Estimating Bayesian Decision Problems with Heterogeneous Priors*

Stephen Hansen
Universitat Pompeu Fabra
Barcelona GSE
stephen.hansen@upf.edu

Michael McMahon
University of Warwick
CAGE, CEP (LSE), and CAMA (ANU)
m.mcmahon@warwick.ac.uk

March 1, 2013

Abstract

In many areas of economics there is a growing interest in how expertise and preferences drive individual and group decision making under uncertainty. Increasingly, we wish to estimate such models to quantify which of these drive decision making. In this paper we propose a new channel through which we can empirically identify expertise and preference parameters by using variation in decisions over heterogeneous priors. Relative to existing estimation approaches, our "Prior-Based Identification" extends the possible environments which can be estimated, and also substantially improves the accuracy and precision of estimates in those environments which can be estimated using existing methods.

Keywords: Bayesian decision making; expertise; preferences; estimation.

JEL Codes: D72, D81, C13

*We would like to thank, without implicating, Greg Crawford and Carlos Velasco Rivera for numerous helpful discussions. A not for publication appendix with supplemental tables and figures is available at http://www2.warwick.ac.uk/fac/soc/economics/staff/academic/mcmahon/research/priors_-_appendix.pdf. Full results are available on request along with all the R code necessary to implement the simulations and estimation exercises.

1 Introduction

How individuals and groups of individuals make decisions under uncertainty is important in many areas of economics and political economy, and there are several areas in which theoretical models emphasize that decision makers differ both in terms of their knowledge of an underlying state of the world and their preferences.[1] Increasingly, we would like to bring such models to the data in order to estimate the decision-making parameters and understand, quantitatively, the role played by different factors in decision making in different contexts. Important recent work by Iaryczower and Shum (2012) - MIMS[2] hereafter - provides a two-step methodology to estimate decision parameters for groups of contemporaneously-serving experts that relies on how decisions vary across states of the world drawn from a given prior distribution ("State-Based Identification"). Our contribution in this paper is to highlight another important channel through which an econometrician can empirically identify the decision-making parameters in a Bayesian decision problem: by using variation in decision-making behavior over heterogeneous priors ("Prior-Based Identification").[3]

We first present a binary choice model of Bayesian decision making in order to illustrate how decision makers with different preferences and/or expertise have different probabilities of choosing one decision (instead of the other). Let the choices be between option 0 and option 1. Our identification can informally be understood by considering that becoming more inclined toward choosing option 1 (a change in preferences toward 1) will cause the probability of choosing 1 to increase for all values of the prior belief (a "shift" in their probability of choosing 1), while decision makers with more expertise will have a lower probability of choosing 1 when the prior favors that choice, but a higher probability of choosing 1 when the prior favors option 0. This "rotation" is driven by the fact that decision makers with more expertise rely more on their own view rather than the prior. It is this distinction between "shifts" and "rotations" which facilitates Prior-Based Identification.

This new channel of identification has two implications that we explore in this paper. The first is that preference and expertise parameters are identifiable without the need for contemporaneous correlation in decisions. This opens up to empirical testing a broader set of decision-making environments than that recognized by MIMS. For example, we

[1]For example, see the literatures on the career concerns of experts (Sorensen and Ottaviani 2000); committee decision making (Gerling, Gruner, Kiel, and Schulte 2005); politicians' behavior (Besley 2006); and social learning (Banerjee 1992, Bikhchandani, Hirshleifer, and Welch 1992).

[2]Standing for Matias Iaryczower and Matt Shum to avoid potential confusion generated by simply using IS.

[3]Another related paper is Li (2012) which proposes a way of separately identifying preferences from expertise among advisors providing policy recommendations in a cheap-talk setting.

show using Monte Carlo simulations that we can estimate the parameters of a single decision maker serving over time using noisy measures of the prior.[4]

The second implication is to show that the specification of MIMS to estimate decision-making parameters for individuals within groups is incomplete as it does not allow different individuals to react differently to changes in the common prior. We argue theoretically that this omission should be particularly problematic when decision makers are more heterogeneous in terms of their expertise. We then show, again using Monte Carlo simulations, that adopting a more flexible specification that allows for both sources of identification substantially improves the accuracy of individual estimates when there is non-negligible expertise heterogeneity. Of particular note is that estimated differences between individuals in terms of preferences, and especially expertise, are inflated by the failure to account for Prior-Based Identification; the bias using the less flexible specification is in the range of 60-90% for reasonable differences in expertise.

To confirm the relevance of the Monte Carlo simulation exercises, we analyze the Supreme Court data used in MIMS and compare the results obtained using our proposed specification with the original MIMS specification. In line with the predictions from our Monte Carlo analysis, using our proposed specification reduces, in some cases markedly, the dispersion of the distribution of the estimated individual parameters. While we do not claim, nor attempt, to overturn MIMS' key result that private signals play an important role in decision making, our results suggest that researchers interested in the level of decision-making parameters, or comparisons between members, would be better served by using both State- and Prior-Based Identification.[5] In a closely-related paper (Hansen, McMahon, and Velasco 2013), we, together with a co-author, make use of the identification channel described in this paper to explore the extent and implications of heterogeneity in preferences and expertise for policymakers on the Bank of England's Monetary Policy Committee (MPC).

2 Theory

In this section, we present the theory underlying the Bayesian decision-making model we (and MIMS) consider and discuss two alternative ways of identifying preference parameters separately from those of of expertise; one is the identification proposed in MIMS

[4]This is in contrast to MIMS who claim that "if there were only one decision maker, for example, it would not be possible to disentangle the independent effects of ideology and information."

[5]For example, Iaryczower, Lewis, and Shum (2013) compare the expertise of elected and appointed judges using State-Based Identification and the MIMS estimator, whereas our results suggest the estimates from an alternative specification incorporating Prior-Based Identification would allow a more accurate comparison.

while the other is our proposal. We discuss how the channel of identification that we propose differs and when it is likely to be important.

2.1 Single decision maker

Consider a single decision maker DM who must take some binary decision $d_t \in \{0,1\}$ at time t. Her utility $u(d_t \mid \omega_t)$ from d_t depends on a binary, unobserved state variable $\omega_t \in \{0,1\}$ drawn from a prior distribution with $\Pr[\omega_t = 1] = q_t$. We assume that DM always prefers to match the decision to the state; i.e., that $u(d_t = \omega_t \mid \omega_t) - u(d_t \neq \omega_t \mid \omega_t) > 0$ for $\omega \in \{0,1\}$.

Before choosing d_t, DM observes a signal $s_t \sim N(\omega_t, \sigma^2)$. σ is an inverse measure of expertise in the sense that when DM has a lower σ, s_t provides a more informative signal of the unknown state. DM chooses $d_t = 1$ only if she is sufficiently convinced of $\omega_t = 1$. Formally, conditional on s_t, DM chooses $d_t = 1$ if and only if

$$\frac{\Pr[\omega_t = 1 \mid s_t]}{\Pr[\omega_t = 0 \mid s_t]} \geq \frac{u(0 \mid 0) - u(1 \mid 0)}{u(1 \mid 1) - u(0 \mid 1)} \equiv \frac{1-\theta}{\theta}. \tag{1}$$

When DM views the wrong decision in state 0 as relatively worse than the wrong decision in state 1, she requires more evidence that the state is 1 in order to choose $d_t = 1$. The parametrization in terms of θ is a common convention in the theoretical voting literature that we adopt for our empirical exercise. We shall refer to θ as DM's preferences.

Applying Bayes' Rule and manipulating the normal density gives the relationship

$$\ln\left(\frac{\Pr[\omega_t = 1 \mid s_t]}{\Pr[\omega_t = 0 \mid s_t]}\right) = \ln\left(\frac{q_t}{1-q_t}\right) + \frac{2s_t - 1}{2\sigma^2}. \tag{2}$$

An examination of (2) reveals a key relationship between signal precision and posterior beliefs: when DM has more expertise (a lower σ), she puts more weight on her signal and less weight on the prior q_t in forming her posterior distribution over states. So, intuitively speaking, the prior is less influential in determining her decision. In the limit as $\sigma \to 0$, q_t is irrelevant for determining d_t, while as $\sigma \to \infty$, q_t alone determines it.

To make these arguments more formally, note that (1) and (2) imply that DM adopts a threshold decision-making rule in which she chooses $d_t = 1$ high whenever

$$s_t \geq \frac{1}{2} - \sigma^2 \left[\ln\left(\frac{\theta}{1-\theta}\right) + \ln\left(\frac{q_t}{1-q_t}\right)\right] \equiv s_t^*(\theta, \sigma, q_t). \tag{3}$$

So, the probability she chooses $d_t = 1$ in state ω_t is

$$P(q_t, \omega_t, \theta, \sigma) \equiv \Pr\left[\, d_t = 1 \mid q_t, \omega_t, \theta, \sigma \,\right] = 1 - \Phi\left[\frac{s_t^*\left(\theta, \sigma, q_t\right) - \omega_t}{\sigma}\right]. \qquad (4)$$

The following limit arguments are useful for understanding how variation in the prior probability generates differential responses in $P(q_t, \omega_t, \theta, \sigma)$ depending on the DM's underlying preference and expertise parameters.

Proposition 1

1. $\lim_{\sigma\to 0} \Pr\left[\, d_t = 1 \mid \omega_t, \theta, \sigma, q_t \,\right] = \mathbb{1}(\omega_t)$.
2. $\lim_{\sigma\to\infty} \Pr\left[\, d_t = 1 \mid \omega_t, \theta, \sigma, q_t \,\right] = \mathbb{1}(q_t \geq 1 - \theta)$.
3. $\lim_{\theta\to 0} \Pr\left[\, d_t = 1 \mid \omega_t, \theta, \sigma, q_t \,\right] = 0 \ \forall q_t \in (0, 1)$.
4. $\lim_{\theta\to 1} \Pr\left[\, d_t = 1 \mid \omega_t, \theta, \sigma, q_t \,\right] = 1 \ \forall q_t \in (0, 1)$.

This proposition, proven by simply taking limits of $P(q_t, \omega_t, \theta, \sigma)$, illustrates that large differences in expertise can generate very different responses to changes in the prior. In either state, as σ gets sufficiently small, $\Pr\left[\, d_t = 1 \mid \omega_t, \theta, \sigma, q_t \,\right]$ is essentially unresponsive to the prior: DM always "knows" the state and simply chooses the decision to match it (from part *1* of the proposition). In contrast, as σ gets sufficiently high, small changes in the prior can generate very large changes in decision making. In both states, when the prior moves from just below $1 - \theta$ to just above it, the probability of choosing $d_t = 1$ high jumps from close to 0 to almost 1 (from part *2* of the proposition). On the other hand, large θ differences do not manifest themselves in terms of different responses to changes in prior (from parts *3* and *4* of the proposition). As preferences become extreme on either end of the $(0, 1)$ interval, decision-making behavior becomes unresponsive to changes in the prior, because DM always selects whatever decision corresponds to her extreme preferences.

These limit arguments suggest that a useful way of distinguishing the effects of changes in the parameters is to think about changes in θ as shifting P, while changes in σ cause P to rotate.[6] Of course, P is non-linear and its range is $(0, 1)$ on $q_t \in (0, 1)$ and so these statements are only heuristically true. Nevertheless, as we show in figure 1, they are by and large correct for reasonable parameter values and non-extreme values for the prior; each sub-figure shows P plotted over $q_t \in (0.01, 0.99)$ with the first (second) column capturing P conditional on the state being 0 (1), and each row shows different

[6]When σ changes in the interior of the parameter space, one can formally show that there exists a $q^*(\omega_t, \theta, \sigma)$ in the neighborhood of which P flattens when σ decreases.

combinations of parameter values. The first (second) row illustrates the situation in which decision makers differ only by preferences (expertise); the final row shows a situation in which both preferences and expertise differ.

These observations lead to our first claim:

Claim 1 *Different behavioral responses to changes in the prior, q_t, can be used to empirically distinguish preferences, θ, from expertise, σ.*

We show using Monte Carlo simulations in section 3 that this claim holds.

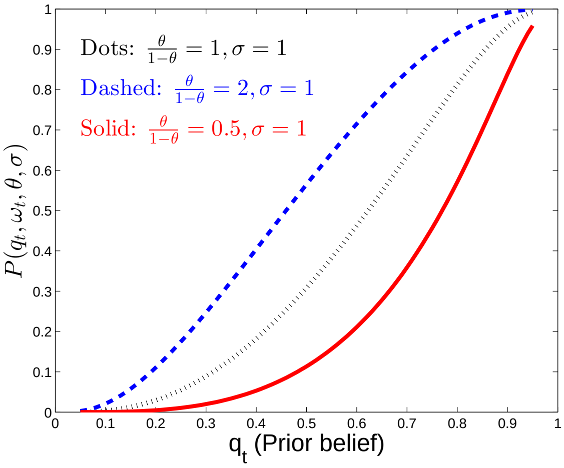
2.2 Several decision makers

While so far we have focused on a single decision maker, in many applications of interest $N \geq 2$ decision makers (who we will hereafter index by i, and each of whom are characterized by θ_i and σ_i) must take binary decisions d_{it}, and all of their utilities are affected by the realization of the same state variable ω_t. For example, one might have panel data on the buy/hold (or sell/hold) recommendations of equity analysts, and wish to estimate the expertise of each analyst separately from his or her inclination towards one alternative.

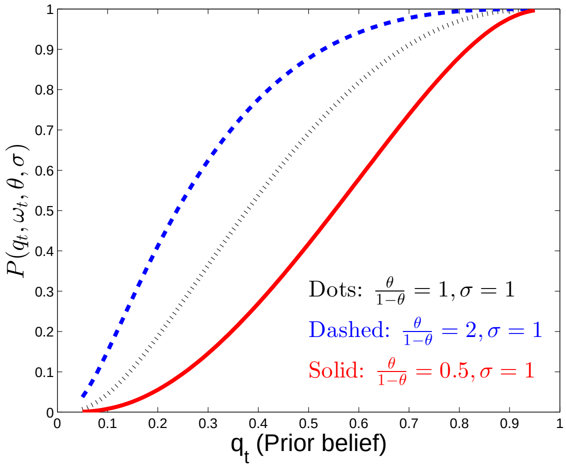
With such data MIMS argue that θ_i and σ_i are separately identifiable for a given prior via unconditional correlation in decisions. Those with high expertise have observed decisions that are highly correlated and on average in line with the prior, while those with less have observed variability in their decisions but are in the minority relatively frequently. Moreover, a large bias is identified via low variability in decisions, which tend to conform to the bias.

MIMS also argue that one should control for period-t characteristics when applying their identification strategy. Within this decision-making model, though, the only reason that the characteristics of a certain time period are relevant is if different periods have different associated values of the prior q_t. And, as we have discussed, whenever different periods have different priors, one can exploit members' differential responses to changes in those priors to better identify whether they differ in terms of preferences or expertise. While the estimator that MIMS construct (and which we discuss below in section 3) indeed incorporates period-t controls, it does *not* allow different members to react differently to changes in those controls, thus shutting down the prior-based identification channel.

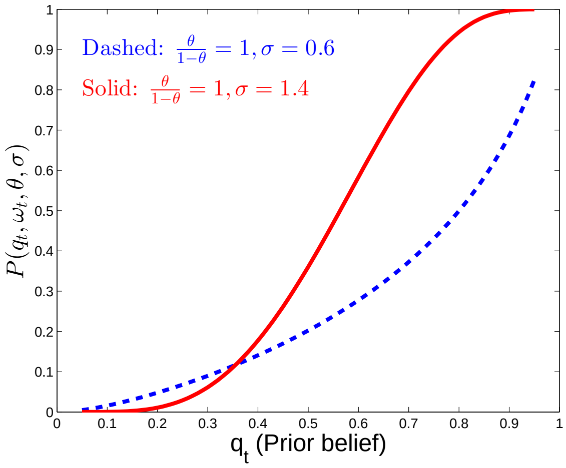
One would expect the failure to account for differential responses to changes in the prior to be particularly problematic when members differ in terms of expertise. As suggested by our previous discussion, when expertise differences between two members are



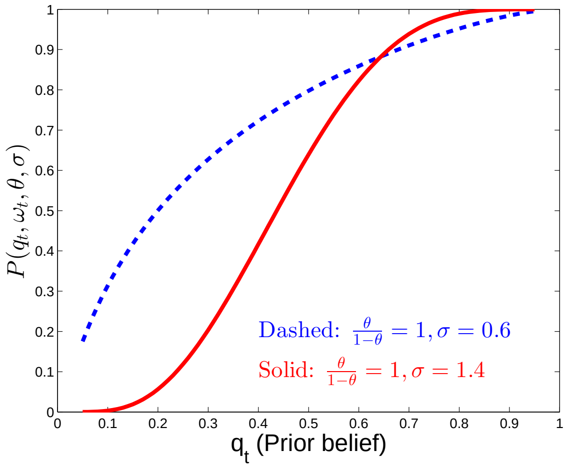
(a) Preference Differences, State=0



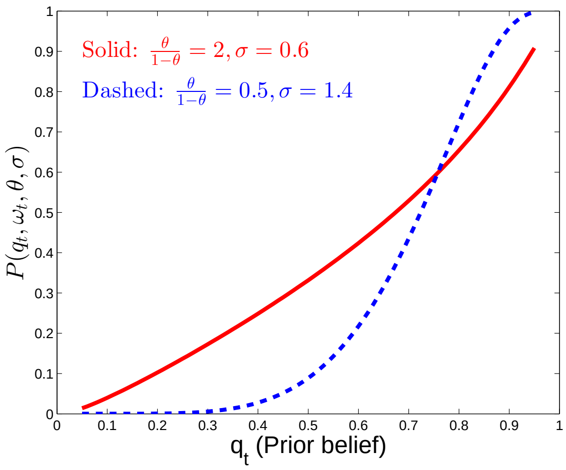
(b) Preference Differences, State=1



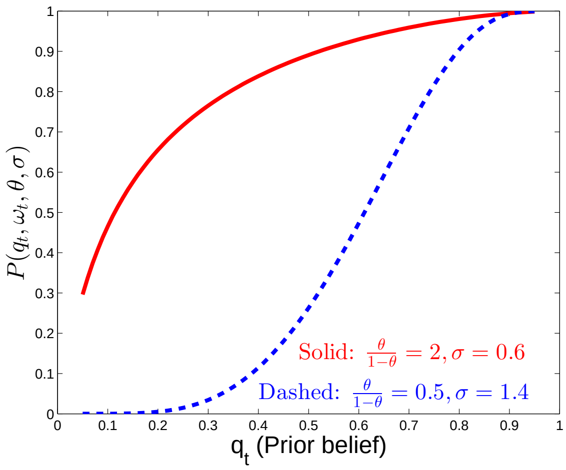
(c) Precision Differences, State=0



(d) Precision Differences, State=1



(e) Preference and Precision Differences, State=0



(f) Preference and Precision Differences, State=1

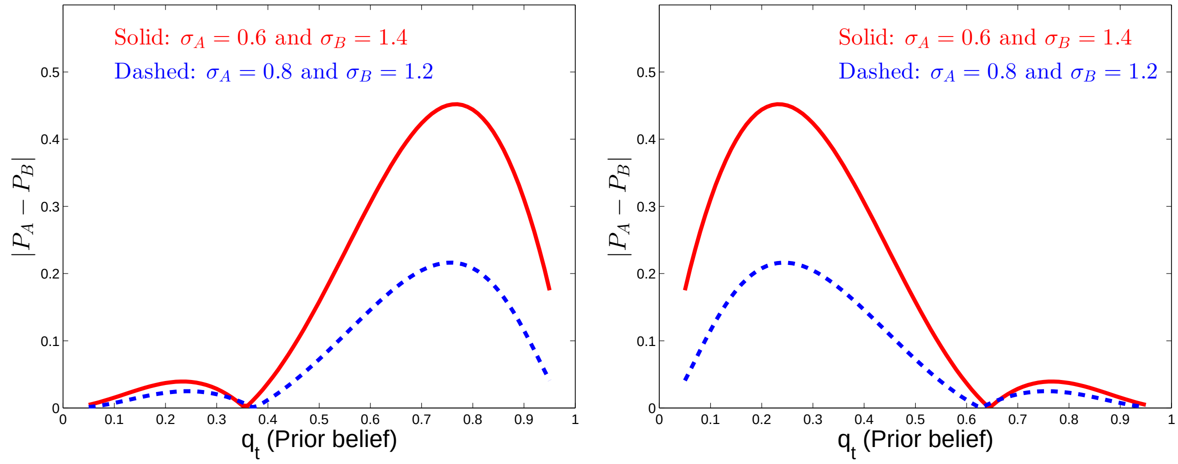
Figure 1: Distinguishing Preferences versus Precisions

Notes: These figures show the theoretical probability that DM chooses $d_t = 1$ as a function of the prior belief that $\omega_t = 1$ (q_t). The different curves represent different combinations of preferences and precisions to show how the probability of selecting $d_t = 1$ changes with the prior.

the most extreme possible (say $\sigma_A \to 0$ and $\sigma_B \to \infty$), member A is totally unresponsive to the prior while member B has a near infinitely high response to changes in prior around $1 - \theta$. In terms of less extreme differences, figure 2 shows how differences in the probability of choosing $d_t = 1$ grow faster over intermediate values of the prior when differences in σ grow. These observations lead to our second claim:

Claim 2 *Controlling for member-specific reactions to changes in q_t is more important for identification of θ and σ differences the larger are σ differences.*

In section 3 below, we show that this claim is true.



(a) State=0 **(b)** State=1

Figure 2: Differences in the Probability of Voting High by State

Notes: These figures show the difference in the theoretical probability that different types of member choose the high decision as a function of the prior belief. In both cases $\theta_A = \theta_B = 0.5$. The different curves represent different differences in the expertise parameter between the members being considered to show how different members react differently to changing priors depending on the gap in their expertise.

We emphasize that we do not challenge the basic identification argument of MIMS, but simply wish to point out that it holds for a fixed prior and exploits differences across states; instead our idea relies on holding fixed a state and exploits differences in the prior. Of course, in actual time-serious data one cannot usually perfectly identify the prior nor the state, so an ideal estimation approach would allow for both sources of identification. To do so, we propose an alternative specification of MIMS' estimator in section 4, which we have applied in Hansen, McMahon, and Velasco (2013).

Before proceeding, it is important to discuss another natural model in which several decision makers' utilities are affected by the same state variable: voting in committees.

Indeed, although their estimator has broader applicability, this is precisely the model that MIMS use to motivate their work. The model we have developed so far is directly applicable to committee voting models if members vote *sincerely*; that is, they behave as if they get utility from matching their vote to the state. In this case, each member votes high ($d_{it} = 1$) whenever $s_{it} \geq s^*_{it}(\theta_i, \sigma_i, q_t)$, as described in (3).

However, if they behave as if they get utility from the committee decision, regardless of what they voted for, and are fully rational, they vote strategically and condition their vote on the probability of changing the decision, or, in the language of jury models, on being pivotal. Although the strategic model is more consistent with economic rationality, the complexity of computing the probability of being pivotal in heterogeneous committees of even modest size under majority is daunting.[7] Rather than undergo such elaborate reasoning, committee members may simply follow the rule of voting for whichever alternative they feel most is most likely to match the state. Also, committee designers sometimes explicitly encourage members to behave sincerely, such as the Bank of England, which tells members of its Monetary Policy Committee to "vote to set interest rates at the level they believe is consistent with meeting the inflation target." For this reason, we feel comfortable interpreting our results as relevant for committee voting in addition to other applications.[8]

3 A New Econometric Methodology to Estimate Independent Decision Data

We begin with an analysis of the case in which there is only variation in the prior that allows us to identify the decision parameters. This is a particularly interesting starting point because it allows us to isolate the effect of our Prior-Based Identification from State-Based Identification. To do this, we focus on estimating the case of a single independent decision maker who acts repeatedly over time; for example, consider a Governor of a central bank who is in sole-charge of making interest-rate decisions such as at the Reserve Bank of New Zealand. This approach also allows us to compare multiple decision makers operating independently at potentially different points in time.

[7] For example, in a nine person committee, each member is pivotal in 70 different events—the number of ways the eight other members can split evenly between voting for 0 and 1. Additionally, within each of these events, the voter must compute the ratios of the probabilities of having arrived at that event in states 1 and 0.

[8] We have also run all simulations under a strategic-voting scenario and generally found the results, including the baseline MIMS specification, so much noisier than for the sincere case that comparing across specifications was difficult. As such, we don't present the results here but they are available on request.

Being able to estimate in such an environment requires one to find good (though not perfect) proxies for the prior because it is not possible to separately estimate the prior q_t (or, econometrically speaking, the mixing probability). Using data to proxy for the prior probability ($\widehat{q_t}$), we can directly estimate the parameters from the model-implied likelihood function L_t^{DM} over the time t decision.

$$L_t^{DM} = q_t \left[1 - \Phi\left(\frac{s_t^*(\cdot) - 1}{\sigma}\right)\right]^{d_t} \left[\Phi\left(\frac{s_t^*(\cdot) - 1}{\sigma}\right)\right]^{1-d_t} + \\ (1 - q_t)\left[1 - \Phi\left(\frac{s_t^*(\cdot) - 1}{\sigma_i}\right)\right]^{d_t} \left[\Phi\left(\frac{s_t^*(\cdot)}{\sigma}\right)\right]^{1-d_t} \tag{5}$$

In order to test this approach, we carry out the following Monte Carlo exercise:

1. Consider a single DM characterized by σ and θ
2. For each of 200 decisions, q_t is drawn from $U[0.2, 0.8]$ (independent across periods)
3. ω_t is drawn from a Bernoulli distribution with $\Pr[\omega_t = 1] = q_t$
4. d_t is drawn from a Bernoulli distribution with $\Pr[d_t = 1] = 1 - \Phi\left(\frac{s_t^* - \omega_t}{\sigma}\right)$ where Φ is the normal cdf and s_t^* is DM's "critical" threshold defined in (3).
5. Generate a proxy $\hat{q}_t = q_t + \varepsilon_t$ where $\varepsilon_t \in U[-x, x]$ for $x \in \{0, 0.05, 0.1\}$
6. Plug in d_t and $\hat{q}_t$ into (5); maximize the corresponding log-likelihood over σ and θ; store values
7. Repeat 1,000 times.

We also explore a simulation with multiple independent decision makers. That is, we modify the above procedure to allow the first 100 decisions to be made by decision maker A, and the second 100 decisions to be made by B; A and B differ in both their preference and expertise parameters. We modify the log-likelihood by allowing the parameters to vary depending on the decision maker we are analyzing.

The results of these simulations are reported in table 1. Although decisions are independent across time and do not exploit State-Based Identification, we generally get reliable estimates if our proxy is sufficiently correlated with the true prior. For example, when $\widehat{q_t}$ has ± 0.05 noise, which we regard as a good proxy, the biases are in the range of 3%-7%. This shows that in order to implement this estimator, it is simply necessary to find data that correlates well with the the probability of the high state having been realized. For example, if studying the behavior of an independent central banker, it would be necessary to get proxies that correlate with the state of the world being inflationary and favoring higher (rather than lower) interest rates. Our results also show that as the level of noise increases (decreases) the performance of our estimator deteriorates (improves). Moreover, for the multiple decision maker case, our approach correctly ranks the different decision-makers' parameters for the vast majority of cases.

4 Incorporating Prior-Based Identification into MIMS's Estimator

In the case of multiple decision makers within a given time-period, one must allow for the correlation of decisions within t and so the period t likelihood of observing the vector of decisions $\mathbf{d}_t$ becomes:

$$\begin{aligned} L_t^{MD} = q_t \prod_i \left[1 - \Phi\left(\frac{s_{it}^*(\cdot) - 1}{\sigma_i}\right)\right]^{d_{it}} \left[\Phi\left(\frac{s_{it}^*(\cdot) - 1}{\sigma_i}\right)\right]^{1-d_{it}} + \\ (1 - q_t) \prod_i \left[1 - \Phi\left(\frac{s_{it}^*(\cdot) - 1}{\sigma_i}\right)\right]^{d_{it}} \left[\Phi\left(\frac{s_{it}^*(\cdot)}{\sigma_i}\right)\right]^{1-d_{it}} \end{aligned} \tag{6}$$

MIMS rewrite (6) as

$$q_t \prod_i (\kappa_{1it})^{d_{it}} (1 - \kappa_{1it})^{1-d_{it}} + (1 - q_t) \prod_i (\kappa_{0it})^{d_{it}} (1 - \kappa_{0it})^{1-d_{it}} \tag{7}$$

where $\kappa_{1it} \equiv 1 - \Phi\left(\frac{s_{it}^*(\cdot)-1}{\sigma_i}\right)$ and $\kappa_{0it} \equiv 1 - \Phi\left(\frac{s_{it}^*(\cdot)}{\sigma_i}\right)$ are the probabilities of deciding $d_{it} = 1$ in states 1 and 0. They then model q_t and the κ terms as functions of observed covariates (potentially time-varying individual characteristics X_{it} and time characteristics Z_t) as follows:

$$q_t = \frac{\exp(\alpha \cdot Z_t)}{1 + \exp(\alpha \cdot Z_t)} \tag{8}$$

and

$$\begin{aligned} \kappa_{0it} &= \frac{\exp(\beta_0 \cdot X_{it} + \beta_1 \cdot Z_t)}{1 + \exp(\beta_0 \cdot X_{it} + \beta_1 \cdot Z_t)} \\ \kappa_{1it} &= \frac{\kappa_{0it} + \exp(\gamma_0 \cdot X_{it} + \gamma_1 \cdot Z_t)}{1 + \exp(\gamma_0 \cdot X_{it} + \gamma_1 \cdot Z_t)}. \end{aligned} \tag{No PBI}$$

The reduced form specification in (No PBI) imposes the restriction that all members, regardless of their underlying heterogeneity, respond to changes in the time specific variables that capture the prior in the same way. As we point out in section 2 above, members with different signal precisions will react differently to changes in the prior. But in this

Table 1: Monte Carlo Estimates of the Single Decision Maker Environment

	Noise Range	$\frac{\theta_A}{1-\theta_A}$	$\frac{\theta_B}{1-\theta_B}$	$\frac{\theta_A}{1-\theta_A} - \frac{\theta_B}{1-\theta_B}$		σ_A	σ_B	$\sigma_A - \sigma_B$	
Single DM									
Type A	**-**	**2**	-	-		**0.6**	-	-	
	0	2.00 *0*				0.61 *2*			
	0.1	2.15 *7*				0.58 *-3*			
	0.2	3.20 *60*				0.46 *-24*			
Type B	**-**	-	**0.5**	-		-	**1.4**	-	
	0		0.50 *0*				1.41 *1*		
	0.1		0.49 *-3*				1.34 *-4*		
	0.2		0.45 *-11*				1.15 *-18*		
2 Ind. DMs	**-**	**2.00**	**0.50**	**1.50**	**Correct Rank (%)**	**0.60**	**1.40**	**0.80**	**Correct Rank (%)**
	0	1.95 *-3*	0.50 *0*	1.45 *-3*	99.6	0.64 *7*	1.45 *3*	0.83 *3*	97.4
	0.1	2.12 *6*	0.49 *-3*	1.63 *9*	99.7	0.60 *0*	1.37 *-2*	0.79 *-2*	96.6
	0.2	2.84 *42*	0.45 *-10*	2.41 *61*	99.4	0.49 *-18*	1.20 *-14*	0.71 *-11*	95.7

Notes: This table shows the Monte Carlo estimates of the decision parameters for different environments characterized by a single decision maker. Bolded numbers are imposed parameters of the decision making problem, while in cells with two numbers the median of the 1000 draws is presented on the left and the percentage deviation from true value is on the right in italics. For the environment in which we compare two independent decision makers, we also list the percentage of the 1000 draws for which the estimated parameters correctly rank the decisions makers in line with the true ranking.

specification there is no Prior-Based Identification, hence the No PBI label.

In order to capture this potentially important channel of identification, we adopt the following more flexible functional forms for the κ terms that include interactions between time and individual characteristics:[9]

$$\begin{aligned}\kappa_{0it} &= \frac{\exp\left(\beta_0 \cdot X_{it} + \beta_1 \cdot Z_t + \beta_2 \cdot X_{it} \cdot Z_t\right)}{1 + \exp\left(\beta_0 \cdot X_{it} + \beta_1 \cdot Z_t + \beta_2 \cdot X_{it} \cdot Z_t\right)} \\ \kappa_{1it} &= \frac{\kappa_{0it} + \exp\left(\gamma_0 \cdot X_{it} + \gamma_1 \cdot Z_t + \gamma_2 \cdot X_{it} \cdot Z_t\right)}{1 + \exp\left(\gamma_0 \cdot X_{it} + \gamma_1 \cdot Z_t + \gamma_2 \cdot X_{it} \cdot Z_t\right)}.\end{aligned} \tag{PBI}$$

This specification incorporates PBI via the γ_2 coefficients that allow responses to changes in Z_t to vary depending on X_{it}.

Under either specification, the estimation of the structural parameters follows a two-step procedure:

1. Estimate the α, β, and γ parameters via the mixture model using maximum likelihood estimation and obtain fitted values $\widehat{q}_t$, $\widehat{\kappa}_{0it}$, and $\widehat{\kappa}_{1it}$.

2. Use the fitted values to recover the structural parameters from the theoretical decision making probabilities. An estimate of period-t expertise comes via

$$\widehat{\sigma}_{it} = \frac{1}{\Phi^{-1}\left(1 - \widehat{\kappa}_{0it}\right) - \Phi^{-1}\left(1 - \widehat{\kappa}_{1it}\right)} \tag{9}$$

and of period-t preferences via

$$\widehat{s}^*_{it} = \frac{\Phi^{-1}\left(1 - \widehat{\kappa}_{0it}\right)}{\Phi^{-1}\left(1 - \widehat{\kappa}_{0it}\right) + \Phi^{-1}\left(\widehat{\kappa}_{1it}\right)} \tag{10}$$

along with $\widehat{q}_t$ and (3) above.

The second stage yields an estimate of preference and precision parameters for each decision maker for each unique value of $\widehat{q}_t$. We consider the median values of these estimates, $\widehat{\theta}_i$ and $\widehat{\sigma}_i$, to be the point estimates.[10]

4.1 Monte Carlo tests of No PBI versus PBI

In order to test the extent to which our Prior-Based Identification matters for the estimation of the multiple decision-maker environment, we again proceed using Monte Carlo analysis. Specifically, we:

[9] We do not claim this is the only possible way of incorporating Prior-Based Identification into the estimation of Bayesian decision problems. It may be the case that more elaborate functional form specifications would outperform our proposal, but we leave this for future research.

[10] For further details on this two-step procedure, see MIMS and Hansen, McMahon, and Velasco (2013).

1. Generate a group of 9 decision makers each making 150 decisions in consecutive time periods.

   (a) 5 members are type A with preferences θ_A and expertise σ_A; 4 members are type B with preferences θ_B and expertise σ_B

   (b) In order to explore the relationship between the importance of Prior-Based Identification and heterogeneity in expertise in light of claim 2 above, we use various parameter values that are "reasonable" in the sense of being in line with estimates in MIMS and Hansen, McMahon, and Velasco (2013). For most of the text we examine $\theta_A = \frac{2}{3}$ and $\theta_B = \frac{1}{3}$, and $\sigma_A = 1 - x$ and $\sigma_B = 1 + x$ for $x \in \{0, 0.05, 0.1, \ldots, 0.5\}$. This means that our baseline comparisons are for eleven unique sets of parameters. As a robustness exercise we keep the same values for σ but take $\theta_A = \frac{1}{3}$ and $\theta_B = \frac{2}{3}$.

2. For each unique set of θ and σ values, we run 1,000 simulations. For each simulation, we generate theoretical decision data according to the following procedure:[11]

   (a) In each period t, q_t is drawn from $U[0.2, 0.8]$ (independent across periods)

   (b) ω_t is drawn from a Bernoulli distribution with $\Pr[\,\omega_t = 1\,] = q_t$

   (c) d_{it} is drawn from a Bernoulli distribution with $\Pr[\,d_{it} = 1\,] = 1 - \Phi\left(\frac{s^*_{it} - \omega_t}{\sigma_i}\right)$ where Φ is the normal cdf and s^*_t is DM's "critical" threshold defined in (3).

3. Given these data, we construct $Z_t = (\mathbf{1}, q_t)$ and $X_{it} = (\mathbf{1}, D_A)$, where D_A is a dummy variable that indicates membership of group A (and thus not actually time-varying). We use these data to estimate two separate specifications of the first-stage regressions given by (No PBI) and (PBI).

4. After we obtain estimates of first-stage coefficients, we use structural equations to back out $\widehat{\theta}_x$ and $\widehat{\sigma}_j$ for $j \in \{A, B\}$ as described above.

Figure 3, which shows the the percentage bias for each value of the expertise difference, summarizes the main results of the simulation exercise.[12] When expertise differences are small, the results confirm that PBI does not outperform state-based identification alone; the estimates of the parameter levels and differences are estimated reasonably accurately in both cases. However, as $\sigma_A - \sigma_B$ increases, the estimates that do not allow for PBI

[11] All estimation is done in R via maximum likelihood using the BFGS algorithm. All code is available on request.

[12] In the online appendix, we present a table which shows the true value, the point estimate and the percentage bias for each parameter as we gradually increase the expertise difference. We also repeat this figure for the level of the bias.

deteriorate quickly, especially in the estimates of the differences between groups, while our proposed specification actually improves in accuracy. For example, when $\sigma_A - \sigma_B = 0.8$, our PBI specification estimates $\frac{\theta_A}{1-\theta_A} - \frac{\theta_B}{1-\theta_B}$ and $\sigma_A - \sigma_B$ to 3% accuracy,[13] whereas the No PBI specification displays biases of 20% and 70% respectively.

Figure 4 shows that this main result is unaffected by our decision to make the type A decision makers more expert. If instead we maintain the assumption that $\theta_A > \theta_B$ but make A less expert, we change the direction of the bias in the levels estimates of θ, but the over-estimation of the difference remains as expertise differences grow.[14]

Finally, we plot the complete distribution of the simulation results for the cases of $\sigma_A = 1$ and $\sigma_B = 1$ (figure 5), and $\sigma_A = 0.6$ and $\sigma_B = 1.4$ (figure 6).[15] With no σ differences, the results are almost identical and PBI may even do slightly worse for some parameters. But even at relatively modest expertise differences, the results show that not only does the PBI specification ensure that the results stay anchored around the true parameters, but also that the distribution around the estimates is less dispersed too. This further supports our argument that this is an important channel of identification.

4.2 Re-estimation of US Supreme Court Justice Characteristics

In order to test our procedure on real data, we conclude this section by re-estimating the structural parameters for the US Supreme Court voting data that MIMS consider. Their dataset contains the vote of every justice (31 in total) on every case from 1953-2008. $d_{it} = 1$ corresponds to a vote for the plaintiff in a legal case, and $d_{it} = 0$ to a vote for the defendant. They run separate regressions on four subsets of cases according to the issue at stake (business, basic rights, criminal, federalism).[16] We focus on the results for economics and basic rights cases, the two subsets MIMS treat as their baseline cases.

The first specification we run on the data is (No PBI), taking X_{it} and Z_t as the same sets of variables that MIMS use; that is, we follow the approach and data of MIMS.[17] The second is to run a modified version of (PBI) in which we interact what appears to us to be the relevant subset of individual and meeting characteristics for influencing justices' prior beliefs.[18]

[13] In the tables we report estimates of $\frac{\theta}{1-\theta}$ so that biases and expertise are presented on the same on the same range $\mathbb{R}_+$.

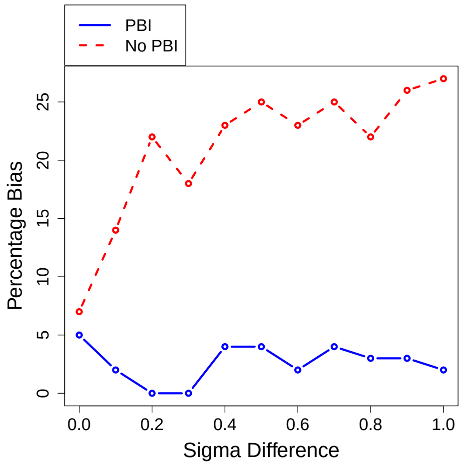
[14] In the online appendix we present the table of these alternative results.

[15] The online appendix contains the complete distirubtions for all the cases.

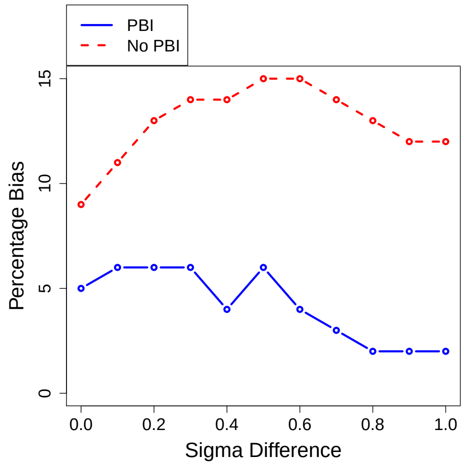
[16] They exclude the small fraction of cases without nine votes

[17] There are two reasons that we cannot simply take results straight from MIMS. First, when we estimate the first-stage coefficients taking their reported estimates as starting values, we obtain new estimates that reduce the value of the log-likelihood. Second, they do not report the median value of the structural parameters across all values of the fitted priors.

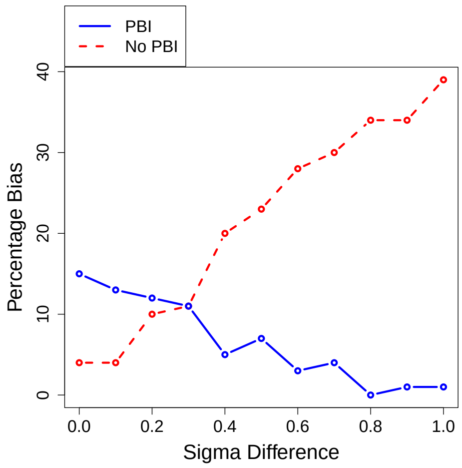
[18] We do not interact the mean value of *other* justices' Segal-Cover ideology or quality scores—covariates



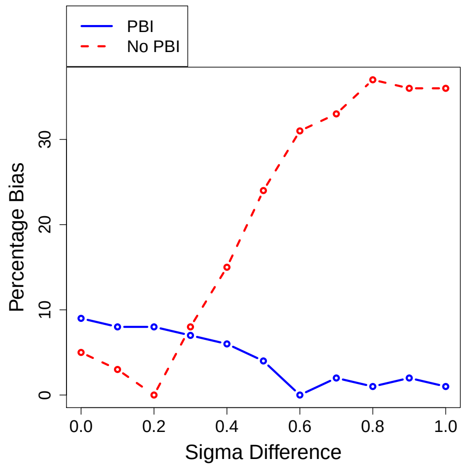
(a) Bias in $\frac{\theta_A}{1-\theta_A}$ Estimates



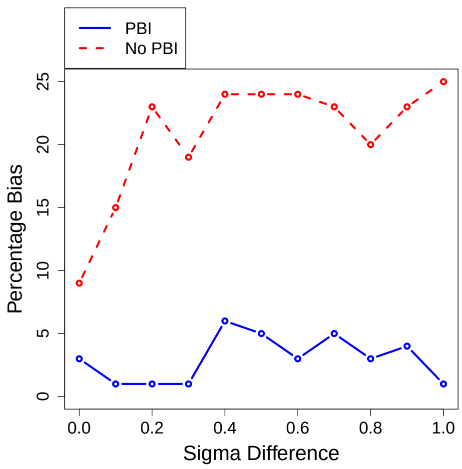
(b) Bias in σ_A Estimates



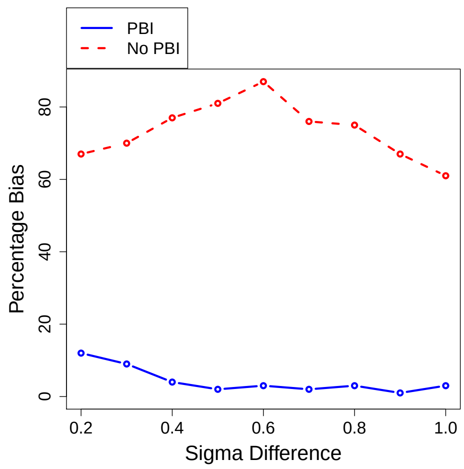
(c) Bias in $\frac{\theta_B}{1-\theta_B}$ Estimates



(d) Bias in σ_B Estimates



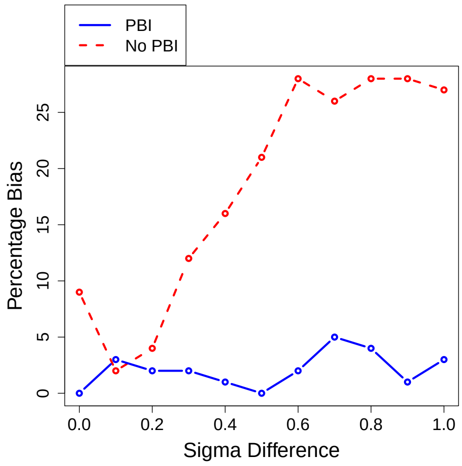
(e) Bias in $\frac{\theta}{1-\theta}$ Difference Estimates



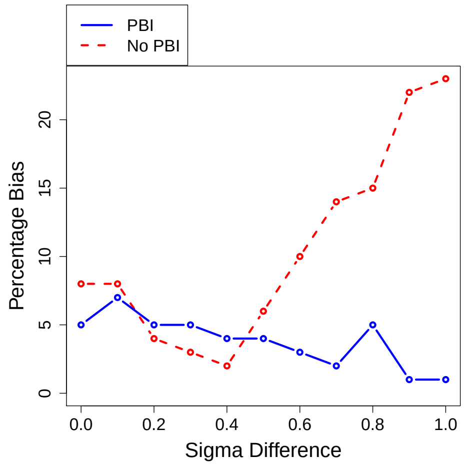
(f) Bias in σ Difference Estimates

Figure 3: Biases of Estimates (Percent)

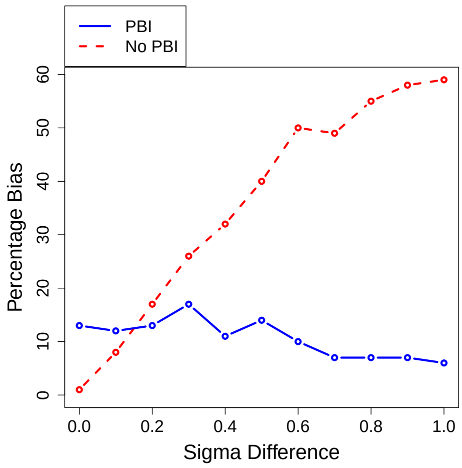
Notes: These figures plot the estimated values as a percentage of the true value (percentage of bias) for the baseline case of $\theta_A = \frac{2}{3}$ and $\theta_B = \frac{1}{3}$ for different values (along the horizontal axis) of the expertise difference (σ_A falls while σ_B increases). The first column reports the results for $\frac{\theta_A}{1-\theta_A}$ (row 1), $\frac{\theta_B}{1-\theta_B}$ (row 2) and the difference between these quantities (row 3). The second column shows σ_A (row 1), σ_B (row 2) and $\sigma_A - \sigma_B$ (row 3).



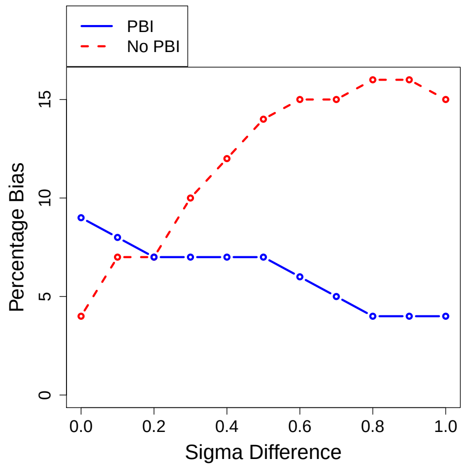
(a) Bias in $\frac{\theta_A}{1-\theta_A}$ Estimates



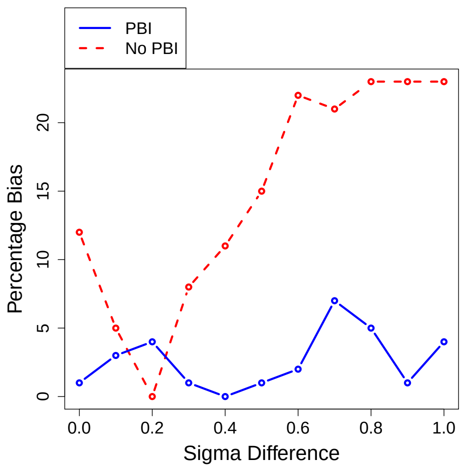
(b) Bias in σ_A Estimates



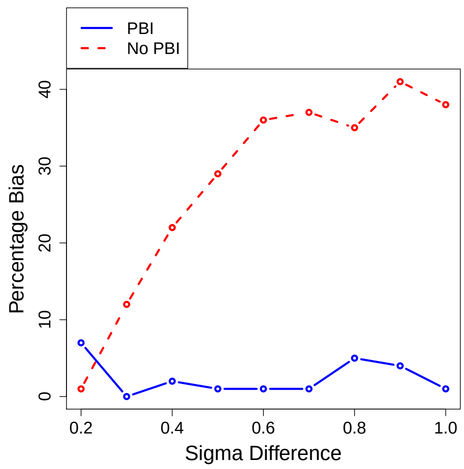
(c) Bias in $\frac{\theta_B}{1-\theta_B}$ Estimates



(d) Bias in σ_B Estimates



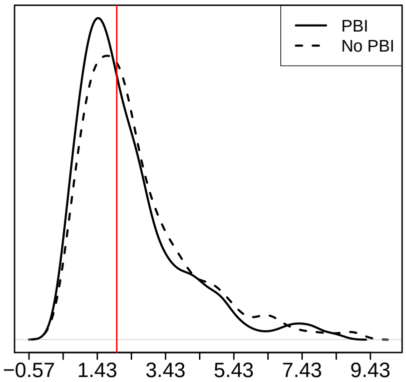
(e) Bias in $\frac{\theta}{1-\theta}$ Difference Estimates



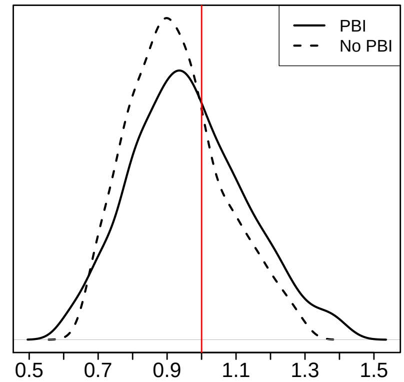
(f) Bias in σ Difference Estimates

Figure 4: Biases of Estimates - Reversed θ

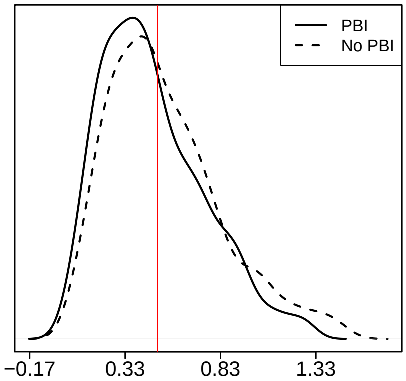
Notes: These figures plot the estimated values as a percentage of the true value (percentage of bias) for the alternative case of $\theta_A = \frac{2}{3}$ and $\theta_B = \frac{1}{3}$ for different values (along the horizontal axis) of the expertise difference (now σ_A increases while σ_B falls). The first column reports the results for $\frac{\theta_A}{1-\theta_A}$ (row 1), $\frac{\theta_B}{1-\theta_B}$ (row 2) and the difference between these quantities (row 3). The second column shows σ_A (row 1), σ_B (row 2) and $\sigma_A - \sigma_B$ (row 3).



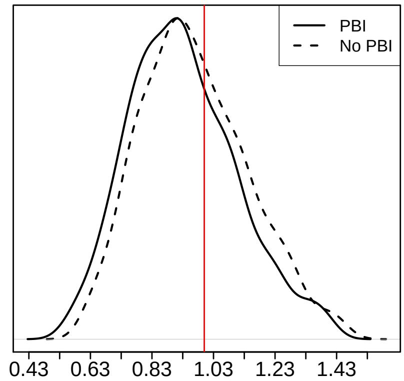
(a) Density of $\frac{\theta_A}{1-\theta_A}$ Estimates



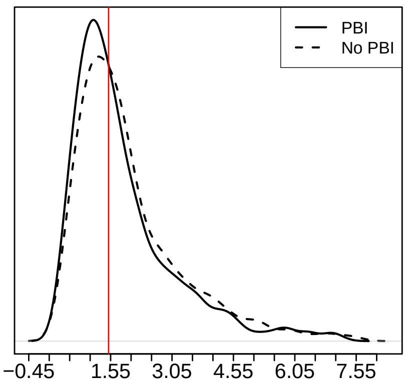
(b) Density of σ_A Estimates



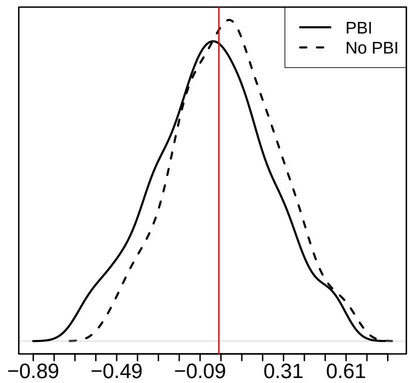
(c) Density of $\frac{\theta_B}{1-\theta_B}$ Estimates



(d) Density of σ_B Estimates



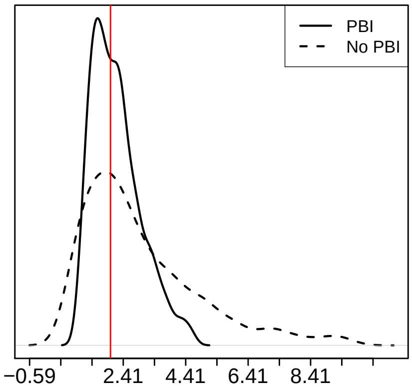
(e) Density of $\frac{\theta_A}{1-\theta_A} - \frac{\theta_B}{1-\theta_B}$ Estimates



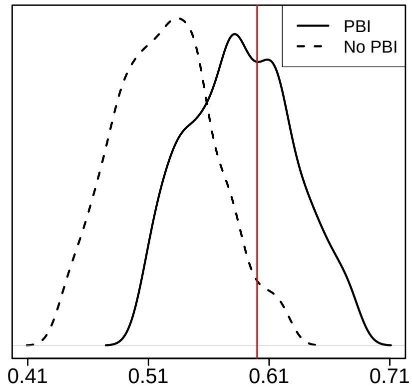
(f) Density of $\sigma_A - \sigma_B$ Estimates

Figure 5: Densities of Estimates with $\sigma_A = 1$ and $\sigma_B = 1$

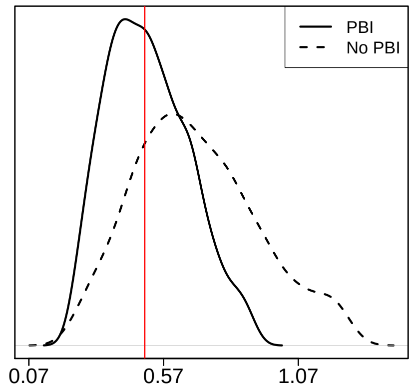
Notes: These figures plot the complete distribution of the simulation results for the case of $\sigma_A = 1$ and $\sigma_B = 1$. The first column reports the results for $\frac{\theta_A}{1-\theta_A}$ (row 1), $\frac{\theta_B}{1-\theta_B}$ (row 2) and the difference between these quantities (row 3). The second column shows σ_A (row 1), σ_B (row 2) and $\sigma_A - \sigma_B$ (row 3).



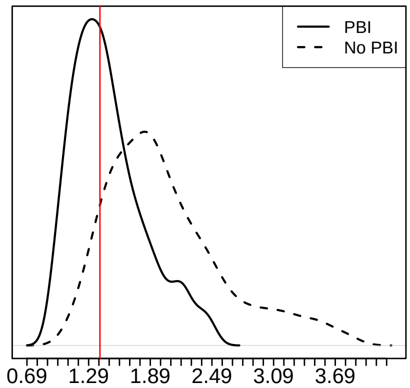
(a) Density of $\frac{\theta_A}{1-\theta_A}$ Estimates



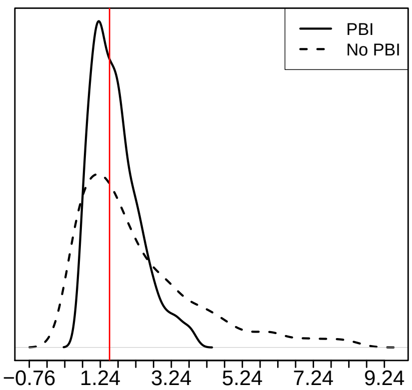
(b) Density of σ_A Estimates



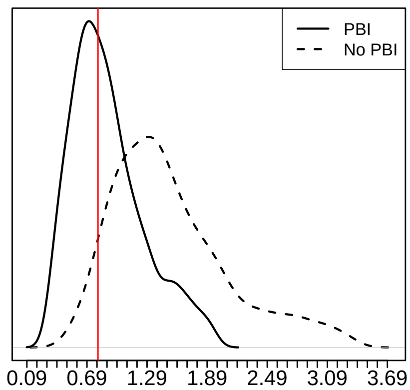
(c) Density of $\frac{\theta_B}{1-\theta_B}$ Estimates



(d) Density of σ_B Estimates



(e) Density of $\frac{\theta_A}{1-\theta_A} - \frac{\theta_B}{1-\theta_B}$ Estimates



(f) Density of σ_D Estimates

Figure 6: Densities of Estimates with $\sigma_A = 0.6$ and $\sigma_B = 1.4$

Notes: These figures plot the complete distribution of the simulation results for the case of $\sigma_A = 0.6$ and $\sigma_B = 1.4$. The first column reports the results for $\frac{\theta_A}{1-\theta_A}$ (row 1), $\frac{\theta_B}{1-\theta_B}$ (row 2) and the difference between these quantities (row 3). The second column shows σ_A (row 1), σ_B (row 2) and $\sigma_A - \sigma_B$ (row 3).

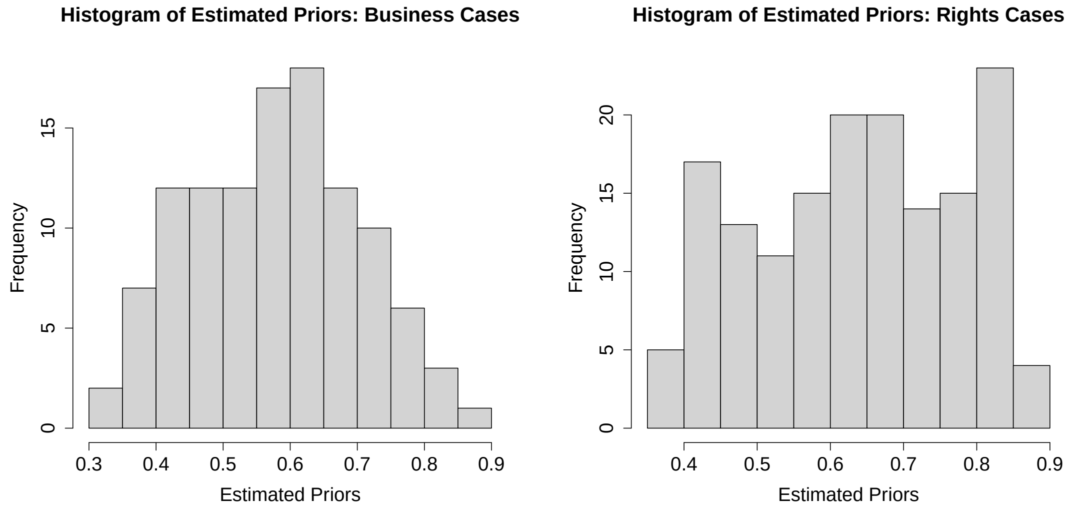


Figure 7: Histograms of Estimated Priors

Notes: This figure plots, for business cases (left figure) and rights cases (right figure), histograms of the estimated priors q_t from the (No PBI) specification.

The first point of interest is that there is a large range of estimated priors in the voting data. Figure 7 plots histograms of the estimated priors from the (No PBI) specification (the results for the (PBI) specification are very similar), and shows they range from around 0.3 to around 0.9, with a fairly dispersed distribution. This variation in the prior indicates that allowing for PBI is potentially important in this dataset.

Our two specifications each produce 31 estimates (corresponding to the number of justices) of θ[19] and σ for business and rights cases. Table 2 displays a number of summary statistics related to the distributions of these estimates. The main message from our simulation exercises is that not using PBI tends to inflate estimated differences between decision makers. As the table shows, PBI reduces justice heterogeneity both in terms of variances and ranges. For rights case this reduction is particularly notable: the variance with PBI is around 1/6 the value of the variance without. This illustrates that PBI can have substantial effects in real world datasets.

In order to compare the distributions of PBI and No PBI estimates more directly, the radar charts in figure 8 are helpful. Justices are ordered lowest to highest moving clockwise based their No PBI estimates. Within this disc we plot both sets of estimates. A distribution with less heterogeneity produces a more circular plot; the PBI estimates,

within X_{it}—with any Z_t variables, nor chief justice dummies—covariates within Z_t—with any X_{it} variables. They are included within X_{it} and Z_t without interactions in the κ and q terms.

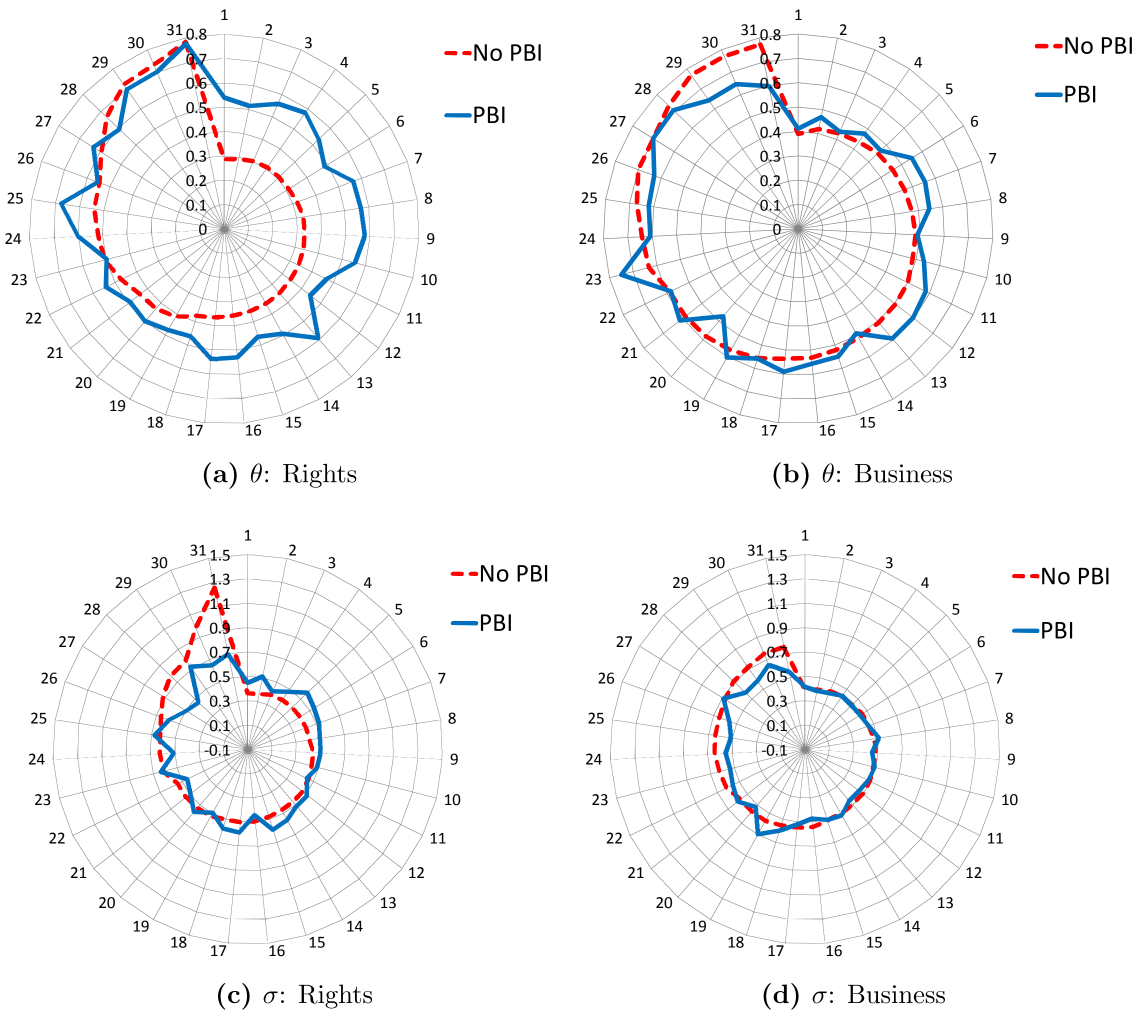
[19]Here we return to reporting θ rather than $\frac{\theta}{1-\theta}$ in order to maintain comparability with MIMS.

Table 2: Re-estimation Exercise

	Business Cases				**Rights Cases**			
	θ estimates		σ estimates		θ estimates		σ estimates	
	No PBI	**PBI**	**No PBI**	**PBI**	**No PBI**	**PBI**	**No PBI**	**PBI**
Variance	0.013	0.006	0.011	0.006	0.020	0.006	0.037	0.006
Inter-quartile Range	0.1597	0.0993	0.1532	0.1051	0.1826	0.0925	0.1925	0.0924
Minimum	0.390	0.411	0.396	0.392	0.285	0.446	0.360	0.415
Median	0.532	0.582	0.543	0.516	0.358	0.540	0.492	0.515
Maximum	0.773	0.750	0.759	0.699	0.784	0.775	1.255	0.726

Notes: This table shows various measures of dispersion of the distribution across judges of the estimated values of θ and σ when we use both the flexible specification (PBI) and the original MIMS specification (No PBI).

particularly for rights cases, display less heterogeneity (are more circular).



(a) θ: Rights

(b) θ: Business

(c) σ: Rights

(d) σ: Business

Figure 8: Radar Plots of Supreme Court Data Re-estimation Exercise

Notes: These figures show, for θ (row 1) and σ (row 2), the estimate of each Justice's parameter using MIMS specification (No PBI) along with the equivalent parameter estimated under the more flexible specification (PBI). In each case, the Justices are ordered lowest to highest moving clockwise based their No PBI estimates. Column 1 refers to Rights Cases and column 2 to Business Cases.

5 Conclusion

Given the high level of interest within economics in how individuals and groups of individuals make decisions under uncertainty, it is important that we can can empirically estimate the different channels that drive the behavior of decision makers to test proposed theories. We have provided an important channel through which it is possible to empirically identify the decision-making parameters in standard Bayesian decision problems; our approach relies on variation in decision-making behavior over heterogeneous priors and differs from the most important existing contribution, which relies on comparing a group of decision makers across different states for a fixed prior.

While there are likely still a number of steps that can be taken within this growing empirical literature, our proposed identification strategy can be viewed as an important contribution in two dimensions. First, unlike the existing method, our approach can be used to estimate decision-making parameters of single decision makers, and of decision makers serving at different points in time or taking independent decisions. Second, we show that where there is greater heterogeneity of expertise amongst decision makers operating contemporaneously, ignoring our proposed channel of identification becomes increasingly costly in terms of accuracy of estimated parameters. Fortunately, there are relatively straightforward ways of augmenting existing approaches and we show that as expertise heterogeneity grows, our proposed specification is increasingly accurate relative to the existing methods.

References

Banerjee, A. V. (1992): "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics*, 107(3), 797–817.

Besley, T. (2006): *Principled Agents?: The Political Economy of Good Government*, The Lindahl Lectures. OUP Oxford.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992): "A Theory of Fads, Fashion, Custom, and Cultural Change in Informational Cascades," *Journal of Political Economy*, 100(5), 992–1026.

Gerling, K., H. P. Gruner, A. Kiel, and E. Schulte (2005): "Information acquisition and decision making in committees: A survey," *European Journal of Political Economy*, 21(3), 563–597.

Hansen, S., M. McMahon, and C. Velasco (2013): "How Experts Decide: Preferences or Private Assessments on a Monetary Policy Committee?," manuscript.

Iaryczower, M., G. Lewis, and M. Shum (2013): "To Elect or to Appoint? Bias, Information, and Responsiveness of Bureaucrats and Politicians," *Journal of Public Economics*, forthcoming.

Iaryczower, M., and M. Shum (2012): "The Value of Information in the Court: Get It Right, Keep It Tight," *American Economic Review*, 102(1), 202–37.

Li, D. (2012): "Information, Bias, and Efficiency in Expert Evaluation: Evidence from the NIH," unpublished manuscript, Kellogg School of Management, Northwestern University.

Sorensen, P., and M. Ottaviani (2000): "Herd Behavior and Investment: Comment," *American Economic Review*, 90(3), 695–704.