Supplementary Appendix to
“Optimal Decision Rules for Weak GMM”
By Isaiah Andrews\(^1\) and Anna Mikusheva\(^2\)

Abstract

This supplementary appendix provides additional results for the invariance argument referenced in the main text, on the properties of Bayes and quasi-Bayes decision rules in the limit experiment, and details of the simulation designs and application discussed in the main text.

Keywords: Limit Experiment, Quasi Bayes, Weak Identification, Nonlinear GMM

JEL Codes: C11, C12, C20

This draft: December 2020.

S1 Invariance Argument

As discussed in the main text, we seek default priors on \(\mu\) that yield reasonable decision rules when combined with many different priors on \(\theta^*\), including priors with restricted support \(\tilde{\Theta} \subset \Theta_0\). For any such prior \(\pi(\theta^*)\), define a corresponding restricted model with parameter space \(\tilde{\Theta} \times H_\mu\). Intuitively, by specifying a restricted-support prior \(\pi(\theta^*)\), a researcher limits attention to the restricted model.

We next introduce a group of transformations. Define \(H_{\mu,\tilde{\Theta}}\) to be the linear subspace of functions in \(H_\mu\) that are zero everywhere on \(\tilde{\Theta}\),

\[
H_{\mu,\tilde{\Theta}} = \left\{ \mu \in H_\mu : \mu(\theta) = 0 \text{ for all } \theta \in \tilde{\Theta} \right\}.
\]

The decision problem in the restricted model is invariant with respect to the group of transformations that for \(\mu \in H_{\mu,\tilde{\Theta}}\) takes \((\xi, h(\cdot)) \rightarrow (\xi, h(\cdot) + \tilde{\mu}(\cdot))\) in the sample space, and \((\theta^*, \mu(\cdot)) \rightarrow (\theta^*, \mu(\cdot) + \tilde{\mu}(\cdot))\) in the parameter space. See Chapter 3 of Lehmann and Casella (1998) for an introduction to invariance in decision problems.

\(^1\)Harvard Department of Economics, Littauer Center M18, Cambridge, MA 02138. Email iandrews@fas.harvard.edu. Support from the National Science Foundation under grant number 1654234, and from the Sloan Research Fellowship is gratefully acknowledged.

\(^2\)Department of Economics, M.I.T., 50 Memorial Drive, E52-526, Cambridge, MA, 02142. Email: amikushe@mit.edu. Financial support from the Castle-Krob Career Development Chair and the Sloan Research Fellowship is gratefully acknowledged. We thank Jiafeng Chen, David Hughes, and Ashesh Rambachan for research assistance, and the editor and two anonymous referees for helpful comments.
A maximal invariant in the parameter space under this group of transformations is \( (\theta^*, \mu (\cdot) I\{\cdot \in \tilde{\Theta}\}) \). The statistic \( (\xi, h(\cdot) I\{\cdot \in \tilde{\Theta}\}) \) is also invariant, and is sufficient for \( (\theta^*, \mu (\cdot) I\{\cdot \in \tilde{\Theta}\}) \). Hence, once we restrict ourselves to invariant decision rules, it is without loss (in terms of attainable risk) to limit attention to decision rules that depend on the data only through \( (\xi, h(\cdot) I\{\cdot \in \tilde{\Theta}\}) \).

By definition, Bayes decision rules in our setting minimize the quasi-posterior risk

\[
\min_{a \in A} \frac{\int L(a, \theta) \ell^*(\theta) d\pi(\theta)}{\int \ell^*(\theta) d\pi(\theta)}
\]

for almost every realization of the data. Motivated by the invariance of the restricted model, we seek default priors \( \pi(\mu) \) such that for all priors \( \pi(\theta^*) \) with restricted support \( \tilde{\Theta} \subset \Theta_0 \), the joint prior \( \pi(\theta^*)\pi(\mu) \) admits Bayes decision rules depending only on \( (\xi, h(\cdot) I\{\cdot \in \tilde{\Theta}\}) \).

**Definition S1.1** A prior \( \pi(\mu) \) is *invariance-compatible* for action space \( A \) and loss function \( L \) if for all \( \tilde{\Theta} \subseteq \Theta_0 \) and all priors \( \pi(\theta^*) \) with support \( \tilde{\Theta} \), there exists a Bayes decision rule that depends on the data only through \( (\xi, h(\cdot) I\{\cdot \in \tilde{\Theta}\}) \).

Priors \( \pi(\mu) \) such that \( \ell^*(\theta^*) \) depends on the data only through \( (\xi, h(\theta)) \) are invariance-compatible for any \( (A, L) \). For unrestricted action spaces and loss functions, the converse (up to scale) holds as well.

**Theorem S1.1** A prior \( \pi(\mu) \) is invariance-compatible for all \( (A, L) \) pairs if and only if for any \( \theta \in \Theta_0 \) the integrated likelihood \( \ell^*(\theta) \), defined in equation (6) in the main text, depends on the data only through \( (\xi, h(\theta)) \), up to scale.

As discussed in the main text, for Gaussian process priors on \( \mu \), \( \ell^*(\theta) \) corresponds to a Gaussian likelihood for \( \xi \) with mean equal to the best linear predictor based on \( h(\cdot) \). In this case there is no scope for a data-dependent constant of proportionality, and \( \pi(\mu) \) is invariance-compatible if and only if \( \ell^*(\theta) \) depends on the data only through \( (\xi, h(\theta)) \).
Proof of Theorem S1.1  The “if” part of the statement is immediate. For “only if,” first consider the case where \( \tilde{\Theta} = \{\theta_1, \theta_2\} \), \( \mathcal{A} = \{a_1, a_2\} \), and

\[
L(a, \theta) = \begin{cases} 
  l_j & \text{if } \theta = \theta_j \text{, and } a \neq a_j \text{ for } j \in \{1, 2\}; \\
  0 & \text{otherwise,}
\end{cases}
\]

so an incorrect decision incurs a loss \( l_j \). Bayes decision rules must take

\[
a \in \begin{cases} 
  \{1\} & \text{if } \frac{\ell^*(\theta_1)}{\ell^*(\theta_2)} > \frac{l_2 \pi(\theta_2)}{l_1 \pi(\theta_1)}; \\
  \{1, 2\} & \text{if } \frac{\ell^*(\theta_1)}{\ell^*(\theta_2)} = \frac{l_2 \pi(\theta_2)}{l_1 \pi(\theta_1)}; \\
  \{2\} & \text{if } \frac{\ell^*(\theta_1)}{\ell^*(\theta_2)} < \frac{l_2 \pi(\theta_2)}{l_1 \pi(\theta_1)},
\end{cases}
\]

for almost every realization of the data. Thus, we see that the optimal action depends on the data only through \( (\xi, h(\theta_1), h(\theta_2)) \) for all values of \( l_1, l_2 \) if and only if the integrated likelihood ratio \( \ell^*(\theta_1) / \ell^*(\theta_2) \) depends on the data only through \( (\xi, h(\theta_1), h(\theta_2)) \).

Since we can repeat this argument for all \( \theta_1, \theta_2 \in \Theta_0 \), we conclude that for \( \pi(\mu) \) to be invariance-compatible, it must be the case that for all \( \theta_1, \theta_2 \in \Theta_0 \)

\[
\frac{\ell^*(\theta_1)}{\ell^*(\theta_2)} = \tilde{r}(\xi, h(\theta_1), h(\theta_2); \theta_1, \theta_2),
\]

for some function \( \tilde{r} \). This implies that for all \( \theta_1, \theta_2, \theta_3 \in \Theta_0 \),

\[
\tilde{r}(\xi, h(\theta_1), h(\theta_2); \theta_1, \theta_2) = \frac{\tilde{r}(\xi, h(\theta_1), h(\theta_3); \theta_1, \theta_3)}{\tilde{r}(\xi, h(\theta_3), h(\theta_2); \theta_3, \theta_2)}.
\]

Hence, the right hand side does not depend on the value of \( h(\theta_3) \), and in particular is the same as if \( h(\theta_3) = 0 \). For a fixed value \( \tilde{\theta} \in \Theta_0 \), define \( r(\xi, h(\theta_1), \theta_1) = \tilde{r}(\xi, h(\theta_1), 0; \theta_1, \tilde{\theta}) \), and note that

\[
\ell^*(\theta_1) = \frac{\ell^*(\theta_2)}{r(\xi, h(\theta_2), \theta_2)} r(\xi, h(\theta_1), \theta_1),
\]

where the right side cannot depend on \( \theta_2 \). Thus \( \ell^*(\theta) \propto r(\xi, h(\theta), \theta) \). □
S2 Properties of Bayes and Quasi-Bayes Rules

This section establishes additional properties for the Bayes decision rules studied in the main text.

S2.1 Admissibility of Proportional Prior Rules

In finite-dimensional settings (specifically, settings where the covariance function $\tilde{\Sigma}$ has a finite number of nonzero eigenvalues) choosing $\Omega = \lambda \tilde{\Sigma}$ implies that $\pi(\mu)$ has support $H_\mu$. In infinite-dimensional settings, by contrast, $\pi(\mu)$ assigns probability zero to $H_\mu$—see Section 3.1 of Berlinet and Thomas-Agnan (2004). It may not therefore be obvious that the resulting Bayes decision rules are admissible.

This section shows that admissibility continues to hold under a weak continuity condition. Specifically, for $\|\cdot\|$ the Euclidian norm, let $\|\mu\|_\infty = \sup_{\theta \in \Theta_0} \|\mu(\theta)\|$ be the sup norm, and define $\overline{H}_\mu$ as the closure of $H_\mu$ under $\|\cdot\|_\infty$. For a metric $d_\theta$ on $\Theta_0$, define the metric

$$d((\theta, \mu), (\theta', \mu')) = d_\theta(\theta, \theta') + \|\mu - \mu'\|_\infty$$

on $\Theta_0 \times \overline{H}_\mu$. Let $S$ be the class of decision rules whose risk functions $\mathbb{E}_{\theta, \mu}[L(s(g), \theta)]$ are continuous with respect to $d$.

**Theorem S2.1** For $\pi(\theta, \mu) = \pi(\theta) \pi(\mu)$, where $\pi(\theta)$ has support $\Theta_0$ and $\pi(\mu)$ corresponds to $\mu \sim \mathcal{GP}(0, \lambda \tilde{\Sigma})$ with $\lambda > 0$, any decision rule $s_\pi \in S$ with

$$\int \mathbb{E}_{\theta, \mu}[L(s_\pi(g), \theta)] d\pi(\theta, \mu) = \min_{s \in S} \int \mathbb{E}_{\theta, \mu}[L(s(g), \theta)] d\pi(\theta, \mu)$$

is admissible relative to $S$ and parameter space $\Theta_0 \times \overline{H}_\mu$.

**Proof of Theorem S2.1** Suppose there exists $\tilde{s} \in S$ that dominates $s_\pi$. Since $\tilde{s}$ dominates $s_\pi$, and the risk function is continuous, the set

$$\{(\theta, \mu) \in \Theta \times \overline{H}_\mu : \mathbb{E}_{\theta, \mu}[L(s_\pi(g), \theta)] - \mathbb{E}_{\theta, \mu}[L(\tilde{s}(g), \theta)] > 0\}$$

is admissible relative to $S$ and parameter space $\Theta_0 \times \overline{H}_\mu$. 

4
is nonempty and open, while the set

$$\left\{ (\theta, \mu) \in \Theta \times \mathcal{H}_\mu : \mathbb{E}_{\theta, \mu} [L(s_\pi(g), \theta)] - \mathbb{E}_{\theta, \mu} [L(\bar{s}(g), \theta)] < 0 \right\} \quad (3)$$

is empty. By Lemma 5.1 in van der Vaart and van Zanten (2008), \( \pi(\mu) \) has support \( \mathcal{H}_\mu \). From the definition of \( d \), \( \pi(\theta, \mu) \) must therefore assign positive mass to (2). Since (3) is empty, this implies that \( \int \mathbb{E}_{\theta, \mu} [L(s_\pi(g), \theta)] d\pi(\theta, \mu) \) strictly exceeds \( \int \mathbb{E}_{\theta, \mu} [L(\bar{s}(g), \theta)] d\pi(\theta, \mu) \). We have obtained a contradiction with (1).

### S2.2 Limit-of-Bayes

As discussed in the main text, we obtain the quasi-posterior of Chernozhukov and Hong (2003) as the limit of a sequence of posteriors for proper priors. Here, we show that under the conditions similar to Theorem 2 in the main text, quasi-Bayes decision rules are likewise the pointwise limit of the corresponding Bayes decision rules, and that the same holds for their risk functions. To state this result, consider any sequence of finite values \( \lambda_r \to \infty \) as \( r \to \infty \), define \( \pi_r(\theta, \mu) = \pi(\theta) \pi_r(\mu) \) to be the corresponding sequence of priors, and let \( s_{\pi_r} \) be the corresponding sequence of Bayes decision rules,

$$s_{\pi_r}(g) \in \arg\min_{a \in A} \frac{\int L(a, \theta) \ell_r^*(\theta) d\pi(\theta)}{\int \ell_r^*(\theta) d\pi(\theta)} = \arg\min_{a \in A} \int L(a, \theta) \ell_r^*(\theta) d\pi(\theta),$$

and \( s_{\pi_\infty}(g) \) the quasi-Bayes decision rule.

**Theorem S2.2** Suppose that \( A \) is compact and convex, that \( L(a, \theta) \) is uniformly bounded as well as continuous and strictly convex in \( a \) for every \( \theta \), and that \( \Sigma(\theta, \theta) \) is everywhere full rank. Then \( s_{\pi_r}(g) \to s_{\pi_\infty}(g) \) for every \( g \) and

$$\mathbb{E}_{\theta, \mu} [L(s_{\pi_r}(g), \theta)] \to \mathbb{E}_{\theta, \mu} [L(s_{\pi_\infty}(g), \theta)] \text{ for each } (\theta, \mu) \in \Theta_0 \times \mathcal{H}_\mu. \quad (4)$$

**Proof of Theorem S2.2** Recall that \( \ell_r^*(\theta) = |\Lambda_r(\theta)|^{-\frac{1}{2}} \cdot \exp \left(-\frac{1}{2} u_r(\theta) \Lambda_r(\theta)^{-1} u_r(\theta) \right), \)

where \( u_r(\theta) = \frac{\lambda}{1+\lambda_r} \psi(\theta)^{-1} g(\theta) + \frac{1}{1+\lambda_r} \xi, \ \Lambda_r(\theta) = \frac{\lambda}{1+\lambda_r} [\psi(\theta)^{-1}] \Sigma(\theta, \theta) [\psi(\theta)^{-1}]' + \frac{1}{1+\lambda_r} \text{Var}(\xi). \) Since the minimum eigenvalue of \( \Sigma(\theta, \theta) \) is bounded away from zero by
compactness of $\Theta_0$, and $g(\cdot)$ is bounded, we have:

$$
\ell_r^* (\theta) \to \ell_\infty^* (\theta) = |\psi(\theta)| \cdot |\Sigma(\theta, \theta)|^{-\frac{1}{2}} \cdot \exp \left( -\frac{1}{2} g(\theta)^\prime \Sigma(\theta, \theta)^{-1} g(\theta) \right)
$$

uniformly on $\Theta_0$. Since the loss function is bounded,

$$
\int L (a, \theta) \ell_r^* (\theta) \, d\pi (\theta) \to \int L (a, \theta) \ell_\infty^* (\theta) \, d\pi (\theta) \quad \text{uniformly on } A,
$$

which, since the loss is strictly convex, implies that $s_{\pi_r} (g) \to s_{\pi_\infty} (g)$. Finally, (4) follows from another application of the dominated convergence theorem.

**S3 Numerical Results**

**S3.1 Simulation Design**

This section describes the simulation design used to produce Figure 1 in Section 2.1 of the paper. We base our simulations on the Graddy (1995) data. In particular, for our data-calibrated simulations, we:

1. Estimate $\hat{\theta}$ using continuously updating GMM.

2. Draw $(W, Y, Z)$ from the empirical distribution, and generate new outcomes $Y^*$ by adding normal noise with standard deviation equal to one-tenth that of $Y$, i.e. $Y^* = Y + \varepsilon$, with $\varepsilon \sim N \left( 0, \frac{1}{100} \text{Var} (Y) \right)$.

3. Use exponential tilting for find weights $\omega_i$ on the observations $\{W_i, Y_i, Z_i\}$ in the original sample such that the quantile IV moments, evaluated with the observations constructed in step 2, hold exactly at $\hat{\theta}$.

4. Draw samples $\{W, Y^*, Z\}$ with sampling weights $\omega_i$ as in step 3, and outcomes $Y^*$ as in step 2.

This construction ensures that (i) $Y$ is continuously distributed, as necessary for conditional quantiles to be uniquely defined and (ii) the (over-identified) GMM moments hold
exactly at the value $\hat{\theta}$ estimated on the original data. Denote the resulting distribution by $P^*$.

To construct the distribution $P_0$, we take the distribution $P^*$ and, as described in Section 2 of the paper, multiply all entries of $Z$ but the constant by a mean-zero (specifically, Rademacher) random variable. This construction ensures that $P_0$ dominates $P^*$ and, as discussed in the text, that $\theta^*$ is set-identified under $P_0$.

Finally, to construct $P_{n,f}$, we draw from a mixture between $P_0$ and $P^*$, with weight $\sqrt{\frac{111}{n}}$ on $P^*$. Hence, for the original sample size ($n = 111$) $P_{n,f} = P^*$, while as $n \to \infty$, $P_{n,f}$ converges to $P_0$. In particular, $P_{n,f}$ satisfies equation (1) in the paper for $f \propto \frac{dP^*}{dP_0} - 1$.

### S3.2 Additional Empirical Results

Figure 1 plots the 95% conditional confidence set formed by inverting the weighted average power optimal tests described in the paper, along with the 95% highest posterior density set. The conditional confidence sets do have asymptotically correct frequentist coverage and use quasi-posterior only to form a powerful test statistics. The 95% highest posterior density set does not have frequentist guarantees and is a Bayesian object. In this application the two sets have a quite similar shape, but the confidence set is slightly smaller, covering 4.74% of the parameter space as compared to 4.82% for the highest posterior density set.

Figure 2 plots the small sample distribution of the GMM estimate for the median, $\tau = 0.5$ and weak- and large-sample asymptotic distributions. We again see that the weak-asymptotic approximation appears substantially more accurate.
Figure 1: Quasi-Bayes 95% highest posterior density set, and 95% conditional frequentist confidence set, for $\tau = 0.75$ based on Graddy (1995) data.

Figure 2: Finite sample distribution of GMM estimator in simulations calibrated to Graddy (1995) data, along with weak- and strong-asymptotic large-sample distributions. Based on 2500 simulation draws.