1n}(0) + \frac{cd}{n}B_{2n}(c)\right),
\end{aligned}
$$

where V is a random variable distributed as $J^{-1}(U - E_\theta[U])$ and U has density $f(\cdot;\theta)$ as defined in (2.1).

Comment 2.1 *Note that $\lambda_n(c)$ is different than the quantity with the same notation defined in [17] and we provide a technical discussion about the differences in the Appendix. For now we note that (2.6) is always smaller than its counterpart in [17] which leads to weaker requirements. In the specific applications of interest we develop bounds on (2.6). In the case that the density (2.1) is log-concave in the data, we provide generic bounds for (2.6) in Appendix D.*

We will need to impose some regularity conditions on the prior π.

Assumption P(c_n). For the specified positive sequence c_n, the prior density function π satisfies:

$$\sup_{\theta\in\Theta} \ln[\pi(\theta)/\pi(\theta_0)] \le O(d) \text{ and } |\ln\pi(\theta) - \ln\pi(\theta_0)| \le K_n(c_n)\|\theta-\theta_0\|$$

for any θ s.t. $\|\theta-\theta_0\| \le \sqrt{c_n\|F^{-1}\|d/n}$, with $K_n(c_n)\sqrt{c_n\|F^{-1}\|d/n} = o(1)$.

In what follows the sequence c_n will typically remain uniformly bounded in n. These conditions differ from the ones imposed in [17]. Although the same Lipschitz condition is assumed, we require only a relative lower bound on the value of the prior on the true parameter instead of an absolute bound. Thus this condition requires that the true parameter does not have an exponentially small prior value relative to other parameter values. We note that such conditions allow for improper priors which were not allowed in [17]. Importantly, the uninformative prior trivially satisfies Assumption P.

Next we state the main results of this section.

Theorem 1 *For any fixed value $c > 0$, suppose that (i) $B_{1n}(c)\sqrt{d/n} = o(1)$, (ii) $\lambda_n(c)d = o(1)$, (iii) $\|F^{-1}\|d/n = o(1)$, and (iv) Assumption P(c) hold.*

Then we have asymptotic normality of the posterior density function

$$\int_{\mathcal{U}} |\pi_n^*(u) - \phi_d(u;\Delta_n, I_d)|du = o_P(1).$$

Theorem 1 has different assumptions on the prior that Theorem 3 of [17] has. However, Theorem 1 does not require additional technical assumptions used in [17], as discussed in Appendix A, and the growth condition of d with relative to the sample size n is improved by at least $\ln d$ factors.

In some applications it might be desired to have stronger convergence properties than simply asymptotic normality. The following theorem provides sufficient conditions for the α-moment convergence. In what follows, for sequences of α and d, let $M_{d,\alpha} := (d+\alpha)\left(1 + \frac{\alpha \ln(d+\alpha)}{d+\alpha}\right)$.

Theorem 2 *In addition to the conditions (i) and (iii) of Theorem 1, suppose that the following hold for any fixed* $\bar{c}$*:*
(ii′) $\lambda_n(\bar{c}M_{d,\alpha}/d)\,[\bar{c}M_{d,\alpha}]^{1+\alpha/2} = o(1)$*;*
(iv′) Assumption $P(\bar{c}M_{d,\alpha}/d)$ *and* $K_n(\bar{c}M_{d,\alpha}/d)\sqrt{\|F^{-1}\|\left[\bar{c}M_{d,\alpha}\right]^{1+\alpha}/n} = o(1)$*.*
Then we have

$$\int_{\mathcal{U}} \|u\|^\alpha |\pi_n^*(u) - \phi_d(u;\Delta_n, I_d)|du = o_P(1). \tag{2.7}$$

Conditions (ii') and (iv') lead to the strengthening of conditions (ii) and (iv) of Theorem 1 respectively. We emphasize that Theorem 2 allows for α and d to grow as the sample size increases. Our conditions highlight the polynomial trade off between n and d but an exponential trade off between n and α. This suggests that the estimation of higher moments in increasing dimensions applications could be very delicate. Conditions (ii') and (iv') simplify significantly if $\alpha \ln d = o(d)$, in which case $M_{d,\alpha} = d(1+o(1))$.

Comment 2.2 *Suppose that we are interested in allowing* α *to grow with the sample size as well. If* d *is growing in a polynomial rate with respect to* n*, our results do not allow for* $\alpha = O(\ln n)$*. Some limitation along these lines should be expected since there is an exponential trade off between* α *and* n*. However, it is definitely possible to let both the dimension and* α *to grow with the sample size with the rate* $\alpha = O(\sqrt{\ln n})$*. Such slow growth conditions illustrate the potential limitations for the practical estimation of higher order moments.*

3. Curved Exponential Family. Next we consider the curved exponential family. Let $X_1, X_2, \ldots, X_n$ be i.i.d. observations from a d-dimensional curved exponential family with density given by

$$f(x;\theta) = h(x)\exp\left(\langle x, \theta(\eta)\rangle - \psi(\theta(\eta))\right),$$

where $\eta \in \Psi \subset \mathbb{R}^{d_1}$, $\theta : \Psi \to \Theta$, an open subset of $\mathbb{R}^d$, and $d \to \infty$ as $n \to \infty$ as before. In this section we assume that $J = I_d$ for notational convenience.

The parameter of interest is η, whose true value η_0 lies in the interior of the set $\Psi \subset \mathbb{R}^{d_1}$. The true value of θ induced by η_0 is given by $\theta_0 = \theta(\eta_0)$. The mapping $\eta \mapsto \theta(\eta)$ takes values from $\mathbb{R}^{d_1}$ to $\mathbb{R}^d$ where $d_1 \leq d$. Moreover, assume that η_0 is the unique solution to the system $\theta(\eta) = \theta_0$.

Thus, the parameter θ corresponds to a high-dimensional parametrization, and η describes the lower-dimensional parametrization of the density. We require the following regularity conditions on the mapping $\theta(\cdot)$.

Assumption A. For every κ, and uniformly in $\gamma \in B(0, \kappa\sqrt{d})$, there exists a linear operator $G : \mathbb{R}^{d_1} \to \mathbb{R}^d$ such that $G'G$ has eigenvalues bounded from above and away from zero, and for every n

$$\sqrt{n}\left(\theta(\eta_0 + \gamma/\sqrt{n}) - \theta(\eta_0)\right) = r_{1n} + (I + R_{2n})G\gamma, \tag{3.8}$$

where

$$\|r_{1n}\| d^{1/2} \to 0 \quad \text{and} \quad \|R_{2n}\| d \to 0. \tag{3.9}$$

Assumption B. There exist a strictly positive constants ε_0 such that for every $\eta \in \Psi$ (uniformly on n) we have

$$\|\theta(\eta) - \theta(\eta_0)\| \geq \varepsilon_0 \|\eta - \eta_0\|. \tag{3.10}$$

Assumption P'. The prior density function π satisfies:

$$\sup_{\eta \in \Psi} \ln[\pi(\theta(\eta))/\pi(\theta(\eta_0))] \leq O(d_p).$$

Thus the mapping $\eta \mapsto \theta(\eta)$ is allowed to be nonlinear and discontinuous. For example, the additional condition of $\delta_{1n} = 0$ implies the continuity of the mapping in a neighborhood of η_0. More generally, condition (3.9) impose that the map admits an (uniform) approximate linearization in the neighborhood of η_0. An example of a kind of map allowed in this framework is given in Figure 1.

A prior π on Θ induces a prior over Ψ as $\pi(\eta) = \pi(\theta(\eta))/\int_\Psi \pi(\theta(\tilde{\eta}))d\tilde{\eta}$. Alternatively the prior can be placed directly over Ψ. Assumption P' also bounds the maximum log-likelihood given by the prior to any η different

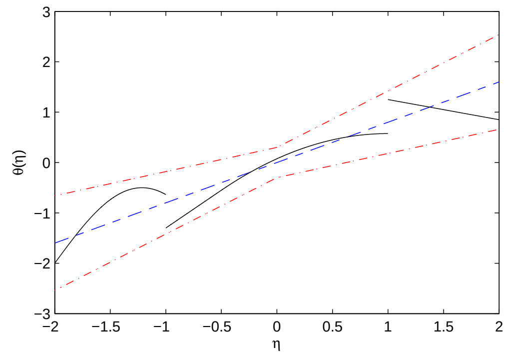

FIG 1. *This figure illustrates the mapping $\theta(\cdot)$. The (discontinuous) solid line is the mapping while the dash line represents the linear map induced by G. The dash-dot line represents the deviation band controlled by r_{1n} and R_{2n}.*

than η_0 to be of the order d_p which can grow with n. If Assumption P holds we have $d_p \leq d$. However, if the prior is placed directly on Ψ we typically have $d_p = d_1$. Finally, if the prior is uninformative we trivially have $d_p = 1$. The posterior of η given the data is denoted by

$$\pi_n(\eta) \propto \pi(\theta(\eta)) \cdot \prod_{i=1}^{n} f(x_i;\theta(\eta)) = \pi(\theta(\eta)) \cdot \exp\left(n\left\langle \bar{X}, \theta(\eta)\right\rangle - n\psi(\theta(\eta))\right)$$

where $\bar{X} = (1/n)\sum_{i=1}^{n} X_i$.

Under this framework, we also define the local parameter space to describe contiguous deviations from the true parameter as

$$\gamma = \sqrt{n}(\eta - \eta_0), \quad \text{and let} \quad s = (G'G)^{-1}G'\sqrt{n}(\bar{X} - \mu)$$

be a first order approximation to the normalized maximum liklelihood/extremum estimate. It follows that the following bounds hold for s:

$$E[s] = 0, \quad E[ss'] = (G'G)^{-1}, \quad \text{and} \quad \|s\| = O_P(\sqrt{d_1}).$$

The posterior density evaluated at $\gamma \in \Gamma := \sqrt{n}(\Psi - \eta_0)$ is given by $\pi_n^*(\gamma) = \ell(\gamma)/\int_\Gamma \ell(\gamma)d\gamma$, where

(3.11)
$$\ell(\gamma) = \exp\left(n\left\langle \bar{X}, \theta(\eta_0 + n^{-1/2}\gamma) - \theta(\eta_0)\right\rangle - n\left[\psi(\theta(\eta_0 + n^{-1/2}\gamma)) - \psi(\theta(\eta_0))\right]\right)$$
$$\times \pi\left(\theta\left(\eta_0 + n^{-1/2}\gamma\right)\right).$$

In order to formally state our results we need the following additional definition

$$a_n = \sup\{c : \lambda_n(c) \le 1/16\}.$$

Because $\lambda_n(c)$ can be used to bound deviations between the posterior distribution from a suitable Gaussian distribution (Lemma 1) it follows that in a neighborhood of size $\sqrt{a_n d}$ we can still bound the posterior $\ell(\cdot)$ by above with a proper Gaussian density.

Next we address the consistency question for the maximum likelihood estimator associated with the curved exponential family.

Theorem 3 *Suppose that Assumptions A, B and P' hold, $a_n \to \infty$. Then the maximum likelihood estimator $\widehat{\eta}$ satisfies*

$$\|\widehat{\eta} - \eta_0\| = O_P\left(\sqrt{\frac{d_1 + d_p}{n}}\right).$$

Two remarks regarding Theorem 3 are worth mentioning. First, we note that the condition $\lambda_n(c) = o(1)$ implies $a_n \to \infty$. However, $\lambda_n(c) = o(1)$ is stronger than the condition $\sqrt{d/n}B_{1n}(c) = o(1)$ used for consistency for the exponential case obtained in Ghosal [17]. Second, consistency result in Theorem 3 relies on the dimension of the larger model d and very mild assumptions on the prior were made in Theorem 3. If the prior used is defined over the full space, it might place an exponentially small (in the dimension d) weight in η_0 relative to other points. In the case $d_1 \sim d$ this is not problematic but it could impact the rates if $d_1 = o(d)$. In cases a prior can be placed directly over Γ so that $d_p = O(d_1)$, we obtain the standard rate of convergence of $\sqrt{d_1/n}$.

Finally, we can state the asymptotic normality result for the curved exponential family.

Theorem 4 *Suppose that Assumptions A, B, and P' hold, $\log d = o(a_n)$, and conditions (i)-(iv) of Theorem 1 hold. Then, asymptotic normality for the posterior density associated with the curved exponential family holds,*

$$\int |\pi_n^*(\gamma) - \phi_{d_1}(\gamma; s, (G'G)^{-1})| d\gamma = o_P(1).$$

4. Applications to Selected Econometric Models. In this section we verify the conditions that lead to asymptotic normality in a variety of econometric problems covering both exponential and curved exponential families under increasing dimension. Most examples are motivated by the classical work of Zellner [28] on Bayesian econometrics.

4.1. *Multivariate Linear Model.* Next we consider a multivariate linear model. The response variable Y_i is a d_r-dimensional vector, the disturbances U_i are normally distributed with mean zero and covariance matrix Σ_0. The covariates Z_i are d_c-dimensional and the parameter matrix of interest Π_0 is $d_c \times d_r$,

$$Y_i = Z_i \Pi_0 + U_i \quad i = 1, \ldots, n. \tag{4.12}$$

For notational convenience, let Y and Z denote the matrices whose rows are given by y_i and Z_i respectively. Note that the dimension of the model is $d = d_r^2 + d_c d_r$.

Conditioning on the covariates Z, this model can be cast as an exponential family model by the following parametrization, see for instance [26],

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2}\Sigma^{-1} \\ \Pi\Sigma^{-1} \end{pmatrix}, \quad \bar{X} = \frac{1}{n}\begin{pmatrix} \bar{X}_1 \\ \bar{X}_2 \end{pmatrix} = \frac{1}{n}\begin{pmatrix} Y'Y \\ Z'Y \end{pmatrix} \tag{4.13}$$

and using the (trace) inner product $\langle \theta, X \rangle = \text{trace}(X_1'\theta_1) + \text{trace}(X_2'\theta_2)$. This parametrization leads to the normalizing function

$$\psi(\theta) = -\frac{1}{4n}\text{trace}(Z\theta_2\theta_1^{-1}\theta_2' Z') - \frac{1}{2}\log\det(-2\theta_1). \tag{4.14}$$

We make the following assumptions on the design. The covariates z_i satisfy $\max_{i \le n} \|Z_i\| = O(d_c^{1/2})$, the matrices $Z'Z/n$ and Σ_0 have eigenvalues bounded away from zero and from above, and the matrix Π_0 has full rank with singular values also bounded away from zero and from above uniformly in n.

Under these assumptions, by Lemma 10 it follows that $\|F^{-1}\| = O(1)$. Also, Lemma 11 bounds the quantities $B_{1n}(c) = O(d_c)$ and $B_{2n}(c) = O(d_c^2)$. Therefore we have asymptotic normality by Theorem 1 provided that the condition $d(d_c\sqrt{d/n} + d_c^2 d/n) = o(1)$ holds.

4.2. *Seemingly Unrelated Regression Equations.* The seemingly unrelated regression model (Zellner [27]) considers a collection of d_r models

$$y_k = X_k \beta_k + u_k, \quad k = 1, \ldots, d_r, \tag{4.15}$$

each having n observations, and the dimension of β_k is d_k. Let d_c denote the total number of distinct covariates. The d_r-dimensional vector of disturbances u has zero mean and covariance Σ_0. This model can be written in the form of (4.12) by setting $\Pi = [\pi_1(\beta_1); \pi_2(\beta_2); \cdots; \pi_{d_r}(\beta_{d_r})]$. Note that the vector $\pi_i(\beta_i)$ has zeros for regressors that do not appear in the ith model. Garderen [26] shows that this model is a curved exponential model provided that the matrix Π has some zero restrictions.

Consider the assumptions of Section 4.12. In this case we have that

$$\eta = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} \Sigma^{-1} \\ \Pi \end{pmatrix}, \quad \theta(\eta) = \begin{pmatrix} -\frac{1}{2}\Sigma^{-1} \\ \Pi\Sigma^{-1} \end{pmatrix}. \tag{4.16}$$

We restrict the space of Σ to consider $\lambda_{min}(\Sigma) > \lambda_{min}$ a fixed constant (note that this induces $\lambda_{max}(\Sigma^{-1}) < 1/\lambda_{min}$ which leads to a convex region in the parameter space), and that operator norm of Π is bounded by a constant, $\|\Pi\| \leq M$ a fixed constant.

The mapping $\theta(\cdot)$ is twice differentiable and Lemma 12 establishes that condition (3.8) holds with $R_{2n} = 0$ and $\|r_{1n}\| \leq O(d/\sqrt{n})$. This implies that the requirement of $d_r^6 + d_c^3 d_r^3 = o(n)$ suffices for Assumption A to hold.

In order to verify Assumption B, we have

$$\|\theta(\eta) - \theta(\eta_0)\| \quad \geq \quad \max\{\|\eta_1 - \eta_{01}\|, \|\eta_2\eta_1 - \eta_{02}\eta_{01}\|\}.$$

By setting $\varepsilon_0 = \lambda_{min}(\Sigma_0)/4M$ we can assume that $\|\eta_1 - \eta_{01}\| < \varepsilon_0\|\eta - \eta_0\|$ otherwise Assumption B holds. This assumption leads to $\|\eta_2 - \eta_{02}\| \geq (1/2)\|\eta - \eta_0\|$. In this case, since the operator norm of η_2 satisfies $\|\eta_2\| \leq M$, we have

$$\begin{aligned} \|\eta_2\eta_1 - \eta_{02}\eta_{01}\| &= \|\eta_2(\eta_1 - \eta_{01}) + (\eta_2 - \eta_{02})\eta_{01}\| \\ &\geq \|\eta_2 - \eta_{02}\|\lambda_{min}(\Sigma_0) - \|\eta_2\|\|\eta_1 - \eta_{01}\| \\ &\geq \|\eta - \eta_0\|\lambda_{min}(\Sigma_0)/2 - M\varepsilon_0\|\eta - \eta_0\| \\ &\geq \|\eta - \eta_0\|\lambda_{min}(\Sigma_0)/4 \geq \varepsilon_0\|\eta - \eta_0\| \end{aligned}$$

which implies Assumption B.

4.3. *Single Structural Equation Model.* Next we consider the single structural equation,

$$y^{(1)} = Y^{(2)}\beta + z^{(1)}\gamma + v \tag{4.17}$$

for which the associated reduced form system, given by the multivariate linear model in (4.12), can be partitioned as

$$(y^{(1)} \quad Y^{(2)}) = (z^{(1)} \quad Z^{(2)}) \begin{pmatrix} \pi_{11} & \Pi_{12} \\ \pi_{21} & \Pi_{22} \end{pmatrix} + (u^{(1)} \quad U^{(2)}). \tag{4.18}$$

We assume full column rank of Z and $\text{rank}(\pi_{21} \;\; \Pi_{22}) = \text{rank}(\Pi_{22}) = d_r - 1$ where d_r is the dimension of $(y^{(1)} \;\; Y^{(2)})$. The compatibility between the models (4.17) and (4.18) requires that

$$\pi_{11} = \Pi_{12}\beta + \gamma, \quad \pi_{21} = \Pi_{22}\beta, \quad \text{and} \quad u^{(1)} = U^{(2)}\beta + v.$$

The model can also be embedded in (4.12) as follows

(4.19)

$$\eta = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} = \begin{pmatrix} \Sigma^{-1} \\ \begin{pmatrix} \Pi_{12} \\ \Pi_{22} \end{pmatrix} \\ \begin{pmatrix} \gamma \\ \beta \end{pmatrix} \end{pmatrix}, \quad \theta(\eta) = \begin{pmatrix} -\frac{1}{2}\Sigma^{-1} \\ \begin{pmatrix} \gamma + \Pi_{12}\beta & \Pi_{12} \\ \Pi_{22}\beta & \Pi_{22} \end{pmatrix} \Sigma^{-1} \end{pmatrix}.$$

Similar arguments to those used in Section 4.2 show that Assumptions A and B holds.

4.4. *Multinomial Model.* This example of multinomial model was also analyzed in [17]. Our goal is to weaken some of the conditions required previously using the techniques proposed here.

Let $\mathcal{X} = \{x^0, x^1, \ldots, x^d\}$ be the known finite support of a multinomial random variable X where d is allowed to grow with sample size n. For each j denote by p_j the probability of the event $\{X = x^j\}$ which is assumed to satisfy $\max_{0\leq j\leq d} 1/p_j = O(d)$. The parameter space is given by $\theta = (\theta_1, \ldots, \theta_d)$ where $\theta_j = \log(p_j/(1 - \sum_{k=1}^d p_k))$. It follows that under the assumption on the p_j's the true value of θ_j's are bounded. The Fisher information matrix is given by $F = P - pp'$ where $P = \text{diag}(p)$. In this case we have $B_{1n}(c) = O(d^{3/2})$ and $B_{2n}(c) = O(d^2)$. We refer to [17] for detailed calculations.

The growth condition $d^6(\log d)/n \to 0$ was imposed in [17] to obtain the asymptotic normality results (the case of $\alpha = 0$). We weaken this growth requirement by combining the derivation in [17] with the analysis in Section 2 with an uninformative (improper) prior. In this case we have $K_n(c) = 0$ and our definition of λ_n remove the logarithmic factors. As a result, Theorem 1 leads to a weaker growth condition $d^4/n \to 0$. For α-moment estimation, the conditions of Theorem 2 are satisfied with the condition that $d^{4+\alpha+\delta}/n \to 0$ for any strictly positive value of δ. Recently another approach based on Le Cam's proof that is specific to discrete probability distributions allows for further improvements, see [6].

4.5. *Multinomial Model with Moment Restrictions.* In this subsection we provide a high-level discussion of the multinomial model with moment restrictions. Let $\mathcal{X} = \{x^0, x^1, x^2, \ldots, x^d\}$ be the known finite support of a multinomial random variable X which was described in Section 2. Conditions $(i) - (iv)$ are verified as in Section 4.4.

As discussed in the introduction, it is of interest to incorporate moment restrictions into this model, see Chamberlain [7] and Imbens [16] for discussions. This will lead to a curved exponential model as studied in Section 3.

The parameter of interest is $\eta \in \Psi \subset \mathbb{R}^{d_1}$ a compact set. Consider a (twice continuously differentiable) vector-valued moment function $m : \mathcal{X} \times \Psi \to \mathbb{R}^M$ such that

$$E[m(X,\eta)] = 0 \text{ for a unique } \eta_0 \in \Psi.$$

The log-likelihood function associated with this model

$$\begin{aligned} & l(q,\eta) = \sum_{i=1}^{n}\sum_{j=0}^{d} I\{X_i = x_j\} \ln q_j \\ & \text{for some } \theta \text{ and } \eta \text{ such that } \sum_{j=0}^{d} q_j m(x_j,\eta) = 0,\ \sum_{j=0}^{d} q_j = 1. \end{aligned} \tag{4.20}$$

and $l(q,\eta) = -\infty$ if the probability distribution q violates any of the moments conditions. The log-likelihood function (4.20) induces the mapping $q : \Psi \to \Delta^{d-1}$ formally defined as

$$\begin{aligned} q(\eta) = \arg\max_{q} \quad & l(q,\eta) \\ & \sum_{j=0}^{d} q_j m(x_j,\eta) = 0,\ \sum_{j=0}^{d} q_j = 1,\ q \geq 0. \end{aligned} \tag{4.21}$$

In this case, the function $\theta_j(\eta) = \log(q_j(\eta)/q_0(\eta))$ (for $j = 1,\ldots,d$) is the mapping from $\Psi \to \Theta$ discussed in Section 3. Assuming that the matrix $E\left[m(X,\eta)m(X,\eta)'\right]$ is uniformly positive definite over η, Qin and Lawless [25] use the inverse function theorem to show that $\theta(\cdot)$ is a twice continuous differentiable mapping of η in a neighborhood of η_0. In particular this implies that we can take $R_{2n} = 0$ and $\|r_{1n}\| = O\Big(dd_1^2(d/n)\Big)$. Thus, Assumption A holds provided that $d^{4.5} = o(n)$.

Assumption B is satisfied if η belongs in a compact set Ψ and that the mapping $\theta(\cdot)$ is injective (over a set that contains Ψ in its interior). We refer to Newey and McFadden [23] for a discussion of primitive assumptions for identification with moment restrictions.

APPENDIX A: TECHNICAL RESULTS

In this section we prove the technical lemmas needed to prove our main result in the following section. Our exposition follows the work of Ghosal [17]. For the sake of completeness we include Proposition 1, which can be found in Portnoy [24], and a specialized version of Lemma 1 of [17]. All the remaining proofs use different techniques and rely on weaker assumptions. In particular, we no longer require the prior to be proper, no bounds on the growth of $\det(\psi''(\theta_0))$ are imposed, and $\ln n$ and $\ln d$ do not need to be of the same order.

Using the notation in Section 2, let

$$H(a) = \{u \in \mathcal{U} : \|u\| \leq a\}. \tag{A.22}$$

Moreover, for $u \in \mathcal{U}$ let

$$\tilde{Z}_n(u) = \exp\left(\langle u, \Delta_n\rangle - \|u\|^2/2\right) \quad \text{and} \tag{A.23}$$

(A.24)

$$Z_n(u) = \exp\left(\frac{1}{\sqrt{n}}\left\langle \sum_{i=1}^{n} X_i, J^{-1}u\right\rangle - n\left[\psi\left(\theta_0 + n^{-1/2}J^{-1}u\right) - \psi(\theta_0)\right]\right),$$

otherwise (if $\theta_0 + n^{-1/2}J^{-1}u \notin \Theta$), let $Z_n(u) = \tilde{Z}_n(u) = 0$. The quantity (A.24) denotes the likelihood ratio associated with f as a function of u. In a parallel manner, (A.23) is associated with a standard Gaussian density. We note that (A.23) and (A.24) are logconcave functions.

We start recalling a result on the Taylor expansion of ψ which is key to control deviations between $\tilde{Z}(u)$ and $Z(u)$.

Proposition 1 (Portnoy [24]) *Let ψ' and ψ'' denote respectively the gradient and the Hessian of ψ. For any $\theta, \theta_0 \in \Theta$, there exists $\tilde{\theta} = \lambda\theta + (1-\lambda)\theta_0$, for some $\lambda \in [0,1]$, such that*

$$\begin{aligned} \psi(\theta) &= \psi(\theta_0) + \langle \psi'(\theta_0), \theta - \theta_0\rangle + \tfrac{1}{2}\langle \theta - \theta_0, \psi''(\theta_0)(\theta - \theta_0)\rangle + \\ &+ \tfrac{1}{6}E_{\theta_0}\left[\langle \theta - \theta_0, W\rangle^3\right] \\ &+ \tfrac{1}{24}\left\{E_{\tilde{\theta}}\left[\langle \theta - \theta_0, W\rangle^4\right] - 3\left(E_{\tilde{\theta}}\left[\langle \theta - \theta_0, W\rangle^2\right]\right)^2\right\} \end{aligned} \tag{A.25}$$

where $E_\theta[g(W)]$ denotes the expectation of $g(U - E_\theta[U])$ with $U \sim f(\cdot;\theta)$.

Based on Proposition 1 we control the pointwise deviation between Z_n and $\tilde{Z}_n$ in a neighborhood of zero (i.e., in a neighborhood of the true parameter).

Lemma 1 (Essentially in Ghosal [17] or Portnoy [24]) *For all u such that $\|u\| \leq \sqrt{cd}$, we have*

$$|\ln Z_n(u) - \ln \tilde{Z}_n(u)| \leq \lambda_n(c)\|u\|^2 \ \ \textit{and} \ \ \ln Z_n(u) \leq \langle \Delta_n, u\rangle - \frac{1}{2}\|u\|^2(1 - 2\lambda_n(c)).$$

Proof. Let $(I) = |\ln \tilde{Z}_n(u) - \ln Z_n(u)| = n|\psi(\theta_0 + n^{-1/2}J^{-1}u) - \psi(\theta_0)|$. Using Proposition 1, by the mean value theorem, for some $\tilde{\theta} \in [\theta_0, \theta_0 + n^{-1/2}J^{-1}u]$ we have

$$
\begin{aligned}
(I) \quad &\leq \quad n\left|\tfrac{1}{6}E_{\theta_0}\left[\left\langle \tfrac{u}{n^{1/2}}, V\right\rangle^3\right]\right| \\
&+ \quad \tfrac{1}{24}\left|\left\{E_{\tilde{\theta}}\left[\left\langle \tfrac{u}{n^{1/2}}, V\right\rangle^4 - 3\left(E_{\tilde{\theta}}\left[\left\langle \tfrac{u}{n^{1/2}}, V\right\rangle^2\right]\right)^2\right]\right\}\right| \\
&\leq \quad \tfrac{1}{6}\left(n^{-1/2}\|u\|^3 B_{1n}(0) + n^{-1}\|u\|^4 B_{2n}(c)\right) \leq \lambda_n(c)\|u\|^2.
\end{aligned}
$$

The second inequality follows directly from the first result. ■

Next we show how to bound the integrated deviation between the quantities in (A.23) and (A.24) restricted to the neighborhood of zero.

Lemma 2 *For any $c > 0$ we have*

$$
\left(\int \tilde{Z}_n(u)du\right)^{-1} \int_{\{u:\|u\|\leq\sqrt{cd}\}} |Z_n(u) - \tilde{Z}_n(u)|du \leq cd\lambda_n(c)e^{2cd\lambda_n(c)}
$$

Proof. Using $|e^x - e^y| \leq |x-y|\max\{e^x, e^y\}$ and Lemma 1, since $\|u\| \leq \sqrt{cd}$ we have

$$
\begin{aligned}
|Z_n(u) - \tilde{Z}_n(u)| \quad &\leq \quad |\ln Z_n(u) - \ln \tilde{Z}_n(u)| \exp\left(\langle \Delta_n, u\rangle - \tfrac{1}{2}(1 - 2\lambda_n(c))\|u\|^2\right) \\
&\leq \quad \lambda_n(c)\|u\|^2 \exp\left(\langle \Delta_n, u\rangle - \tfrac{1}{2}(1 - 2\lambda_n(c))\|u\|^2\right).
\end{aligned}
$$

By integrating over the set $H(\sqrt{cd})$ as defined in (A.22), we obtain

$$
\begin{aligned}
&\int_{H(\sqrt{cd})} |Z_n(u) - \tilde{Z}_n(u)|du \\
&\leq \int_{H(\sqrt{cd})} \lambda_n(c)\|u\|^2 \exp\left(\langle \Delta_n, u\rangle - \frac{1}{2}(1 - 2\lambda_n(c))\|u\|^2\right) du \\
&\leq cd\lambda_n(c) \int_{H(\sqrt{cd})} \exp\left(\langle \Delta_n, u\rangle - \frac{1}{2}(1 - 2\lambda_n(c))\|u\|^2\right) du \\
&\leq cd\lambda_n(c)e^{2cd\lambda_n(c)} \int_{H(\sqrt{cd})} \exp\left(\langle \Delta_n, u\rangle - \frac{1}{2}\|u\|^2\right) du \\
&\leq cd\lambda_n(c)e^{2cd\lambda_n(c)} \int \tilde{Z}_n(u)du.
\end{aligned}
$$

■

The next lemma controls the tail of Z_n relatively to $\tilde{Z}_n$. In order to achieve that it makes use of a concentration inequality for log-concave densities functions developed by Lovász and Vempala in [22]. The lemma is stated with a given bound on the norm of Δ_n which is allowed to grow with the dimension. Such bound on Δ_n can be easily obtained with probability arbitrary close to one by standard arguments.

Lemma 3 *Suppose that* $\|\Delta_n\|^2 < C_1 d$ *and* $\lambda_n(c) < 1/8$ *for some* $c > 16[4C_1 \vee 2]$. *Then for every* $k \geq 1$ *we have*

$$\begin{aligned} &\int_{\{u:\|u\|\geq k\sqrt{cd}\}} \pi(\theta_0 + n^{-1/2}J^{-1}u) Z_n(u)du \\ &\leq \sup_u \pi(u) \left(e^{cd\lambda_n(c)} \int \tilde{Z}_n(u)du \right) e^{-kcd/16}. \end{aligned}$$

Proof. Firs we prove some technial results. By definition of c, we have $\Delta_n \in H(\sqrt{cd})$ and $u \in H^c(k\sqrt{cd})$. Define $\tilde{u} = \sqrt{cd}u/\|u\|$. Since Z_n and $\tilde{Z}_n$ are logconcave functions we have

$$\begin{aligned} \log Z_n(u) &\leq \log Z_n(\tilde{u}) + \nabla[\log(Z_n(\tilde{u}))]'(u - \tilde{u}) \\ &= \log Z_n(\tilde{u}) + (\|u\| - \sqrt{cd})\nabla[\log(Z_n(\tilde{u}))]'u/\|u\| \end{aligned} \tag{A.26}$$

where ∇ is the gradient. Next note by Lemma 1

$$\begin{aligned} \log Z_n(\tilde{u}) &\leq \log \tilde{Z}_n(\tilde{u}) + \lambda_n(c)\|\tilde{u}\|^2 \\ &= \Delta_n'\tilde{u} - (1/2)(1 - 2\lambda_n(c))\|\tilde{u}\|^2 \\ &\leq \sqrt{C_q d}\sqrt{cd} - (1/2)(1 - 2\lambda_n(c))\|\tilde{u}\|^2 \\ &= -cd(1/2 - \lambda_n(c) - \sqrt{C_1/c}) = -cd\varphi \end{aligned} \tag{A.27}$$

where $\varphi := (1/2 - \lambda_n(c) - \sqrt{C_1/c}) \geq (1/2) - (1/8) - (1/8) = 1/4$.

Using the logconcavity of Z_n and Lemma 1 again we have

$$\begin{aligned} 0 = \log Z_n(0) &\leq \log Z_n(\tilde{u}) + \nabla[\log(Z_n(\tilde{u}))]'(0 - \tilde{u}) \\ &\leq \sqrt{C_1 d}\sqrt{cd} - cd(1 - 2\lambda_n(c))/2 - \sqrt{cd}\nabla[\log(Z_n(\tilde{u}))]'u/\|u\| \end{aligned}$$

so that

$$\nabla[\log(Z_n(\tilde{u}))]'u/\|u\| \leq -\sqrt{cd}\varphi. \tag{A.28}$$

Using (A.27) and (A.28) in the bound (A.26), since $\|u\| - \sqrt{cd} \geq (k-1)\sqrt{cd}$, we have for any $u \in H^c(k\sqrt{cd})$

$$\log Z_n(u) \leq -cd\varphi - \sqrt{cd}\varphi(k-1)\sqrt{cd} \leq -cdk\varphi.$$

Thus we can apply Lemma 5.16 [22] with $\beta := \frac{d}{d-1}ck\varphi \geq 2k \geq 2$ and $M_f \geq Z_n(0) = 1$, and we obtain

$$\begin{aligned} \int_{H^c(k\sqrt{cd})} Z_n(u)du &\leq (e^{1-\beta}\beta)^{d-1} \int Z_n(u)du \\ &\leq (e^{1-\beta}\beta)^{d-1} 2 \int_{H(\sqrt{cd})} Z_n(u)du \\ &\leq (e^{1-\beta}\beta)^{d-1} 2e^{cd\lambda_n(c)} \int_{H(\sqrt{cd})} \tilde{Z}_n(u)du \end{aligned} \tag{A.29}$$

where we used that $\int Z_n(u)du \leq 2\int_{H(\sqrt{cd})} Z_n(u)du$ (note that k does not appear).

To prove the statement of the lemma, we have

$$\int_{H(k\sqrt{cd})^c} \pi(\theta_0 + n^{-1/2}J^{-1}u)Z_n(u)du \leq \sup_{H(k\sqrt{cd})^c} \pi(\theta_0 + n^{-1/2}J^{-1}u) \int_{H(k\sqrt{cd})^c} Z_n(u)du$$

and the result follows by (A.29) and noting

$$\begin{aligned}(e^{1-\beta}\beta)^{d-1}2e^{cd\lambda_n(c)} &= 2e^{d-1}([d/(d-1)]ck\varphi)^{d-1}e^{cd\lambda_n(c)}e^{-ckd\varphi} \\ &\leq \exp\left(d + d\log(2ck\varphi) + cd\lambda_n(c) - ckd\varphi\right) \\ &\leq \exp\left(-ckd\varphi/4\right)\end{aligned}$$

since $1 + \log(2x) \leq x/4$ for $x \geq 19$ and $\lambda_n(c) \leq \varphi/2$. ■

We note that the value of c in the previous lemma could depend on n as long as the condition is satisfied. In fact, we can have c as large as

$$a_n := \sup\{c : \lambda_n(c) < 1/16\}. \tag{A.30}$$

(A.30) characterizes a neighborhood of size $\sqrt{a_n d}$ on which the quantity $Z_n(\cdot)$ can still be bounded by a proper Gaussian. Essentially, Lemma 3 bounds the contribution outside this neighborhood.

We close this section with a technical lemma that combines some of the previous results to be easily applied.

Lemma 4 *Under* $r_{1n} := K_n(c)\sqrt{\|F^{-1}\|cd/n} = o(1)$, $\sup_\theta \pi(\theta)/\pi(\theta_0) \leq \exp(wd)$, $r_{2n} := [1/16 - \lambda_n(c) - w/c] > 0$, $r_{3n} := cd\lambda_n(c) = o(1)$ *we have:*

(1) $\sup_{\|u\|\leq\sqrt{cd}} \left|\frac{\pi(\theta_0+n^{-1/2}J^{-1}u)}{\pi(\theta_0)} - 1\right| \leq (1 + o(1))r_{1n}$
(2) $\int \pi(\theta_0 + n^{-1/2}J^{-1}u)Z_n(u)du = [1 + O(r_{1n} + e^{-r_{2n}d} + r_{3n})] \int \pi(\theta_0)\tilde{Z}_n(u)du,$

Proof. To show (1) note that for $x = o(1)$, we have $\exp(x) = 1+x(1+o(1))$. Since $K_n(c)\sqrt{\|F^{-1}\|cd/n} = o(1)$ we have

$$\begin{aligned}\sup_{\|u\|\leq\sqrt{cd}} \left|\frac{\pi(\theta_0+n^{-1/2}J^{-1}u)}{\pi(\theta_0)} - 1\right| &\leq \left|\exp\left(K_n(c)\sqrt{\|F^{-1}\|cd/n}\right) - 1\right| \\ &= (1 + o(1))K_n(c)\sqrt{\|F^{-1}\|cd/n}.\end{aligned}$$

To show (2), let $r_{1n} := K_n(c)\sqrt{\|F^{-1}\|cd/n}$, $\Lambda = \{u : \|u\| \leq \sqrt{cd}\}$ for c sufficiently large. Using (1) we have

$$\int \pi(\theta_0 + n^{-1/2}J^{-1}u)Z_n(u)du = (1 + r_{1n}[1 + o(1)]) \int_\Lambda \pi(\theta_0)Z_n(u)du+$$

$$+O\left(\sup_u \pi(\theta_0 + n^{-1/2}J^{-1}u)\int_{\Lambda^c} Z_n(u)du\right).$$

By assumption we have $\sup_u \pi(\theta_0 + n^{-1/2}J^{-1}u) \leq \pi(\theta_0)\exp(wd)$, and by Lemma 3 we have that for c large enough

$$\sup_u \pi(\theta_0+n^{-1/2}J^{-1}u)\int_{\Lambda^c} Z_n(d)du \leq \exp(-cd[1/16-\lambda_n(c)-w/c])\int \pi(\theta_0)\tilde{Z}_n(u)du$$

where by $c > 16w$ and $\lambda_n(c) = o(1)$, $\exp(-cd[1/16 - \lambda_n(c) - w/c]) = o(1)$ since $d \to \infty$ (alternatively, if d is fixed we could let $c \to \infty$ slowly enough so that $\lambda_n(c) = o(1)$).

By Lemma 2 we have

$$\int_\Lambda |Z_n(u) - \tilde{Z}_n(u)|du \leq cd\lambda_n(c)\exp(2cd\lambda_n(c))\int \tilde{Z}_n(u)du$$

where $cd\lambda_n(c)\exp(2cd\lambda_n(c)) = O(r_{3n})$ since $r_{3n} = o(1)$. ∎

APPENDIX B: PROOF OF THEOREMS 1 AND 2

Armed with Lemmas 1, 2, 3, and 4, we now show asymptotic normality and moments convergence results (respectively Theorems 1 and 2) under the appropriate growth conditions of the dimension of the parameter space with respect to the sample size.

It is easy to see that Theorem 1 follows from Theorem 2 with $\alpha = 0$, therefore its proof is omitted.

Proof of Theorem 2. By definition of $M_{d,\alpha}$, we use that $\sqrt{\bar{c}M_{d,\alpha}} \geq 4\|\Delta_n\|$ in the analysis which is not restrictive since $\|\Delta_n\| = O_P(\sqrt{d})$.

We will divide the integral of (2.7) in two regions

$$\Lambda = \left\{u \in \mathbb{R}^d : \|u\| \leq \sqrt{\bar{c}M_{d,\alpha}}\right\} \quad \text{and} \quad \Lambda^c,$$

where $\bar{c}$ is a sufficiently large constant (independent of n). We note that the conditions in Lemma 4 are satisfied with $cd = \bar{c}M_{d,\alpha}$ for n sufficiently large. We use the notation $r_{1n} := K_n(\bar{c}M_{d,\alpha}/d)\sqrt{\|F^{-1}\|\bar{c}M_{d,\alpha}/n}$, $\sup_\theta \pi(\theta)/\pi(\theta_0) \leq \exp(wd)$, $r_{2n} := [1/16-\lambda_n(\bar{c}M_{d,\alpha}/d)-w/\bar{c}] > 0$, and $r_{3n} := \bar{c}M_{d,\alpha}\lambda_n(\bar{c}M_{d,\alpha}/d)$.

Note that by construction

$$\pi_n^*(u) = \frac{\pi(\theta_0 + n^{-1/2}J^{-1}u)Z_n(u)}{\int \pi(\theta_0 + n^{-1/2}J^{-1}u')Z_n(u')du'} \quad \text{and} \quad \phi(u; \Delta_n, I_d) = \frac{\tilde{Z}_n(u)}{\int \tilde{Z}_n(u')du'}.$$

To simplify the notation we let $I := \int \pi(\theta_0 + n^{-1/2}J^{-1}u')Z_n(u')du'$, $\tilde{I} := \int \tilde{Z}_n(u')du'$, $\pi_u(u) := \pi(\theta_0 + n^{-1/2}J^{-1}u)$ and $\pi_u(0) := \pi(\theta_0)$. Thus have

$$\begin{array}{l}
\int \|u\|^\alpha \left| \frac{\pi_u(u)Z_n(u)}{I} - \frac{\tilde{Z}_n(u)}{\tilde{I}} \right| du \\
\leq \int_\Lambda \|u\|^\alpha \frac{\pi_u(u)Z_n(u)}{I} \left| \frac{\pi_u(0)}{\pi_u(u)} - 1 \right| du + \\
+ \int_\Lambda \|u\|^\alpha \left| \frac{\pi_u(0)Z_n(u)}{I} - \frac{\pi_u(0)\tilde{Z}_n(u)}{\pi_u(0)\tilde{I}} \right| du \\
+ \sup_u \pi_u(u) \int_{\Lambda^c} \|u\|^\alpha \frac{Z_n(u)}{I} du \\
+ \int_{\Lambda^c} \|u\|^\alpha \frac{\tilde{Z}_n(u)}{\tilde{I}} du
\end{array} \tag{B.31}$$

where we want to prove that the LHS is $o_P(1)$. Next we bound each of the four terms on the RHS.

To bound the first integral in the RHS of (B.31), by Lemma 4, since $r_{1n} = o(1)$, we have

$$\sup_{u \in \Lambda} \left| \frac{\pi(\theta_0)}{\pi(\theta_0 + n^{-1/2}J^{-1}u)} - 1 \right| \leq (1 + o(1))r_{1n}.$$

Thus,

$$\begin{array}{l}
\int_\Lambda \|u\|^\alpha \frac{\pi_u(u)Z_n(u)}{I} \left| \frac{\pi_u(0)}{\pi_u(u)} - 1 \right| du \\
\leq (1 + o(1))r_{1n} \int_\Lambda \|u\|^\alpha \frac{\pi_u(u)Z_n(u)}{I} du \\
\leq (1 + o(1))r_{1n}(\bar{c}M_{d,\alpha})^{\alpha/2} \int_\Lambda \frac{\pi_u(u)Z_n(u)}{I} du \\
\leq (1 + o(1))r_{1n}(\bar{c}M_{d,\alpha})^{\alpha/2} = o(1)
\end{array}$$

where the last relation follows from (iv').

To bound the second integral in (B.31), note that

$$\begin{array}{ll}
\int_\Lambda \|u\|^\alpha \left| \frac{\pi_u(0)Z_n(u)}{I} - \frac{\pi_u(0)\tilde{Z}_n(u)}{\pi_u(0)\tilde{I}} \right| du & \leq \int_\Lambda \|u\|^\alpha \frac{\pi_u(0)Z_n(u)}{I} \left| \frac{\pi_u(0)\tilde{I} - I}{\pi_u(0)\tilde{I}} \right| du + \\
& + (1/\tilde{I}) \int_\Lambda \|u\|^\alpha \left| Z_n(u) - \tilde{Z}_n(u) \right| du.
\end{array}$$

By Lemma 4 part (2), $I = [1 + O(r_{1n} + e^{-r_{2n}d} + r_{3n})]\pi(\theta_0)\tilde{I}$. By Lemma 2 we have

$$\int_\Lambda \left| Z_n(u) - \tilde{Z}_n(u) \right| du \leq r_{3n} \exp(2r_{3n})\tilde{I}.$$

Also using Lemma 4 part (1) we have

$$\begin{array}{l}
\int_\Lambda \|u\|^\alpha \left| \frac{\pi_u(0)Z_n(u)}{I} - \frac{\pi_u(0)\tilde{Z}_n(u)}{\pi_u(0)\tilde{I}} \right| du \\
\leq (\bar{c}M_{d,\alpha})^{\alpha/2} O(r_{1n} + e^{-r_{2n}\bar{c}M_{d,\alpha}} + r_{3n}) = o(1)
\end{array}$$

under our conditions.

To bound the third term in the RHS of (B.31), let

$$\Lambda_k^c := \left\{u : \|u\| \in \left[k\sqrt{\bar{c}M_{d,\alpha}}, (k+1)\sqrt{\bar{c}M_{d,\alpha}}\right]\right\}.$$

For each k, using Lemma 3, and subsequently Lemma 4 part (2), we have

$$\begin{aligned}\int_{\Lambda_k^c} Z_n(u)du &\leq \exp(r_{3n} - k\bar{c}M_{d,\alpha}/16)\int \tilde{Z}_n(u)du \\ &\leq \exp(r_{3n} - k\bar{c}M_{d,\alpha}/16)[1 + O(r_{1n} + e^{-r_{2n}d} + r_{3n})]I/\pi_u(0).\end{aligned}$$

Thus,

(B.32)

$$\begin{aligned}&\sup_u \pi_u(u)\int_{\Lambda^c} \|u\|^\alpha \frac{Z_n(u)}{I}du \leq \sup_u \pi_u(u)\sum_{k=1}^{\infty}\left\{(k+1)^\alpha \bar{c}^{\alpha/2}M_{d,\alpha}^{\alpha/2}\int_{\Lambda_k^c} Z_n(u)du\right\} \\ &\leq \sup_u \frac{\pi_u(u)}{\pi_u(0)}\bar{c}^{\alpha/2}M_{d,\alpha}^{\alpha/2}W\sum_{k=1}^{\infty}(k+1)^\alpha \exp(-k\bar{c}M_{d,\alpha}/16)\end{aligned}$$

where $W = [1 + O(r_{1n} + e^{-r_{2n}d} + r_{3n})]\exp(r_{3n}) = 1 + o(1)$. Note also that $\sup_u \pi_u(u)/\pi_u(0) \leq \exp(wd)$. Since $M_{d,\alpha} > \max\{1, \alpha\}$, by choosing $\bar{c}$ large enough we have

$$\sum_{k=1}^{\infty}(k+1)^\alpha e^{-k\bar{c}M_{d,\alpha}/16} \leq e^{-\bar{c}M_{d,\alpha}/20}.$$

Moreover, our definition of $M_{d,\alpha}$ also implies that the RHS of (B.32) satisfies

$$\begin{aligned}&W\exp(wd)\bar{c}^{\alpha/2}M_{d,\alpha}^{\alpha/2}e^{-\bar{c}M_{d,\alpha}/20} \\ &= W\exp\left(wd + \tfrac{\alpha}{2}(\ln\bar{c} + \ln M_{d,\alpha}) - \bar{c}M_{d,\alpha}/20\right) \\ &\leq W\exp\left(-\bar{c}M_{d,\alpha}/40\right) = o(1)\end{aligned}$$

provided that $\bar{c}$ is large enough.

Finally, the last integral in (B.31) converges to zero by standard bounds on Gaussian densities for an appropriate choice of the constant $\bar{c}$ (note that $\bar{c}$ can be chosen independently of d and α). ■

APPENDIX C: PROOFS OF SECTION 3

For $\gamma \in \Gamma = \sqrt{n}(\Psi - \eta_0)$ let $u_\gamma = n^{1/2}[\theta(\eta_0 + n^{-1/2}\gamma) - \theta(\eta_0)] \in \mathcal{U} \subset \mathrm{I\!R}^d$, and we write

$$Z_n(u_\gamma) := Z_n\left(n^{1/2}\left[\theta\left(\eta_0 + n^{-1/2}\gamma\right) - \theta(\eta_0)\right]\right) = \frac{\ell(\gamma)}{\pi\left(\theta\left(\eta_0 + n^{-1/2}\gamma\right)\right)}$$

where Z_n is defined in (A.24).

In the next lemma it is required that $\log d = o(a_n)$ which is a substantially weaker condition than the one used in [17] for establishing asymptotic normality for the posterior of (regular) exponential densities, $\lambda_n(c\log d)d = o(1)$.

Lemma 5 *Assume that requirements (i),(ii), (iii), and (iv) hold. Suppose Assumptions A, B and P'(0) holds. In addition, suppose that* $\log d = o(a_n)$. *Then, for some constant* $\bar{k}$ *independent of* n, *we have*

$$\int_{\Gamma\setminus B(0,\bar{k}\sqrt{d_1+d_p})} \pi\left(\theta(\eta_0)+n^{-1/2}u_\gamma\right) Z_n(u_\gamma)d\gamma$$
$$\leq o\left(\int_\Gamma \pi\left(\theta(\eta_0)+n^{-1/2}u_\gamma\right) Z_n(u_\gamma)d\gamma\right)$$

Comment C.1 *The only assumption made on* d_1 *in Lemma 5 was that* $d_1 \leq d$. *If* $d_1\log d = o(d)$ *the proof simplifies significantly since there is no need to define region (II) in the proof. Moreover, we see the role of the prior via* $d_p \leq d$. *Strengthening the assumption on the prior the same proof allow for the integrand on the left hand side to be* $\Gamma\setminus B(0,\bar{k}\sqrt{d_1})$ *instead of potentially* $\Gamma\setminus B(0,\bar{k}\sqrt{d})$ *in the result.*

Proof of Lemma 5. Divide Γ into three regions:

$$(I) := B\left(0,\bar{k}N_G\right), \quad (II) := \{\gamma : \max\{\|\gamma\|,\|u_\gamma\|\} \leq \bar{k}N_T\}\setminus B\left(0,\bar{k}N_G\right),$$
$$(III) := \Gamma\setminus\Big((I)\cup(II)\Big),$$

where $\bar{k}$ is chosen later to be large enough independent of the dimensions d or d_1. Region (I) is defined to be the region where the linear approximation G for $\theta(\cdot)$ is valid in the sense of Assumption A. Region (III) represents the tail of the distribution; either γ or u_γ has large norm. Finally, region (II) is an intermediary region for which G is not a valid approximation but we still have interesting guarantees for deviations from normality. We point out that regions (II) or (III) might be non-convex. We will derive sufficient conditions on the values of N_G and N_T as a function of d and d_1. It will be sufficient to set $N_G = \sqrt{d_1+d_p}/\sqrt{2}$ and $N_T = \sqrt{d}\log d$.

For notational convenience we define $c_G = \bar{k}^2N_G^2/d$ and $c_T = \bar{k}^2N_T^2/d$. Our assumptions are such that

$$d\lambda_n(c_G)\to 0 \ \text{ and } \ \lambda_n(c_T) < 1/16.$$

We first bound the contribution of region (III). For any $\gamma\in(III)$, define $\tilde{u}_\gamma = \bar{k}N_G\frac{u_\gamma}{\|u_\gamma\|}\in\mathcal{U}$. Using Lemma 1 we have

$$\ln Z_n(\tilde{u}_\gamma) \leq \langle\Delta_n,\tilde{u}_\gamma\rangle - \frac{1}{2}(1-2\lambda_n(c_G))\|\tilde{u}_\gamma\|^2.$$

Since $\log Z_n(\cdot)$ is globally concave in $\mathcal{U}$ and $\log Z_n(0) = 0$, we have
(C.33)

$$\begin{array}{l} \log Z_n(u_\gamma) \leq \frac{\|u_\gamma\|}{\|\tilde{u}_\gamma\|} \ln Z_n(\tilde{u}_\gamma) \\ \leq \frac{\|u_\gamma\|}{\bar{k}N_G} \left(\langle \Delta_n, r_{1n} + (I + R_{2n})G\gamma \rangle - \frac{1-2\lambda_n(c_G)}{2} \|\tilde{u}_\gamma\|^2 \right) \\ \leq \frac{\|u_\gamma\|}{\bar{k}N_G} \left(\|\Delta_n\| \|r_{1n}\| + \|\Delta_n\| \|GR_{2n}\gamma\| + \langle G'\Delta_n, \gamma \rangle - \frac{1-2\lambda_n(c_G)}{2} \|\tilde{u}_\gamma\|^2 \right) \\ \leq \frac{\|u_\gamma\|}{\bar{k}N_G} \left(\frac{\bar{k}N_G}{5} \bar{k}N_G - \frac{1-2\lambda_n(c_G)}{2} \bar{k}^2 N_G^2 \right) \\ \leq -\|u_\gamma\| N_G \ \frac{\bar{k}}{5} \end{array}$$

by choosing $\bar{k}$ large enough such that $\|G'\Delta_n\| < \bar{k}N_G/5$, $\|\Delta_n\|\|r_{1n}\| = O(d^{1/2}\|r_{1n}\|) = o(1)$, $\|\Delta_n\|\|GR_{2n}\gamma\| = O(d^{1/2}N_G\|R_{2n}) = o(1)$, and using that $\lambda_n(c_G) \leq \lambda_n(c_T) \leq 1/16$. The contribution of (III) can be bounded by

$$\begin{array}{l} \displaystyle\int_{(III)} \pi\left(\theta(\eta_0) + n^{-1/2} u_\gamma\right) Z_n(u_\gamma) d\gamma \\ \displaystyle\leq \pi(\theta_0) \sup_{\eta \in \Psi} \frac{\pi(\theta(\eta))}{\pi(\theta(\eta_0))} \int_{(III)} \exp\left(-\frac{\bar{k}N_G}{5}\|u_\gamma\|\right) d\gamma, \end{array}$$

where $\sup_{\eta\in\Psi} \pi(\theta(\eta))/\pi(\theta(\eta_0)) \leq \exp(wd_p)$ for a constant w associated with Assumption P'.

The integral on the right can be bounded as follows

$$\begin{array}{rcl} \int_{(III)} \exp\left(-\frac{\bar{k}N_G}{5}\|u_\gamma\|\right) d\gamma & \leq & \int_{B(0,\bar{k}N_T)\cap(III)} \exp\left(-\frac{\bar{k}N_G}{5}\|u_\gamma\|\right) d\gamma \\ & + & \int_{B(0,\bar{k}N_T)^c} \exp\left(-\frac{\bar{k}N_G}{5}\|u_\gamma\|\right) d\gamma \end{array}$$

and recall that u_γ is a function of γ. However, by definition of the regions, $\gamma \in B(0, N_T) \cap (III)$ implies $\|u_\gamma\| \geq \bar{k}N_T$, and $\gamma \in B(0, N_T)^c$ implies that $\|u_\gamma\| \geq \varepsilon_0 N_T$ by Assumption B. Direct calculations to bound the integrals yields

$$\begin{array}{l} \int_{(III)} \exp\left(-\frac{\bar{k}N_G}{5}\|u_\gamma\|\right) d\gamma \\ \leq \exp(-\bar{k}^2 N_G N_T/5)(\bar{k}N_T)^{d_1} \text{Vol}_{d_1}(B_{d_1}(0,1)) + \\ +(5/\bar{k}N_G)^{d_1-1} \text{Vol}_{d-1}(S^{d-1}(0,1)) \exp(-\bar{k}\varepsilon_0 N_G N_T/3 + 2)\Gamma(d_1) \\ \leq \exp\left(-\frac{\bar{k}^2}{5} N_G N_T + d_1 \ln(\bar{k}N_T)\right) + \\ +\exp\left(-\frac{\bar{k}}{3}\varepsilon_0 N_G N_T + d_1 \ln d_1\right) \end{array}$$

where Vol_k denote the k-dimensional volume of a set and $\Gamma(\cdot)$ is the gamma function so that $\text{Vol}_{d_1-1}(S^{d_1-1}(0,1)) = 2\pi^{d_1/2}/\Gamma(d_1/2)$, $\text{Vol}_{d_1}(B_{d_1}(0,1)) = \pi^{d_1/2}/\Gamma(d_1/2+1)$, and $\text{Vol}_{d_1-1}(S^{d_1-1}(0,1)) = d_1\text{Vol}_{d_1}(B_{d_1}(0,1))$.

Using the assumption on the prior, we can bound the contribution of (III) by

$$\pi(\theta_0)\exp\big(3 + wd_p + d_1\log d_1 + d_1\log(\bar{k}N_T) - \bar{k}\varepsilon_0 N_G N_T/3\big) \tag{C.34}$$

if $\bar{k} \geq 2\varepsilon_0$.

Next consider $\gamma \in (II)$. By definition $\gamma \in B(0, \bar{k}N_T) \setminus B(0, \bar{k}N_G)$. Under the assumption that $\lambda_n(c_T) < 1/16$, we have that

$$\ln Z_n(u_\gamma) \leq \langle \Delta_n, u_\gamma\rangle - \frac{1}{2}\frac{7}{8}\|u_\gamma\|^2.$$

Moreover, by Assumption B, we have $\|u_\gamma\| \geq \varepsilon_0\|\gamma\|$. Therefore, by choosing $\bar{k}$ such that $\bar{k}N_G > 8\|G'\Delta_n\|$, Assumption A, we have

$$\begin{array}{l}\int_{(II)} \pi\left(\theta(\eta_0) + n^{-1/2}u_\gamma\right) Z_n(u_\gamma)d\gamma \\ \leq \pi(\theta_0)\left(\sup_{\eta\in\Psi}\frac{\pi(\theta(\eta))}{\pi(\theta(\eta_0))}\right)\int_{(II)}\exp\left(\langle\Delta_n, u_\gamma\rangle - \frac{1}{2}\frac{7}{8}\|u_\gamma\|^2\right)d\gamma \\ \leq \pi(\theta_0)\left(\sup_{\eta\in\Psi}\frac{\pi(\theta(\eta))}{\pi(\theta(\eta_0))}\right)\int_{(II)}\exp\left(-\frac{1}{2}\frac{7}{16}\|u_\gamma\|^2\right)d\gamma \\ \leq \pi(\theta_0)\left(\sup_{\eta\in\Psi}\frac{\pi(\theta(\eta))}{\pi(\theta(\eta_0))}\right)\exp(-\varepsilon_0^2\bar{k}^2N_G^2/5)(4/\varepsilon_0)^{d_1}\end{array}$$

where we used standard bounds to Gaussian densities. Using our assumption on the prior, we can bound the contribution of (II) by

$$\pi(\theta_0)\exp\Big(wd_p + d_1\ln(4/\varepsilon_0) - \varepsilon_0^2\bar{k}^2N_G^2/5\Big). \tag{C.35}$$

Finally, we show a lower bound on the integral over (I). First note that for any $\gamma \in (I)$ condition (3.9) holds and we have $u_\gamma = r_{1n} + (I + R_{2n})G\gamma$. Therefore, $u_\gamma \in B(0, \|G\|(1 + \|R_{2n}\|)\bar{k}N_G + \|r_{1n}\|) \subset B(0, 2\|G\|\bar{k}N_G)$. For simplicity, let $c_{(I)} = 4\|G\|^2N_G^2/d$ so that

$$\begin{array}{l}\int_{(I)} \pi\left(\theta(\eta_0) + n^{-1/2}u_\gamma\right) Z_n(u_\gamma)d\gamma \\ \geq \pi(\theta_0)\exp\left(-K_n(c_{(I)})\sqrt{\frac{c_{(I)}d}{n}}\right)\int_{(I)} Z_n(u_\gamma)d\gamma.\end{array}$$

Under our assumptions $\exp\left(-K_n(c_{(I)})\sqrt{\frac{c_{(I)}d}{n}}\right) \to 1$. Furthermore, using (3.10), $\|\Delta_n\| = O(\sqrt{d})$, and $\|\gamma\| \leq \bar{k}N_G$, we have

$$\begin{array}{rcl}\ln Z_n(r_{1n} + (I + R_{2n})G\gamma) & = & \langle\Delta_n, r_{1n} + (I + R_{2n})G\gamma\rangle - \\ & & -\frac{1+2\lambda_n(c_{(I)})}{2}\|r_{1n} + (I + R_{2n})G\gamma\|^2 \\ & \geq & o(1) + \langle\Delta_n, G\gamma\rangle - \frac{1+2\lambda_n(c_{(I)})}{2}\|G\gamma\|^2.\end{array}$$

Therefore we have

$$\begin{array}{l}\int_{(I)} \pi\left(\theta(\eta_0)+n^{-1/2}u_\gamma\right) Z_n(u_\gamma)d\gamma \\ \geq \pi(\theta_0) O\left(\int_{(I)} \exp\left(\langle \Delta_n, G\gamma\rangle - \frac{1+2\lambda_n(\bar{k}^2)}{2}\|G\gamma\|^2\right) d\gamma\right) \\ \geq \pi(\theta_0) O\left((1-2\lambda_n\left(c_{(I)}\right))^{d_1/2} \det(G'G)^{-1/2}\right) \\ \geq \pi(\theta_0) O\left(\exp(-\|G\|d_1)\right).\end{array}$$

The choices stated in the beginning $N_G = \sqrt{d_1+d_p}/\sqrt{2}$ and $N_T = \sqrt{d}\log d$ yields the result since we have $d \geq \max\{d_1, d_p\}$ and $\bar{k}$ can be taken to be sufficiently large (independent of d and n). ■

Proof of Theorem 3. Note that $\sup_{\eta\in\Psi} |\log[\pi(\theta(\eta))/\pi(\theta(\eta_0))]| \leq w d_p$. Let $\widehat{\gamma}$ be such that $\widehat{\eta} = \eta_0 + n^{1/2}\widehat{\gamma}$. We will show that $\ln Z(u_\gamma) < -c\{d_1+d_p\}$ with $c > 2w$ for any $\gamma \notin B(0, \bar{k}\sqrt{d_1+d_p})$ where $\bar{k}$ is sufficiently large. Therefore, since $\log Z_n(0) = 0$, the MLE $\widehat{\gamma} \in B(0, \bar{k}\sqrt{d_1+d_p})$ and the result follows.

Take any $\gamma \notin B(0, \bar{k}\sqrt{d_1+d_p})$, using (C.33) where $N_G = \sqrt{d_1+d_p}/\sqrt{2}$, we have

$$\ln Z(u_\gamma) < -\|u_\gamma\| N_G \bar{k}/5 \leq -\varepsilon_0 \bar{k}^2\{d_1+d_p\}/5\sqrt{2}$$

where the last inequality follows from Assumption B so that $\|u_\gamma\| \geq \varepsilon_0\|\gamma\| \geq \varepsilon\bar{k}\sqrt{d_1+d_p}$. As stated earlier, the result follows by choosing $\bar{k}$ sufficiently large. ■

Proof of Theorem 4. We have that

$$\begin{array}{rl}\int |\pi_n^*(\gamma) - \phi_{d_1}(\gamma; s, (G'G)^{-1})| d\gamma & \leq \int_{B(0,\bar{k}\sqrt{d})} |\pi_n^*(\gamma) - \phi_{d_1}(\gamma; s, (G'G)^{-1})| d\gamma + \\ & \int_{\Gamma\setminus B(0,\bar{k}\sqrt{d})} \pi_n^*(\gamma) + \phi_{d_1}(\gamma; s, (G'G)^{-1}) d\gamma.\end{array}$$

The main step of the proof is to show that the second term is negligible for $\bar{k}$ large enough. This follows from Lemma 5 and known results for Gaussian densities under the assumed conditions. Thus, we can restrict our analysis to $B(0, \bar{k}\sqrt{d})$.

The remaining of the proof follows the same steps in the proof of Theorem 2 since Assumption A ensures that the linearization in (3.8) is sufficiently precise in the region $B(0, \bar{k}\sqrt{d})$ under (3.9).

■

APPENDIX D: BOUND ON $\lambda_n(c)$ WHEN f IS LOGCONCAVE

In this section we derive a new bound on the fundamental quantity

$$\lambda_n(c) = \frac{1}{6}\left(\sqrt{\frac{cd}{n}} B_{1n}(0) + \frac{cd}{n} B_{2n}(c)\right)$$

when the density function (2.1) is log-concave in the data. We start by restating the following theorem for log-concave distributions.

Lemma 6 (Lovász and Vempala [22]) *If X is a random vector from a log-concave distribution in $\mathbb{R}^d$ then*

$$E\left[\|X\|^k\right]^{1/k} \leq 2kE\left[\|X\|\right] \leq 2kE\left[\|X\|^2\right]^{1/2}.$$

This result provides a reverse direction of the Holder inequality which will allow us to control higher moments based on the second moment. Since we will be bounding moments from random variables in the exponential family we can apply Lemma 6.

In what follows we consider $\theta \in \mathcal{R}_c = \{\theta \in \Theta : \|J^{-1}(\theta - \theta_0)\| \leq \sqrt{cd/n}\}$, $U \sim f_\theta = f(\cdot;\theta)$, and let $H_\theta = E_\theta[(U - E_\theta[U])(U - E_\theta[U])']^{1/2}$. In this notation $J = H_{\theta_0}$.

We first bound the third moment term $B_{1n}(0)$. In this case, since the variable of interest $\langle a, V\rangle$ is properly normalized to have unit variance, its third moment is bounded by a constant.

Lemma 7 (Bound on B_{1n}) *Suppose that $f(\cdot;\theta_0)$ is a logconcave distribution. Then we have that $B_{1n}(0) \leq 6^3$.*

Proof. Let $V = J^{-1}(U - E[U])$ where $U \sim f_{\theta_0}$. Therefore V has a logconcave density function, $E[V] = 0$, and $E[VV'] = I_d$. Using Lemma 6, we have

$$B_{1n}(0) \leq \sup_{\|a\|=1} E_{\theta_0}\left[|\langle a, V\rangle|^3\right] \leq 6^3 \sup_{\|a\|=1} E\left[|\langle a, V\rangle|^2\right]^{3/2} = 6^3.$$

■

Before we proceed to bound the term B_{2n} in λ_n we state and prove the following technical lemma.

Lemma 8 *Let X be a random vector in $\mathbb{R}^d$ and M be a $d \times d$ matrix. We have that*

$$\sup_{\|a\|=1} E\left[|\langle a, MX\rangle|^k\right] \leq \|M\|^k \sup_{\|a\|=1} E\left[|\langle a, X\rangle|^k\right]$$

Proof. Let $\bar{a}$ achieve the supremum on the left hand side. Then we have

$$\begin{array}{rcl} E\left[|\langle \bar{a}, MX\rangle|^k\right] & = & E\left[|\langle M'\bar{a}, X\rangle|^k\right] = \|M'\bar{a}\|^k E\left[|\left\langle \frac{M'\bar{a}}{\|M'\bar{a}\|}, X\right\rangle|^k\right] \\ & \leq & \|M'\|^k\|\bar{a}\|^k E\left[|\left\langle \frac{M'\bar{a}}{\|M'\bar{a}\|}, X\right\rangle|^k\right] \\ & \leq & \|M\|^k \sup_{\|a\|=1} E\left[|\langle a, X\rangle|^k\right] \end{array}$$

since $\|\bar{a}\| = 1$ and $M'\bar{a}/\|M'\bar{a}\| = 1$. ■

Unlike Lemma 7, we need to bound the forth moment in a vanishing neighborhood of θ_0. This will require an additional assumption that H_θ becomes sufficiently close to J for any θ in this neighborhood of θ_0. This additional condition is c than the conditions of Theorem 2.4 in [17].

Lemma 9 *Suppose that $f(\cdot;\theta)$ is a logconcave distribution and assume that $\|I - H_\theta^{-1}J\| < 1/2$ (in the operator norm). Then we have that*

$$\sup_{\|a\|=1} E_\theta\left[|\langle a, V\rangle|^k\right] \leq 2^{2k}\cdot k^k.$$

Proof. By convexity of $t \mapsto t^k$ ($k \geq 1$) we have $(t+s)^k \leq 2^{k-1}\left(t^k + s^k\right)$, and Lemma 8 yields

$$\begin{aligned}
\sup_{\|a\|=1} E_\theta\left[|\langle a, V\rangle|^k\right] &= \sup_{\|a\|=1} E_\theta\left[|\left\langle a, (I - H_\theta^{-1}J + H_\theta^{-1}J)V\right\rangle|^k\right] \\
&\leq 2^{k-1}\sup_{\|a\|=1} E_\theta\left[|\left\langle a, (I - H_\theta^{-1}J)V\right\rangle|^k\right] + \\
&+ 2^{k-1}\sup_{\|a\|=1} E_\theta\left[|\left\langle a, H_\theta^{-1}JV\right\rangle|^k\right] \\
&\leq 2^{k-1}\|I - H^{-1}J\|^k \sup_{\|a\|=1} E_\theta\left[|\langle a, V\rangle|^k\right] + \\
&+ 2^{k-1}\sup_{\|a\|=1} E_\theta\left[|\left\langle a, H_\theta^{-1}JV\right\rangle|^k\right].
\end{aligned}$$

Using that $\|I - H_\theta^{-1}J\| < 1/2$ we have

$$\sup_{\|a\|=1} E_\theta\left[|\langle a, V\rangle|^k\right] \leq 2^k \sup_{\|a\|=1} E_\theta\left[|\left\langle a, H_\theta^{-1}JV\right\rangle|^k\right].$$

Now we invoke Lemma 6 to obtain

$$\sup_{\|a\|=1} E_\theta\left[|\langle a, V\rangle|^k\right] \leq 2^k\cdot(2k)^k \sup_{\|a\|=1} E_\theta\left[|\left\langle a, H_\theta^{-1}JV\right\rangle|^2\right]^{k/2} = 2^{2k}\cdot k^k$$

since $E_\theta\left[(H_\theta^{-1}JV)(H_\theta^{-1}JV)'\right] = I$.
■

Corollary 1 (Bound on $B_{2n}(c)$) *Suppose that $f(\cdot;\theta)$ is a logconcave distribution and assume that $\|I - H_\theta^{-1}J\| < 1/2$ for any $\theta \in \mathcal{R}_c$. Then we have that $B_{2n}(c) \leq 2^{16}$.*

APPENDIX E: AUXILIARY RESULTS FOR SECTION 4

Lemma 10 *In the multivariate linear model, the information matrix satisfies*

$$\lambda_{min}(F) \geq \frac{1}{4}\frac{\lambda_{min}^4(\Sigma_0)\lambda_{min}^2(Z'Z/n)}{1 + 1\vee[\lambda_{max}^2(\Sigma_0)\lambda_{max}^2(4\Pi_0'Z'Z/n)]}.$$

Proof. For a direction $\gamma = (\gamma_1, \gamma_2)$, we have $F[\gamma,\gamma] = \nabla^2\psi(\theta_0)[\gamma,\gamma] = \nabla(\nabla\psi(\theta_0)[\gamma])[\gamma]$. Since

$$\psi(\theta) = -\frac{1}{4n}\mathrm{trace}(Z\theta_2\theta_1^{-1}\theta_2'Z') - \frac{1}{2}\log\det(-2\theta_1)$$

by direct calculations we have

$$\begin{aligned}\nabla\psi(\theta)[\gamma] &= (1/[4n])\mathrm{trace}(\gamma_1'\theta_1^{-1}\theta_2'Z'Z\theta_2\theta_1^{-1}) - (1/[2n])\mathrm{trace}(\gamma_2'\theta_1^{-1}\theta_2'Z'Z)+\\ &\qquad +(1/2)\mathrm{trace}(\gamma_1'\theta_1^{-1})\end{aligned}$$

and

$$\begin{aligned}F[\gamma,\gamma] &= -(1/[2n])\mathrm{trace}(\gamma_1'\theta_1^{-1}\theta_2'Z'Z\theta_2\theta_1^{-1}\theta_1^{-1}\gamma_1')+\\ &+(1/n)\mathrm{trace}(\gamma_2'\theta_1^{-1}\gamma_1\theta_1^{-1}\theta_2'Z'Z)-\\ &\quad -(1/[2n])\mathrm{trace}(\gamma_2'Z'Z\gamma_2'\theta_1^{-1}) + (1/2)\mathrm{trace}(\gamma_1'\theta_1^{-1}\gamma_1'\theta_1^{-1})\\ &= (4/n)\mathrm{trace}(\gamma_1'\Pi_0'Z'Z\Pi_0\gamma_1') + (4/n)\mathrm{trace}(\gamma_1'\Pi_0'Z'Z\gamma_2'\Sigma_0)+\\ &\quad +(1/n)\mathrm{trace}(\Sigma_0^{1/2}\gamma_2'Z'Z\gamma_2'\Sigma_0^{1/2}) + 2\mathrm{trace}(\Sigma_0^{1/2}\gamma_1'\Sigma_0\gamma_1'\Sigma_0^{1/2})\end{aligned}$$

where we used that $\theta_1 = -(1/2)\Sigma_0^{-1}, \theta_2 = \Pi_0\Sigma_0^{-1}$. Next note that only the second term in the expression above can be negative. However, we can bound its magnitude by

$$\begin{aligned}|(4/n)\mathrm{trace}(\gamma_1'\Pi_0'Z'Z\gamma_2'\Sigma_0)| &= |(4/n)\mathrm{trace}(\Pi_0'Z'Z\gamma_2'\Sigma_0\gamma_1')|\\ &\leq \|4\Pi_0'(Z'Z/n)\gamma_2\|\|\Sigma_0\gamma_1'\|\\ &\leq \lambda_{max}(4\Pi_0'Z'Z/n)\|\gamma_2\|\lambda_{max}(\Sigma_0)\|\gamma_1\|.\end{aligned}$$

Moreover, we also note that $-\frac{1}{4n}\mathrm{trace}(Z\theta_2\theta_1^{-1}\theta_2'Z')$ is a convex function in the relevant range.[1] Thus,

$$\begin{aligned}&(4/n)\mathrm{trace}(\gamma_1'\Pi_0'Z'Z\Pi_0\gamma_1') + (4/n)\mathrm{trace}(\gamma_1'\Pi_0'Z'Z\gamma_2'\Sigma_0)+\\ &\qquad +(1/n)\mathrm{trace}(\Sigma_0^{1/2}\gamma_2'Z'Z\gamma_2'\Sigma_0^{1/2}) \geq 0.\end{aligned} \tag{E.36}$$

To bound $\min_\gamma F[\gamma,\gamma]/\|\gamma\|^2$ from below let

$$\mu = 1 \vee \lambda_{max}(4\Pi_0'Z'Z/n)\lambda_{max}(\Sigma_0)/[\lambda_{min}(\Sigma_0)\lambda_{min}(Z'Z/n)].$$

If $\|\gamma_2\| \geq 2\mu\|\gamma_1\|$ we have

$$\begin{aligned}F[\gamma,\gamma] &\geq (1/2)\lambda_{min}(\Sigma_0)\lambda_{min}(Z'Z/n)\|\gamma_2\|^2\\ &\geq (1/2)\lambda_{min}(\Sigma_0)\lambda_{min}(Z'Z/n)(1/(1+[1/\mu^2])\|\gamma\|^2.\end{aligned}$$

[1]Indeed, over $\{X = (X_1, X_2) : X_1 \succeq 0\}$, we have

$$\mathrm{trace}(AX_2X_1^{-1}X_2'A') = \min_M \mathrm{trace}(M) : \begin{bmatrix} M & X_2'A' \\ AX_2 & X_1 \end{bmatrix} \succeq 0.$$

Otherwise, we can assume that $\|\gamma_2\| \leq 2\mu\|\gamma_1\|$. In this case, by (E.36) we have

$$\begin{array}{ll} F[\gamma,\gamma] & \geq \lambda_{min}(\Sigma_0)^2\|\gamma_1\|^2 \\ & \geq \lambda_{min}(\Sigma_0)^2\|\gamma\|^2(1/[1+4\mu^2]). \end{array}$$

The result follows. ■

Lemma 11 *In the multivariate linear model, we have*

$$B_{1n}(c) = O(d_c) \quad \textit{and} \quad B_{2n}(c) = O(d_c^2)$$

where the constants can depend on the maximal eigenvalues of J^{-1} *and* Σ_0*, and maximal singular value of* Π_0*.*

Proof. Let $y \in \mathbb{R}^{d_r}$ $X = (yy', z_i y')$ we have

$$yy' = (u_i + z_i\Pi)(u_i' + \Pi' z_i') = u_i u_i' + u_i \Pi' z_i' + z_i \Pi u_i' + z_i \Pi\Pi' z_i'$$

so that if $u \sim N(0,\Sigma)$, we have $yy' - E[yy'] = u_i u_i' - \Sigma + u_i\Pi' z_i' + z_i \Pi u_i'$. Similarly,

$$z_i y' = z_i(u_i' + \Pi' z_i') = z_i u_i' + z_i \Pi' z_i'$$

so that if $u \sim N(0,\Sigma)$, we have $z_i y' - E[z_i y'] = z_i u_i'$. Thus, for $a = (a_1', a_2')'$ and $J^{-1} = [J_1^{-1}; J_2^{-1}]$,

$$\begin{array}{ll} \left\langle a, J^{-1}(X - E[X])\right\rangle & = \text{trace}(a_1' J_1^{-1}[u_i u_i' - \Sigma + u_i \Pi' z_i' + z_i \Pi u_i']) \\ & \quad + \text{trace}(a_2' J_2^{-1}[z_i u_i']) \end{array}$$

and by symmetry of the probability distribution of u, we have

$$\begin{array}{l} |E\left[\left\langle a, J^{-1}(X - E[X])\right\rangle^3\right]| \\ = |E\left[\text{trace}^3(a_1' J_1^{-1}[u_i u_i' - \Sigma])\right] + \\ \quad +3E\left[\text{trace}(a_1' J_1^{-1}[u_i u_i' - \Sigma])\text{trace}^2(a_2' J_2^{-1}[z_i u_i'])\right]| \\ \leq 2^3|E\left[\text{trace}^3(a_1' J_1^{-1} u_i u_i')\right]| + 2^3|\text{trace}^3(a_1' J_1^{-1}\Sigma)| + \\ \quad +3E\left[\text{trace}(a_1' J_1^{-1} u_i u_i')\text{trace}^2(a_2' J_2^{-1} z_i u_i')\right]| + \\ \quad +3\text{trace}(a_1' J_1^{-1}\Sigma)E\left[\text{trace}^2(a_2' J_2^{-1} z_i u_i')\right]| \\ \leq 2^3|E\left[(u_i' a_1' J_1^{-1} u_i)^3\right]| + 2^3|\text{trace}^3(a_1' J_1^{-1}\Sigma)| + \\ \quad +3E\left[(u_i' a_1' J_1^{-1} u_i)(u_i' a_2' J_2^{-1} z_i)^2\right]| + \\ \quad +3\text{trace}(a_1' J_1^{-1}\Sigma)E\left[(u_i' a_2' J_2^{-1} z_i)^2\right]| \\ = O(d_c) \end{array}$$

where we used the Gaussianity of u, Lemma 6, and that J has bounded eigenvalues.

To obtain the second result, we have

$$E\left[\left\langle a, J^{-1}(X-E[X])\right\rangle^4\right] \leq 2^4 E\left[\mathrm{trace}^4(a_1' J_1^{-1}[u_i u_i' - \Sigma + u_i \Pi' z_i' + z_i \Pi u_i'])\right] + +2^4 E\left[\mathrm{trace}^4(a_2' J_2^{-1} z_i u_i')\right].$$

Similar calculations yield

$$E\left[\left\langle a, J^{-1}(X-E[X])\right\rangle^4\right] = O(d_c^2).$$

■

Lemma 12 *In the seemingly unrelated regressors model, for every $\kappa > 0$, uniformly in $\gamma = (\gamma_1, \gamma_2) \in B(0, \kappa\sqrt{d})$ we have*

$$\sqrt{n}(\theta(\eta_0 + \gamma/\sqrt{n}) - \theta(\eta_0)) = \begin{pmatrix} -\gamma_1/2 \\ \Pi_0\gamma_1 + \gamma_2\Sigma_0^{-1} \end{pmatrix} + r_{1n}(\gamma)$$

where $\|r_{1n}(\gamma)\| \leq \kappa^2 d/\sqrt{n}$.

Proof. By direct calculations we have

$$\nabla\theta(\eta_0)[\gamma] = \begin{pmatrix} -\gamma_1/2 \\ \Pi_0\gamma_1 + \gamma_2\Sigma_0^{-1} \end{pmatrix} \quad \text{and} \quad \nabla^2\theta(\eta)[\gamma,\gamma] = \begin{pmatrix} 0 \\ 2\gamma_2\gamma_1 \end{pmatrix}$$

so that $\|\nabla^2\theta(\eta)[\gamma,\gamma]\| \leq \|\gamma\|^2$ for all η. Thus we can set $R_{2n} = 0$ and note that we can set

$$\|r_{1n}(\gamma)\| \leq \sqrt{n}\sup_{\eta}\|\nabla^2\theta(\eta)[\gamma/\sqrt{n}, \gamma/\sqrt{n}]\| \leq \|\gamma\|^2/\sqrt{n} \leq \kappa^2 d/\sqrt{n}.$$

■

REFERENCES

[1] O. Barndorff-Nielsen (1978). Information and exponential families in statistical theory. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Ltd., Chichester.

[2] A. Belloni and V. Chernozhukov (2009). On the Computational Complexity of MCMC-based Estimators in Large Samples, Ann. Statist. 37, 2011-2055.

[3] P. J. Bickel and B. J. K. Kleijn (2012). The semiparametric Bernsteinvon Mises theorem, Ann. Statist. Volume 40, Number 1 (2012), 206-237.

[4] P. J. Bickel and J. A. Yahav (1969). Some contributions to the asymptotic theory of Bayes solutions, Z. Wahrsch. Verw. Geb 11, 257-276.

[5] D. Bontemps (2011). Bernstein-von Mises theorems for Gaussian regression with increasing number of regressors, Ann. Statist. Vol. 39, No. 5, 2557–2584.

[6] S. Boucheron and E. Gassiat (2009). A Bernstein-Von Mises Theorem for discrete probability distributions, Electron. J. Statist. Volume 3 (2009), 114-148.

[7] G. Chamberlain (1987). Asymptotic efficiency in estimation with conditional moment restrictions, Journal of Econometrics 34, no. 3, 305–334

[8] V. Chernozhukov and H. Hong (2003). An MCMC approach to classical estimation, Journal of Econometrics 115, 293-346.

[9] P. Diaconis, S. Holmes and M. Shahshahani (2012). Sampling From A Manifold, arXiv:1206.6913

[10] S. G. Donald, G. W. Imbens, W. K. Newey (2003). Empirical likelihood estimation and consistent tests with conditional moment restrictions. Journal of Econometrics, 117, no. 1, 55–93.

[11] R. Dudley (2000). Uniform Cental Limit Theorems, Cambridge Studies in advanced mathematics.

[12] B. Efron (1975). Defining the curvature of a statistical problem (with applications to second order efficiency)(with discussion), Annals of Statistics, Vol. 3, No. , 1189-1242.

[13] B. Efron (1978). The Geometry of Exponential Families, Annals of Statistics, Vol. 6, No. 2, 362-376.

[14] L. P. Hansen and K. J. Singleton (1982). Generalized instrumental variables estimation of nonlinear rational expectations models, Econometrica 50, no. 5, 1269–1286.

[15] I. Ibragimov and R. Has'minskii (1981). Statistical Estimation: Asymptotic Theory, Springer, Berlin.

[16] G. W. Imbens (1997). One-step estimators for over-identified generalized method of moments models. Rev. Econom. Stud. 64, no. 3, 359–383.

[17] S. Ghosal (2000). Asymptotic normality of posterior distributions for exponential families when the number of parameters tends to infinity, Journal of Multivariate Analysis, vol 73, 49-68.

[18] J. Heckman (1974). Shadow prices, market wages, and labor supply Econometrica 42, 679–694.

[19] D. R. Hunter (2007). Curved exponential family models for social networks, Social Networks 29, pp. 216–230.

[20] D. R. Hunter and M. S. Handcock (2006). Inference in curved exponential family models for networks, Journal of Graphical and Computational Statistics, Vol. 15, Number 3, pp. 565–583.

[21] E. L. Lehmann and G. Casella (1998). Theory of point estimation. Second edition. Springer Texts in Statistics. Springer-Verlag, New York.

[22] L. Lovász and S. Vempala, The Geometry of Logconcave Functions and Sampling Algorithms, To appear in Random Structures and Algorithms.

[23] W. K. Newey and D. McFadden (1994). Large Sample Estimation and Hypothesis Testing, Handbook of Econometrics, Volume IV, Chapter 36, Editedby R.F. Engle and D.L. McFadden, Elsevier Science.

[24] S. Portnoy (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. Ann. Statist. 16, no. 1, 356–366.

[25] J. Qin and J. Lawless (1994). Empirical Likelihood and General Estimating Equations, The Annals of Statistics, Vol. 22, No. 1, 300–325.

[26] K. J. van Garderen (1997). Curved Exponential Models in Econometrics, Econometric Theory, 13, 771–790.

[27] A. Zellner (1962). An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias, Journal of the American Statistical Association, 57, 348-368.

[28] A. Zellner (1971). An Introduction to Bayesian Inference in Econometrics, John Wiley & Sons: New York, NY.

100 Fuqua Drive
Durham, NC 27708
E-mail: abn5@duke.edu

50 Memorial Drive
Cambridge , MA 02142
E-mail: vchern@mit.edu