

# Accounting for uncertainty: an application of Bayesian methods to accruals models

Matthias Breuer<sup>1</sup> . Harm H. Schütt<sup>2</sup>

Accepted: 16 September 2021 Published online: 19 October 2021  $\ensuremath{\textcircled{O}}$  The Author(s) 2021

## Abstract

We provide an applied introduction to Bayesian estimation methods for empirical accounting research. To showcase the methods, we compare and contrast the estimation of accruals models via a Bayesian approach with the literature's standard approach. The standard approach takes a given model of normal accruals for granted and neglects any uncertainty about the model and its parameters. By contrast, our Bayesian approach allows incorporating parameter and model uncertainty into the estimation of normal accruals. This approach can increase power and reduce false positives in tests for opportunistic earnings management as a result of better estimates of normal accruals and more robust inferences. We advocate the greater use of Bayesian methods in accounting research, especially since they can now be easily implemented in popular statistical software packages.

**Keywords** Bayes  $\cdot$  Prediction  $\cdot$  Accruals  $\cdot$  Earnings management  $\cdot$  Measurement uncertainty

JEL Classification  $\,C11\cdot C53\cdot M40$ 

# 1 Introduction

A central theme in accounting is the measurement of latent constructs (e.g., economic earnings, managerial effort, or abnormal accruals). Such measurement is inherently uncertain (e.g., Barker et al. 2020). To account for the uncertainty, the Bayesian framework provides a formal approach (e.g., Gelman et al. 2013) that is commonly used in accounting theory (e.g., Johnstone 2018). In empirical accounting research, by contrast, the formal Bayesian approach is uncommon, despite its conceptual

Matthias Breuer mb4468@gsb.columbia.edu

<sup>&</sup>lt;sup>1</sup> Columbia University, 3022 Broadway, New York, NY 10027, USA

<sup>&</sup>lt;sup>2</sup> Tilburg University, Warandelaan 2, 5000 LE, Tilburg, Netherlands

appeal.<sup>1</sup> An important barrier to adoption has been the Bayesian framework's computational requirements. But as a result of increasing computational power, this barrier has diminished in recent years. A remaining barrier is unfamiliarity.

We provide an applied introduction to key methods of Bayesian estimation relevant for the measurement of latent constructs, including Bayesian regression, Bayesian hierarchical modelling, and Bayesian model averaging. To illustrate the methods, we apply them to a central concern of accounting researchers, investors, and regulators: the detection of accruals-based earnings management (e.g., Healy and Wahlen 1999; Dechow et al. 2010; Lewis 2012). A key challenge in detecting managed abnormal accruals is researchers' uncertainty about the determinants of normal accruals, absent earnings management (Ball 2013). Faced with this uncertainty, researchers default to models of normal accruals to detect accruals-based earnings management. Several studies, for example, use a variant of the seminal Jones (1991) model. But reliance on any given model and its estimates risks overstating the confidence of conclusions by producing false positives. The likely false positives are not only troubling to the literature (e.g., Ball 2013) but also concerning investors allocating scarce capital and regulators allocating enforcement resources (e.g., Lewis 2012).<sup>2</sup>

Using the popular Jones (1991) model, we first conceptually discuss and then empirically document how Bayesian methods increase power and reduce false positives in tests of accruals-based earnings management by combining various sources of information (e.g., priors and data), according to their relative (un)certainty. The Jones model provides a useful example because its OLS-based estimation is well known, allowing a contrast with the Bayesian approach. In addition, the Jones model, while widely employed, is strongly criticized for producing false positives (e.g., Dechow et al. 2010; Ball 2013), allowing us to tackle a first-order problem of this important model.

Bayesian estimation *per se* does not solve all the issues with the Jones model, nor does it eliminate the need for model fixes and alternative models. It instead allows combining and assessing the importance of various fixes and alternatives via Bayesian model averaging, a technique increasingly applied to fundamental questions in economics (for a review, see Steel 2020). Given this feature, we use the basic modified Jones model (Dechow et al. 1995) for the purpose of exposition, but note that the conceptual benefits of the proposed Bayesian estimation only increase with the incorporation of more and more recent models used in the literature and practice (Piironen and Vehtari 2017). In this spirit, we apply our Bayesian approach to an extended set of models, including popular and recent ones (e.g., Dechow et al.

<sup>&</sup>lt;sup>1</sup>While Glaeser and Guay (2017) advocate a Bayesian approach to inference in accounting research, empirical studies using formal Bayesian methods have only recently emerged (e.g., Du et al. 2020; Zhou 2021).

<sup>&</sup>lt;sup>2</sup>Ball (2013) voices concerns that limited knowledge of the determinants of accruals results in bias and false positives. To address the issue, several studies have proposed improvements to the Jones model (e.g., Kothari et al. 2005; Hribar and Collins, 2002) or new models (e.g., Dechow and Dichev 2002; Gerakos and Kovrijnykh 2013; Bloomfield et al. 2017; Beyer et al. 2019; Nikolaev 2018; Du et al. 2020). The issue of false positives, arising from the neglect of model and parameter uncertainty, by contrast, remains largely unaddressed.

1995; McNichols 2002; Ball and Shivakumar 2006; Collins et al. 2017; Frankel and Sun 2018). The resulting model-averaged estimates of abnormal accruals can readily be used by researchers interested in explaining or controlling for accruals-based earnings management.<sup>3</sup>

To conceptually discuss the benefits of our Bayesian approach, we compare and contrast it with the literature's standard approach to detecting accruals-based earnings management. The standard approach entails three steps. First, researchers choose a model for normal accruals (i.e., accruals explained by fundamental forces), usually the modified Jones model (Dechow et al. 1995).<sup>4</sup> Next, they estimate the model's parameters (i.e., the regression coefficients). To account for heterogeneity in the accruals process, they typically estimate the model separately for different groups of observations—usually industry-years with a minimum amount of observations. Finally, researchers use the estimated model parameters to generate a point prediction of normal accruals for a given firm-year. Any deviation from this prediction is classified as abnormal (i.e., accruals potentially affected by managerial discretion). A crucial point is that this approach shoves any uncertainty about the predicted level of normal accruals into the abnormal accruals measure.

The Bayesian approach allows for an explicit accounting for researchers' prediction uncertainty in abnormal accruals estimation. It can account for the two central components of prediction uncertainty: uncertainty about the parameters of a chosen model (e.g., the coefficient values of the Jones model) and uncertainty about the model to choose. The Bayesian estimation of a given model's parameters produces a posterior distribution for each of the parameter values. Equipped with these distributions, we can obtain a distribution of normal accruals for each firm-year, instead of merely a point prediction. This distribution, called posterior predictive distribution, incorporates the parameter uncertainty, given a chosen model. Notably, this distribution can be calculated for several variants of the Jones model and alternatives. Using Bayesian averaging techniques, the various models' accruals predictions can be combined to yield a model-averaged predictive distribution of normal accruals. This distribution explicitly incorporates the uncertainty about the best model (within the set of considered models). We can use this distribution of normal accruals to discern the extent of plausibly abnormal accruals. This extent can be expected to be substantially reduced, compared to the extent implied by the standard approach, which conflates abnormal accruals and prediction uncertainty.

Following the conceptual discussion, we implement the Bayesian approach to empirically assess its benefits. We explicitly consider two model variants in our empirical implementation: models with varying levels of parameter heterogeneity and those with varying extents of controls. With respect to parameter heterogeneity (across groups), we implement a model allowing for parameters varying at the industry-year level (consistent with the literature, e.g., DeFond and Jiambalvo 1994)

<sup>&</sup>lt;sup>3</sup>To facilitate the adoption of our approach, we provide the code for our Bayesian estimation publicly. We also provide a dataset containing the means and standard deviations of abnormal accruals produced by our averaged model for each firm-year in our sample.

<sup>&</sup>lt;sup>4</sup>Following the literature, we use the modified Jones model, which includes the change in cash revenues, instead of the change in total revenues, as a determinant of accruals, in all of our analyses.

and a model allowing for firm-level heterogeneity (in the spirit of Owens et al. 2017). We implement the parameter heterogeneity using Bayesian hierarchical modelling, which uses two sources of information about a group's parameters, that is, withinand across-group data. The less within-group information available for a particular group, the more that group's coefficient estimates rely on the information provided by other groups. This cross-group learning obviates any ad hoc cutoffs and allows accommodating even firm-level heterogeneity. With respect to the extent of controls, we implement a model with the basic determinants included in the modified Jones model and one model with performance, financing, and volatility controls (in the spirit of prominent model critiques, e.g., Kothari et al. 2005).

Our empirical implementation confirms the conceptual benefits of the Bayesian approach and yields additional insights. First, we find that the firm-level model with extensive controls receives the greatest weight in the Bayesian model's averaging stage. This finding highlights that it is not only important to control for fundamental factors, such as performance but also to allow for firm-level heterogeneity. To date, firm-level heterogeneity in accruals models has been neglected, because of concerns about the required number of observations (per firm). The Bayesian hierarchical approach circumvents this constraint, allowing an accounting for a relevant level of heterogeneity in firms' accruals processes.

Second, we find that the distribution of plausible normal accruals, after accounting for parameter and (selected) model uncertainty, is quite wide. This finding suggests the large extent of abnormal accruals (e.g., in terms of percentage of total assets) observed in the literature largely reflects parameter and model uncertainty, rather than truly unexpected or abnormal accruals.

Third, we document that the predicted normal (abnormal) accruals of the Bayesian averaged model are more (less) correlated with total accruals than the normal (abnormal) accruals obtained using the literature's standard approach, even after controlling for performance, financing, and volatility. This finding suggests that combining various models, including models with firm-level heterogeneity, substantially improves our ability to model firms' accruals. As long as the vast amount of firms' accruals is driven by factors other than opportunistic reporting incentives, such improved model fit translates into a reduced risk of false positives (e.g., Dechow et al. 2010).

Lastly, we find that, in simulations with randomly assigned true and false earnings management indicators, the Bayesian approach increases the power of earnings management tests and reduces the incidence of false positives, relative to the standard approach. This improvement is particularly evident in cases when accruals processes vary at the firm level, instead of the industry-year level, *and* when there is high uncertainty (e.g., among young firms with few observations and volatile fundamentals). Notably, our earlier evidence suggests accruals processes indeed primarily vary at the firm, not industry-year, level. Moreover, the literature often uses opportunistic incentive variables correlated with uncertainty. In these cases (missspecified model and limited or noisy data), our simulations suggest that the standard approach can frequently yield false positives, while a Bayesian approach provides more powerful and robust inferences.

In a final step, we extend our Bayesian estimation to incorporate accruals models besides the Jones model. Following a review of the recent literature, we consider industry-year and firm-level variants of the models developed by Dechow et al. (1995), Kothari et al. (2005), Dechow and Dichev (2002), McNichols (2002), Ball and Shivakumar (2006), Collins et al. (2017), and Frankel and Sun (2018). The averaged model gives nonzero weights to all models, documenting the importance of including various empirical models. The greatest weight is put on more recent models, especially Frankel and Sun (2018), and model variants allowing for firm-level heterogeneity, consistent with our previous finding.

We assess the usefulness of the averaged model by examining its ability to explain future performance and popular earnings management indicators (AAERs, restatements, and comment letters). We find that the normal accruals produced by our averaged model explain future performance better than those of the individual (OLSbased) models, while its (absolute) abnormal accruals exhibit a stronger relation with earnings management indicators than those of the individual models. Collectively, these findings indicate the usefulness of the normal and abnormal accruals estimates produced by our averaged model.

Our paper contributes to the accounting literature by illustrating key methods of Bayesian estimation of relevance to empirical accounting research. Bayesian methods, while conceptually appealing, have received limited attention in applied work until recently, because they require computational power and methodological expertise. These barriers to their application, however, are vanishing, as a result of the rapid expansion of computational power and the implementation of ready-made Bayesian methods into popular applied statistics programs (e.g., R and STATA). Taking advantage of these trends, recent work in accounting has adopted select Bayesian methods (e.g., hierarchical modelling) to uncover latent constructs, such as a new earnings quality measure (Du et al. 2020) and investor learning (Zhou 2021). We add to this emerging stream of the literature by providing a timely illustration of various Bayesian methods, including Bayesian regression, Bayesian hierarchical modelling, and, in particular, Bayesian model averaging. This illustration should help make the new empirical tools more accessible and reduce barriers to adoption.<sup>5</sup>

Our paper also contributes to the extensive literature on accruals-based earnings management. We suggest a novel estimation approach to address a central, unaddressed criticism of the literature: researchers' uncertainty in specifying normal accruals and the corresponding concern about false positives. Our proposed approach provides a natural way to incorporate uncertainty into the prediction of normal accruals derived from accruals models. It improves the literature's standard approach in two key ways. First, it uses more information and models, leading to better predictions and greater power in tests for (opportunistic) abnormal accruals. Second, it accounts for model and parameter uncertainty, reducing the incidence of false positives.

<sup>&</sup>lt;sup>5</sup>We view the Bayesian methods as additional tools for accounting researchers, instead of a substitute for familiar frequentist methods (including recent machine-learning techniques). Indeed, several features of our Bayesian approach can be mimicked by frequentist methods (e.g., regularization and model averaging). We focus on the Bayesian approach because it provides a coherent framework, which nests various methods to account for uncertainty about models and their parameters.

More broadly, Bayesian estimation promises to enhance the measurement and identification of several other constructs of interest, such as productivity (Syverson 2011), investment inefficiency (Biddle and Hilary 2006; Biddle et al. 2009), real earnings management (Roychowdhury 2006), abnormal R&D activity (Kothari et al. 2015), abnormal tone (Huang et al. 2014), and abnormal readability (Bushee et al. 2018). Each of these constructs, like abnormal accruals, is measured using models predicting normal behavior, which researchers are uncertain about. Using the familiar example of abnormal accruals as an application, we provide a relevant and timely introduction to key methods of Bayesian estimation and explain how they can help researchers with their predictions.

Before delving into the details of our approach, we acknowledge that the presented Bayesian methods, while eliminating several ad hoc choices typically made by researchers (e.g., cutoff values for industry-year observations), are not free from subjectivity. Researchers must specify the distributions representing the beliefs and data. In particular, they must explicitly specify prior beliefs. This feature often invokes criticism. We note that prior beliefs are usually chosen such that the data, not the prior, chiefly determine the posterior belief. In cases of limited data and reasonable expectations about plausible values (e.g., regarding the range of economically plausible coefficient values), however, we actually view the ability to inject information into the prior as an advantage, not a drawback, of Bayesian methods. This feature prevents over-fitting the data and enables cross-study learning and truly cumulative empirical work.<sup>6</sup>

# 2 Concepts

## 2.1 Prediction problem

Researchers need to know firms' normal accruals process to discern abnormal deviations, due to managerial discretion and opportunism. However, this sort of knowledge has remained elusive to date. Absent perfect knowledge of the normal process, the modelling of the normal process and measurement of abnormal accruals is akin to a prediction problem. Given researchers' imperfect knowledge and data, the goal is to produce the best prediction of normal accruals possible. This prediction, however, is naturally uncertain, and the uncertainty should be considered in classifying firms' accruals as abnormal. In the presence of prediction uncertainty, not every deviation from predicted normal accruals can confidently be classified as abnormal.

## 2.1.1 Standard approach

The literature's standard approach to predicting normal accruals proceeds in three steps. Researchers first pick a model specifying the relevant determinants (e.g.,

<sup>&</sup>lt;sup>6</sup>In the online appendix, we document the robustness of our inferences to alternative prior choices by varying the informativeness and the functional forms of our priors.

change in revenues) and functional forms (e.g., linear model) of normal accruals, typically the modified Jones model:

$$TA_{i,t} = \beta_{0,g} InvA_{t_{i,t-1}} + \beta_{1,g} \triangle CRev_{i,t} + \beta_{2,g} PPE_{i,t} + e_{i,t}$$
(1)

where  $InvAt_{i,t-1}$  is the inverse of lagged total assets,  $\triangle CRev_{i,t}$  is the change in cash revenues scaled by lagged total assets, and  $PPE_{i,t}$  is property, plant, and equipment scaled by lagged total assets.<sup>7</sup> (For detailed variable definitions, see Table 1.)

The model's parameters (i.e., coefficients) are treated as unknown but estimable quantities. In the second step, these parameters are estimated via ordinary least squares (OLS). Typically, the parameters ( $\Theta_g$ ) are estimated at the industry-year level (g), for industry-years with a minimum number of observations, to allow for heterogeneity in the accruals process across industries (j = 1, ..., J) and time (t = 1, ..., T):

$$\hat{\Theta}_{g} = \{\hat{\beta}_{0,g}, \hat{\beta}_{1,g}, \hat{\beta}_{2,g}\} = \underset{\Theta_{g}}{\operatorname{argmin}} \Big[ (TA_{i,t} - \hat{T}A_{i,t})^{2} \Big] \forall g = 1, ..., JT$$
(2)

In the last step, these parameter estimates are combined with firms' observed determinants in the way described by the chosen model to generate a prediction of normal accruals for a given firm at a given point in time:

$$\hat{TA}_{i,t}^{OLS} = \hat{\beta}_{0,g} InvAt_{i,t-1} + \hat{\beta}_{1,g} \triangle CRev_{i,t} + \hat{\beta}_{2,g} PPE_{i,t}$$
(3)

These point predictions are expected accruals, *conditional* on firms' observed determinants  $(D_g)$ , the estimated parameters  $(\hat{\Theta}_g)$ , and the chosen model (M):

$$\hat{TA}^{OLS}|D_g,\hat{\Theta}_g,M \tag{4}$$

These predictions do not account for the fact that the model parameters were estimated (parameter uncertainty) and that researchers are uncertain about the true model (model uncertainty). Nevertheless, the literature typically neglects prediction uncertainty (model and parameter uncertainty), takes the predicted accruals of the chosen model as the true accruals, and classifies any deviations of firms' observed accruals from the predictions as abnormal. As a result, this classification rule can be expected to lead to excessive evidence of managerial discretion and opportunism.

#### 2.1.2 Bayesian approach

Our proposed Bayesian approach to predicting normal accruals attempts to explicitly incorporate model and parameter uncertainty to reduce the incidence of false positives.

**Bayes** Bayesian statistics provide a convenient way to incorporate uncertainty about firms' accruals processes in predicting normal accruals. Researchers' beliefs about

<sup>&</sup>lt;sup>7</sup>In line with the original Jones (1991) model, we include the inverse of total assets, instead of a constant. In latter sections, we also consider models with (group-specific) constants, instead of (group-specific) coefficients on the inverse of total assets.

unknown quantities (e.g., models, parameters, or normal accruals) can be represented by distributions. As researchers examine data (on firms' observed accruals and their determinants), their initial beliefs change following Bayes's rule. For example, researchers' beliefs about the coefficient values of the modified Jones model after examining the data are a weighted average of researchers' beliefs before examining the data (prior distribution) and the data (likelihood), where the relative weights depend on the precision of the prior belief vis-à-vis the amount and quality of the data:



In the following three steps, we illustrate how Bayesian methods and tools can be applied to the estimation of the modified Jones model to better predict normal accruals and explicitly account for multiple sources of uncertainty.

**Parameter uncertainty** In a first step, we show how uncertainty surrounding the model's parameters (i.e., coefficients) is incorporated into the normal accruals prediction. For this purpose, we take the modified Jones model as given and simply estimate its parameters using Bayesian, instead of OLS regressions. For the Bayesian regression estimation, we must specify distributions describing accruals and our priors about reasonable coefficient values ( $p(\Theta_g|M)$ ).

The Bayesian estimation yields posterior distributions for the coefficient values  $(p(\Theta_g|D_g, M))$ , informed by the data  $(D_g)$ . The wider these distributions, the less precise are the coefficient values. The coefficient distributions serve as a key input into the prediction of normal accruals. Just as in the standard approach, we predict firms' normal accruals by combining the firms' observed determinants with the respective coefficient values. In contrast to the standard approach, however, we do not simply use one point estimate per coefficient but instead calculate predicted accruals by weighting all plausible coefficient values with the respective probabilities assigned by the posterior coefficient distribution  $(p(\Theta_g|D_g, M))$ . As a result, we obtain a full distribution of predicted normal accruals for a given firm at a given time—the so-called posterior predictive distribution  $(p(TA|D_g, M))$ :

$$p(\hat{TA}|D_g, M) \propto \int p(\hat{TA}|D_g, \Theta_g, M) p(\Theta_g|D_g, M) d\Theta_g$$
(6)

This predictive distribution incorporates uncertainty about the true parameters of the model into the prediction of normal accruals. Unlike the OLS point prediction, the predictive distribution does *not* take the coefficient estimates as given.

**Heterogeneity** In a next step, we illustrate how one can efficiently incorporate heterogeneity in parameters across groups into the normal accruals prediction. For this purpose, we again take the modified Jones model as given and estimate its parameters using Bayesian regressions. In addition, we now explicitly model the heterogeneity in parameters across groups using Bayesian hierarchical modelling, instead of merely estimating our coefficients within groups with a given minimum number of observations.

The hierarchical model accounts for the fact that, in a hierarchical setting, the data really contains two sources of information: the within-group data and across-group variation. It splits the model parameters into group-specific parameters (e.g., industry-year-specific coefficients) and hyper-parameters ( $\psi$ ) governing the distribution of the group-specific parameters across groups:

$$p(\hat{TA}|D,M) \propto \int \prod_{g=1}^{JT} \left( \int p(\hat{TA}|D_g,\Theta_g,\psi,M) p(\Theta_g|D_g,\psi,M) d\Theta_g \right) p(\psi|D,M) d\psi$$
<sup>(7)</sup>

The hierarchical model uses the entire sample (D instead of  $D_g$ ) to estimate all the parameters jointly. If there are plenty of observations in a given industry-year, the group-specific coefficient is chiefly determined by these observations. If instead there are only limited observations in a given industry-year, the group-specific coefficient borrows information from the other industry-year coefficients. (This information is contained in the across-group hyper-parameters.) This approach eliminates the need to define ad hoc cutoffs of minimum required observations and discard useful information/data (e.g., cross-group information and information in groups with fewer observations than defined by the cutoff).

**Model uncertainty** In a last step, we illustrate how, using Bayesian model averaging, one can combine different models of normal accruals to achieve better predictions and incorporate model uncertainty into the predictive distribution. Just as with parameters, we can have a prior belief about the plausible models and update this belief based on the data. The resulting posterior distribution of plausible models can be used to weight the predictions of distinct models. These model probabilities can be used to average the model-specific predictive distributions of normal accruals and obtain an averaged predictive distribution of normal accruals:<sup>8</sup>

$$p(\hat{TA}|D) \propto \int \left( \int p(\hat{TA}|D,\Theta,M) p(\Theta|D,M) d\Theta \right) p(M|D) dM$$
(8)

This averaged distribution uses information from all candidate models with nonzero probabilities. Importantly, it incorporates researchers' uncertainty about the true model among the set of candidate models. Unlike the OLS point prediction

<sup>&</sup>lt;sup>8</sup>For notational brevity, we us  $p(\hat{T}A|D, M) \propto \int p(\hat{T}A|D, \Theta, M)p(\Theta|D, M)d\Theta$  as a short form of the hierarchical model where  $\psi$  is already integrated out.

 $(\hat{TA}|D_g, \Theta_g, M)$ , the predictive distribution  $p(\hat{TA}|D)$  does *not* take one ad hoc model for granted anymore.<sup>9</sup>

#### 2.2 Measurement problem

We can use the model-averaged predictive distribution of normal accruals in various ways to classify firms' accruals as abnormal. For example, we can calculate a dichotomous variable indicating when firms' observed accruals fall outside of, for example, the 98% credible interval of the predictive distribution. We can also calculate a continuous variable using scoring rules (i.e., loss functions). The logarithmic scoring rule, for example, closely relates to information theory. It translates the predictive probability assigned to firms' observed accruals into a measure of loss or cost to a decision-maker who relies on the prediction. The negative of the logarithmic scoring function hence can be viewed as a measure of researchers' confidence that observed accruals are indeed abnormal.<sup>10</sup> Alternatively, and similar to the standard approach, we can take the difference between the observed accruals and the mean of the posterior predictive distribution as a point prediction of abnormal accruals but then apply an adjustment for the corresponding prediction uncertainty.

#### 2.3 Testing problem

We can use moments of the predictive distribution of normal accruals to test for abnormal accruals, due to opportunistic accruals in a Bayesian measurement-error model. In the measurement-error model, we can, for example, not only use the mean but also the standard deviation of the predictive distribution to infer opportunistic accruals. In particular, we can consider abnormal accruals as a latent construct:

$$DA_{i\,t}^{*} = TA_{i,t} - NDA_{i\,t}^{*} \tag{9}$$

which can only be measured with error, because we do not observe the normal (or nondiscretionary) accruals  $(NDA_{i,t}^*)$ .

To measure abnormal accruals  $(DA_{i,t})$ , we take the difference between firms' observed accruals  $(TA_{i,t})$  and the expected value of the predictive distribution of

<sup>&</sup>lt;sup>9</sup>Bayesian model averaging relates to two ad hoc approaches to addressing model uncertainty common in the earnings management literature: showing robustness across various models (e.g., Sletten et al. 2018) and using a combined score (e.g., Leuz et al. 2003). Unlike these approaches, Bayesian model averaging explicitly accounts for differences in the predictive ability of the distinct models via the model weights. This theoretically motivated and data-driven weighting increases power to detect earnings management, relative to ad hoc judgments about which models to report and how to combine earnings management indicators.

<sup>&</sup>lt;sup>10</sup>The logarithmic scoring rule uses (the negative of) the logarithm of the posterior probability assigned by the model to the observed accruals value. This rule can be linked to information theory and is commonly used in Bayesian inference. Absent an explicit economic model, it, however, is just one among many possible rules (i.e., it is a purely statistical construct).

normal accruals  $(NDA_{i,t} = TA_{i,t}|D)$ , closely following the literature's approach:

$$DA_{i,t} = (TA_{i,t} - NDA_{i,t}) = DA_{i,t}^* + (NDA_{i,t}^* - NDA_{i,t})$$
  
~  $N(DA_{i,t}^*, \sigma_{NDA,i,t})$  (10)

As our prediction of normal accruals is uncertain, so is our abnormal accruals measure. To reflect the uncertainty, we can model our abnormal accruals measure as normally distributed with the latent abnormal accruals value  $(DA_{i,t}^*)$  as its mean and the extent of measurement error  $(\sigma_{NDA,i,t})$  as its standard deviation.<sup>11</sup> We proxy for the extent of error in our measurement by using the standard deviation of the predictive distribution of normal accruals  $(p(TA_{i,t}|D))$ .

Equipped with our measure of abnormal accruals and its extent of measurement error, we can estimate the Bayesian measurement-error model regressing measured abnormal accruals on variables for managerial opportunism  $(X_{i,t})$ .<sup>12</sup>

$$DA_{i,t}^{*} = a_{0} + a_{1}X_{i,t} + u_{i,t}$$

$$DA_{i,t} \sim N \left( DA_{i,t}^{*}, \sigma_{NDA,i,t} \right)$$
(11)

In estimating the impact of managerial opportunism, this model takes into account that some abnormal accruals are less precisely measured than others (i.e., more uncertain). The model weights observations according to the prediction uncertainty for  $DA_{i,t}$ . This precision-based weighting of abnormal accruals will shrink the influence of volatile firms and firms with limited information. Notably, these firms typically contribute to the many (potentially false) positives in tests for managerial opportunism (e.g., Hribar and Collins 2002; Kothari et al. 2005).<sup>13</sup>

## 3 Data

We closely follow the literature in constructing the sample and variables underlying our tests. We use Compustat North America data from 1988 until 2017. For the estimation of the Jones model and its basic variants (Section 4), we exclude financial firms (three-digit SIC code between 600 and 699) and observations with lagged total assets smaller than \$10 million, leverage (debt over total assets) larger than one, or total asset growth of 200% or more (to exclude M&A activity). We restrict the

<sup>&</sup>lt;sup>11</sup>To implement the measurement-error model, we assume that the unobservable measurement error  $(NDA_{i,t}^* - NDA_{i,t})$  follows a normal distribution. This assumption is a simplifying assumption as the actual distribution is a mixture of unknown form. It is motivated by two observations. First, we observe that the predictive distributions of abnormal accruals appears to approximate normal distributions. Second, we note that the normal distribution is the maximum uncertainty distribution if all we know is an expected value and a standard deviation. Hence our simplifying assumption follows the maximum uncertainty principle that governs our choice of priors (Section 4).

<sup>&</sup>lt;sup>12</sup>Figure A4 in the online appendix illustrates the measurement-error regression framework.

<sup>&</sup>lt;sup>13</sup>For the above two-step approach, it is important to include the determinants used in the normal-accruals estimation (the first step) into the earnings-management regression (the second step) to avoid omitted variable bias, if the incentive variables are correlated with the determinants of normal accruals (Chen et al. 2018). Alternatively, one can use a one-step estimation, regressing total accruals on both their normal determinants and the incentive variables.

sample to observations with nonmissing values for all of required variables. We truncate all continuous variables at the first and 99th percentile. Table 1 summarizes the data. Panel A provides the definition of our variables (following Dechow et al. 1995). Panel B presents descriptive statistics for the variables.

For our application of the Bayesian approach to an extended set of accruals models (Section 5), we additionally require nonmissing values for all the variables specified by the various models (listed in Table 6). These additional requirements (e.g., availability of book-to-market information) further restrict the sample size. As before, we truncate all continuous variables at the first and 99th percentile.

For validation tests (Section 5), we use earnings management indicators provided by Dechow et al. (2011) and Audit Analytics. In our first earnings management test, we employ an indicator ( $I_{AAER}$ ) taking the value of one for firm-years subject to

Table 1 Sample descriptives and variable description

#### Panel A:

Variable Description

$\triangle CRev_{i,t}$	Change in	cash revenu	ues ((sale <sub>i</sub>	$_{,t} - sale_{i,t-}$	$-1 - (rect_{i,t})$	$-rect_{i,t-1}$	$))/at_{i,t-1})$	
$InvAt_{i,t-1}$	Inverse of	lagged tota	l assets (1,	$(at_{i,t-1})$				
$Lev_{i,t-1}$	Financial	leverage ((d	$lltt_{i,t-1} +$	$dlc_{i,t-1})/at$	$t_{i,t-1})$			
$PPE_{i,t}$	Property,	plant, and ed	quipment (	ppegt <sub>i,t</sub> /at	(i,t-1)			
$RoA_{i,t}$	Return on	assets (oid	$adp_{i,t}/at_{i,t}$	$_{t-1})$				
$SdRev_{i,t}$	Revenue	volatility ( $\sigma$	(sale <sub>i,t</sub> , sa	$ale_{i,t-1}, sale$	$e_{i,t-2})/at_{i,t}$	_1)		
$TA_{i,t}$	Total accr	uals (( <i>ib<sub>i,t</sub> -</i>	$-oancf_{i,t}$	$)/at_{i,t-1})$				
Panel B:								
Variable	Ν	Mean	StDev	P05	P25	Median	P75	P95
$\triangle CRev$	139,241	0.075	0.240	-0.237	-0.019	0.039	0.147	0.480
InvAt	139,241	0.013	0.020	0.000	0.001	0.004	0.015	0.059
Lev	139,241	0.242	0.215	0.000	0.037	0.216	0.375	0.652
PPE	139,241	0.634	0.472	0.070	0.253	0.532	0.942	1.477
RoA	139,241	0.035	0.188	-0.315	-0.007	0.066	0.123	0.251
SdRev	139,241	0.163	0.185	0.007	0.045	0.104	0.213	0.519
TA	139,241	-0.070	0.120	-0.264	-0.108	-0.056	-0.018	0.080

The table presents information about the sample used to fit the six accruals models used in the analysis. The sample is based on the Compustat North America File. Panel A shows the computations for all variables used. We generally use the variable definitions by Dechow et al. (1995). Variables inside the brackets refer to the corresponding Compustat column names: *sale* is revenues; *rect* is trade receivables; *at* is total assets; *dltt* is long-term debt; *dlc* is current debt; *ppegt* is gross property, plant, and equipment; *oiadp* is operating income after depreciation; *ib* is income before extraordinary items; and *oancf* is cash from operating activities. The seven final variables are scaled to mean zero and standard deviation one before fitting the models. Panel B presents descriptive statistics for the unscaled final sample. We use all data from 1988 till 2017 that fulfill the following requirements: we exclude financial firms (three-digit sic code between 600 and 699) and observations with lagged total assets smaller \$10 million. To be included an observation must further have nonmissing values for all seven variables. Finally, we truncate all continuous variables at the first and 99th percentiles of their respective distributions and keep only firms with leverage less than 1 and total asset growth of less than 200% (to exclude extreme M&A activity)

upward earnings management, according to an Accounting and Auditing Enforcement Release (AAER) of the SEC (Dechow et al. 2011) (and zero otherwise). In our second test, we use an indicator ( $I_{REST}$ ) taking the value of one for firm-years with a restatement due to accounting rule (GAAP/FASB) application failure, excluding disclosure issues (and zero otherwise) (Du et al. 2020). In our third test, we use an indicator ( $I_{CL}$ ) taking the value of one for firm-years with an SEC comment letter pertaining to a 10-K or 10-Q in the particular year (and zero otherwise) (Du et al. 2020).<sup>14</sup>

## **4** Estimation

## 4.1 Predicting normal accruals

#### 4.1.1 Bayesian approach

**Specification** The Bayesian version of the industry-year-level modified Jones model (with  $g = \{j, t\}$ ) takes the following form:

$$TA_{i,j,t} \sim N(\mu_{i,j,t}, \Sigma) \mu_{i,j,t} = \beta_{0,j,t} InvA_{t_{i,j,t-1}} + \beta_{1,j,t} \triangle CRev_{i,j,t} + \beta_{2,j,t} PPE_{i,j,t}$$
(12)

Similar to the OLS model in Eq. 1, we model total accruals (TA) as a linear function of (inverse) total assets, changes in revenues, and PP&E, plus some noise. For the Bayesian estimation, we need to specify the distribution of the noise and priors over the model parameters. Following standard Bayesian practice for linear regressions, we model the noise of the accruals model as normally distributed, with mean zero and unknown standard deviation  $\Sigma$ . In specifying the priors, we follow guiding principles advocated by leading Bayesian statisticians (e.g., McElreath 2020; van de Schoot et al. 2021). We choose priors that satisfy three criteria. First, they should down-weight unreasonable parameter ranges. Second, they should make minimal assumptions about the prior shape. Third, they should be easy to reason about.<sup>15</sup>

Following the first criterion, we choose weakly informative priors for the unknown standard deviation  $\Sigma$  and the industry-year-level coefficients,  $\beta_{0,j,t}$ ,  $\beta_{1,j,t}$ ,  $\beta_{2,j,t}$ . In contrast to diffuse priors, ours down-weight unreasonable parameter values, reducing overfitting concerns. Following the second criterion, we choose an exponential distribution for the shape of the unknown standard deviation and normal distributions for the shape of the coefficient priors. Following the maximum entropy argument (e.g., Jaynes 2003; McElreath 2020), these distributions encode maximum uncertainty about the shapes of the priors. The exponential distribution, for example, is the

<sup>&</sup>lt;sup>14</sup>We follow Du et al. (2020) in requiring that the filing date of the form (10-K or 10-Q) commented on by the SEC to fall within a 365-calendar-day interval ending 100 days after the fiscal year-end.

<sup>&</sup>lt;sup>15</sup>A popular alternative approach to choosing priors is to consider conjugate priors that, in combination with the likelihood, result in known posterior distributions. The benefit of conjugacy is that it reduces the computational burden. Its drawback is that it may impose undesirable restrictions on the shape of the priors.

maximum uncertainty distribution if all we are willing to specify is the "average deviation." Similarly, the normal distribution is the maximum uncertainty distribution if all we are willing to specify are the mean and standard deviation. Lastly, following the third criterion, we standardize total accruals and all determinants to have means of zero and standard deviations of one. This standardization (and the choice of lowparameter distributions) aids the accessibility of the model estimation by simplifying the reasoning about the priors.

In sum, we specify our priors as follows:

$$\Sigma \sim Exp(1)$$
  

$$\beta_{0,j,t} \sim N(0, 2.5) \; \forall j, t$$
  

$$\beta_{1,j,t} \sim N(0, 2.5) \; \forall j, t$$
  

$$\beta_{2,j,t} \sim N(0, 2.5) \; \forall j, t$$
(13)

The prior for  $\Sigma$  implies that, before seeing the data, we expect the average deviation of the residual to be one. As we standardize our dependent variable, this parameterization allows the residuals to explain most if not all the variation in the dependent variable. Accordingly, it is a conservative but still weakly informative prior (which respects the nonnegativity constraint for values of the unknown standard deviation). Similarly, the prior choices for the coefficients imply that, before seeing the data, we expect that the true (standardized) coefficients lie within the range of -5 to +5 with a 95% probability. This choice essentially says that we view it as unlikely that a one-standard-deviation increase in InvAt,  $\triangle CRev$ , or PPE is associated with a change in accruals of more than five times its standard deviation. These priors are weakly informative because only economically implausible and statistically rare coefficients of magnitudes outside the -5 to +5 range are down-weighted.

We specify and estimate this model using STAN software (Carpenter et al. 2017). The benefit of this software is that it can be called from standard statistical software packages (e.g., R and STATA) and employs various forms of Markov Chain Monte Carlo (MCMC) algorithms in the background to draw the posterior distributions.<sup>16</sup> Accordingly, our Bayesian approach merely requires a conceptual understanding of Bayesian statistics and concepts. It does not require knowledge of new programming languages or expertise in sampling algorithms.

For illustrative purposes, we briefly introduce the conceptual idea behind MCMC sampling algorithms, provide details on necessary estimation choices, and discuss post-estimation diagnostics in the following. In our Bayesian estimation, MCMC algorithms are needed because the posterior parameter distributions of our complex models do not correspond to closed-form expressions of known distributions. In the absence of such distributions, the algorithms allow to numerically approximate the posterior distributions by generating a large sample of parameter values. To obtain this sample, MCMC algorithms perform two operations to ensure that the parameter values are sampled in a way that approximates the underlying posterior of interest.

<sup>&</sup>lt;sup>16</sup>We provide our code publicly to ease the adoption of our approach.

First, starting from current parameter values ( $\Theta^s$ ), they propose a new set of parameter values ( $\Theta^*$ ) by stepping randomly in a new direction in parameter space. Second, they decide whether to accept the proposed step. The algorithms accept the proposed values ( $\Theta^*$ ) into the series ({ $\Theta^{(1)}, ..., \Theta^{(s)}$ }) if the ratio of the probabilities of the proposed vis-à-vis the current parameter values  $(p(y|\Theta^*)p(\Theta^*)/p(y|\Theta^{(s)})p(\Theta^{(s)}))$  is higher than some random amount between 0 and 1. Otherwise, they add the current value ( $\Theta^s$ ) to the series again. As the series or chain grows, the resulting distribution of parameter values converges to the desired parameter distribution. (For a detailed introduction to sampling algorithms, see Gelman et al. (2013).)

MCMC algorithms differ in how they implement the generation of proposal values and the acceptance decision. In our estimation, we use the Hamiltonian MCMC algorithm, the default algorithm in STAN. The algorithm uses Hamiltonian motion dynamics to determine the next proposed value. For complex models, it tends to be more efficient than previous algorithms (e.g., Gibbs or Metropolis-Hastings) as the proposed values tend to have a higher acceptance rate (Betancourt 2018). To parameterize the algorithm (e.g., step size of proposal values), we rely on the default settings in STAN. The STAN implementation auto-tunes the algorithm parameterization to achieve appropriate mixing (e.g., a high acceptance rate of proposal values) via machine learning tools (e.g., using trees to determine the next move).

We estimate all models using two chains with 3,500 iterations each. We discarded the first 1,000 iterations of each chain as our burn-in period. Using multiple chains and discarding the initial values are best-practices, which reduce the dependence of a given series on the starting value.

After the estimation, we examine the chains for convergence and efficiency (i.e., their mixing properties). The chains should not be stuck at any particular value for too long and move through all possible values several times. We visually inspect the chains of the key model parameters using trace plots, which plot the series of accepted values across the iterations. In addition, we examine the *R*-hat convergence diagnostic and effective sample size (Gelman and Rubin 1992). We make sure that our choices of the burn-in period and number of iterations result in *R*-hat values of less than 1.05, which indicates sufficient mixing and ensures an effective sample size of at least 100 samples per chain (Stan Development Team 2018).<sup>17</sup>

**Results** The model estimation produces posterior distributions of the model parameters. Equipped with these distributions, we can generate the posterior predictive distribution of accruals. Figure 1 showcases an example of a predictive distribution for a given firm in a given year. The peak of the distribution represents the most likely accruals value, according to the model. The width of the distribution reflects parameter uncertainty. The figure highlights the importance of accounting for uncertainty in the accruals prediction. The OLS estimate (marked by a thin black line)

<sup>&</sup>lt;sup>17</sup>After checking post-estimation diagnostics (e.g., examining mixing properties), we could leave almost all default settings in place. We only had to increase the maximum tree-depth for the Ball and Shivakumar (2006) and the Frankel and Sun (2018) models estimated in Section 5.



**Fig. 1** Example of Posterior Predictive Distribution. The figure illustrates the posterior predictive distribution. The light-gray area shows the posterior predictive distribution for one illustrative firm-year observation from a Bayesian hierarchical version of the modified Jones model of the form  $TA_{i,t} = a_{1,j,t} InvAt_{i,t} + a_{2,j,t} \triangle CRev_{i,t} + a_{3,j,t} PPE_{i,t}$ . The subscripts j, t denote industry-year-specific coefficients for the three variables. Variable descriptions can be found in Table 1. The dark-gray areas mark accruals values that lie beyond the first and 99th percentile and are thus very unlikely according to the model. More generally, the width of the posterior reflects the model's uncertainty regarding which value of nondiscretionary accruals to expect

and theobserved accruals level (marked by a dashed line), while differing from each otherby around 4% of the standard deviation of accruals, are close to the center of the predictive distribution. Accordingly, neither the observed accruals level nor the difference between the OLS estimate and the observed accruals level appears unusual nor significant, given the extent of prediction uncertainty.

#### 4.1.2 Heterogeneity

**Specification** We augment the industry-year-level model in Eq. 12 by introducing a hierarchical parameter structure featuring hyper-parameters governing the relation of industry-year-level coefficients across industry-years.<sup>18</sup> To this end, we first modify the priors for the coefficients in Eq. 13 to be:

$$\begin{pmatrix} \beta_{0,j,t} \\ \beta_{1,j,t} \\ \beta_{2,j,t} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_0 \\ \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_0 & \rho_{0,1} & \rho_{0,2} \\ \rho_{0,1} & \sigma_1 & \rho_{1,2} \\ \rho_{0,2} & \rho_{1,2} & \sigma_2 \end{pmatrix}\right) \quad \forall j,t$$
(14)

<sup>&</sup>lt;sup>18</sup>In line with the literature, we allow for coefficient heterogeneity. The Bayesian hierarchical modelling approach could also be used to allow heterogeneity in the residual variance ( $\Sigma$ ).

The coefficient priors are now functions of new hyper-parameters ( $\mu$ ,  $\rho$ ,  $\sigma$ ). The hyper-parameters determine the plausible distribution of coefficients across industryyear groups.  $\mu$  reflects the expected average coefficient across industry-years,  $\sigma$  determines the range of plausible coefficients across industry-years, and  $\rho$  allows for cross-coefficient learning. Importantly, and unlike before, the hyper-parameters  $\mu$  and  $\sigma$  are not fixed anymore. Their posterior distributions are estimated from the data, jointly with the industry-year-specific coefficients. Just as the coefficients, these hyper-parameters need (hyper-)priors. We specify these priors as follows:

$$\mu_{d} \sim N(0, 2.5) \quad \forall d$$
  

$$\sigma_{d} \sim Exp(1) \quad \forall d$$
  

$$\rho \sim LKJcorr(2) \tag{15}$$

The parameters of the hyper-priors determine the plausible values of the hyperparameters, before seeing the data.  $\mu_d$  is the prior for the average across all group-specific coefficients  $\beta_{d,g}$  for determinant d (with d = 0, 1, 2 in the modified Jones model).  $\sigma_d$  is the prior for the amount of variation in  $\beta_{d,g}$  across groups. We chose relatively weak priors to have nearly all the shrinkage (i.e., the disciplining of group specific coefficient estimates) come from the data.<sup>19</sup> The prior for the correlation between coefficients  $\rho$  is a special multivariate prior for correlations that honors the dependencies between correlations (Lewandowski et al. 2009).<sup>20</sup> *LK J corr*(2) defines a weakly informative prior that is skeptical of extreme correlations.

We estimate this hierarchical model using the entire dataset. The hierarchical modelling eliminates the need to estimate parameters separately for industry-years and facilitates information sharing across industry-year groups. (For further detail on the prior definition and choice of the Bayesian hierarchical model, please refer to the online appendix.)

**Results** The model estimation produces posterior distributions of the model parameters. Figure 2 illustrates that, compared to the corresponding OLS estimates, the (average) coefficient estimates obtained using the hierarchical model are less dispersed. The hierarchical modelling, in particular, shrinks extreme coefficients observed for industry-year groups with little information (i.e., few observations) toward the typical coefficient values observed across industry-year groups. This shrinkage is based on the amount of available information, and cross-group

<sup>&</sup>lt;sup>19</sup>"Shrinkage" in Bayesian estimation refers to the impact of (weakly) informative priors on posterior coefficient estimates. Priors pull (or shrink) the coefficient estimates obtained from the data toward the priors. This pull is a form of regularization, which addresses issues with outliers in the data (see Leone, Minutti-Meza, and Wasley (2019) for a discussion of outliers in accounting research).

<sup>&</sup>lt;sup>20</sup>Correlations cannot be independent. For example, if the correlation between *PPE* and *InvAt* is large, then this influences the correlations between *PPE* and  $\triangle CRev$  and between *InvAt* and  $\triangle CRev$ . The LKJ prior accounts for this dependence (Lewandowski et al. 2009). Its use for multi-level models is advocated in most software packages. See, for example, the STAN manual (Stan Development Team 2018).



**Fig. 2** Comparison of Industry-Year Coefficient Estimates. The figure plots industry-year coefficient estimates  $(b_{1,j,t}, b_{2,j,t}, b_{3,j,t})$  of the modified Jones model by industry-year sample size. The modified Jones model used is of the form  $TA_{i,t} = b_{1,i,t}InvAt_{i,t} + b_{2,j,t} \triangle CRev_{i,t} + b_{3,j,t}PPE_{i,t}$ . Variable descriptions can be found in Table 1. The three plots on the left of the figure show OLS estimates, whereas the plots on the right show posterior means for each coefficient from a Bayesian hierarchical model. The grey line highlights the requirement of a minimum of 20 observations in an industry-year that is common in the literature. Comparing the OLS (left) and Bayes (right) plots for each coefficient differences are apparent for industry-years with large sample sizes (e.g., greater than 200 obs). By contrast, the dispersion of OLS estimates is much larger for small industry-year samples (e.g. less than 100 observations)

learning eliminates the need to rely on ad hoc cutoffs (e.g., for the number of required observations), opens the possibility of estimating even firm-level models, and can be expected to lead to more plausible coefficient estimates and accruals predictions.

#### 4.1.3 Model uncertainty

**Specification** We incorporate model uncertainty in our estimation and prediction following Yao et al. (2018). They propose a Bayesian model-averaging approach for the realistic case where researchers can specify a set of (*K*) candidate models but do not expect any of those models to be the one true model. Their approach combines the posterior predictive distributions of the various candidate models via model weights,  $\hat{w}_k$ . These weights, which are akin to the posterior probabilities of the various models, provide a discrete distribution over the set of candidate models. They are derived from the models' out-of-sample prediction performance, using approximate leave-one-out-prediction accuracy. For each model, we multiply its weight with its predictive distribution to obtain a weighted-averaged posterior predictive distribution that fits the data best (in the sense of the Kullback–Leibler divergence scoring rule):<sup>21</sup>

$$p(\hat{TA}|D) = \sum_{k=1}^{K} \hat{w}_k p(\hat{TA}|D, M_k)$$
(16)

In our implementation, we use six candidate models (K = 6). The first is the hierarchical modified Jones model with industry-year-specific coefficients. The second includes three additional determinants capturing firms' operating and financing characteristics (e.g., in the spirit of Kothari et al. (2005) and Hribar and Nichols (2007)): return on assets (*RoA*), leverage (*Lev*), and revenue volatility (*SdRev*).

$$TA_{i,j,t} \sim N(\mu_{i,j,t}, \Sigma) \mu_{i,j,t} = \beta_{0,g} InvA_{t_{i,j,t-1}} + \beta_{1,g} \triangle CRev_{i,j,t} + \beta_{2,g} PPE_{i,j,t} + \beta_{3,g} RoA_{i,j,t} + \beta_{4,g} Lev_{i,j,t-1} + \beta_{5,g} SdRev_{i,j,t}$$
(17)

The third model resembles the extended (second) model but includes an industryyear-specific intercept, instead of an industry-year-specific coefficient for the inverse of total assets. Models four to six are versions of the first three models with coefficients varying at the firm, instead of the industry-year, level.

For illustrative purposes, we focus on a narrow set of candidate models, representing slight deviations from the modified Jones model. Our implemented approach accordingly only captures uncertainty about the models within the set of candidate models. It neglects any uncertainty about and information in models excluded from the candidate set. To incorporate model uncertainty more fully and obtain the best normal-accruals prediction, we apply our Bayesian approach to an extended

<sup>&</sup>lt;sup>21</sup>The approach combines models based on their out-of-sample fit, reducing concerns about over-fitting (e.g., selecting more complex models due to better in-sample fit as a result of more parameters). For a detailed discussion and implementation of the Bayesian model-averaging approach, see Yao et al. (2018).

set of models, including some of the most popular and recent ones in a final step (Section 5).<sup>22</sup>

**Results** We fit all six candidate models to the data and compute their weights. Panel A of Fig. 3 shows the weights for each. The extended firm-level model (model 5) receives the highest weight, followed by the extended firm-level model with firm-level intercepts (instead of inverse total asset coefficients) (model 6). By contrast, the industry-year-level models receive only little weight. Only the extended industry-year model with industry-year intercepts (model 3) receives a nonnegligible weight.

The weights make three important points. First, the dominance of the firm-level models highlights that firm-level heterogeneity in accruals processes is more important in explaining the data than industry-year-level heterogeneity (which is consistent with Owens et al. (2017)). Second, the dominance of the extended models highlights the importance of controlling for performance and related determinants (which is consistent with Kothari et al. (2005)). Third, the nonnegligible weight on the industry-year intercept model suggests that time trends in accruals (not necessarily their processes (i.e., coefficients)) appear important to account for (e.g., via industry-year fixed effects).

Panel B of Fig. 3 illustrates, for a given company in a given year, how the modelspecific predictive distributions combine to generate one model-averaged posterior predictive distribution of normal accruals. The left graph depicts the model-specific distributions. The predictive distributions of the industry-year-level models are centered around lower values than the firm-level distributions and exhibit more narrow ranges. The narrower range reflects that the industry-year-level coefficients are estimated with greater precision than the firm-level coefficients. The model weights, however, suggest that this greater precision comes at the cost of significant bias, compared to the firm-level models. As a result, the model-averaged distribution, in the right graph, more closely resembles the firm-level predictive distribution in that it exhibits a relatively wide range and is centered more to the right (compared to the industry-year-level distributions). Notably, the model-averaged distribution is less wide than the firm-level distributions, because it uses information and features from several models. Accordingly, the model-averaged predictive distribution is not only less biased than the individual distributions but also more precise than the widest individual (firm-level) distribution.

 $<sup>^{22}</sup>$ In determining the relevant candidate models, including their determinants and level of heterogeneity, we advocate the use of theory, institutional knowledge, and prior literature. For proper inferences, we suggest to account for first-order determinants of accruals, especially if these determinants are correlated with proxies for opportunistic reporting incentives (e.g., Belloni et al. 2013). As a result of the bias-variance trade-off, we note that this approach may result in a lack of power to detect earnings management. This lack of power (or independent treatment variation), however, does not justify the omission of important correlated factors. Instead, it suggests that, given the data and the existence of confounding influences, we cannot confidently infer whether earnings are managed opportunistically.



**Fig. 3** Illustration of Stacking Weights. Panel A shows stacking weights for six models: a hierarchical modified Jones model with industry-year-specific coefficients (Model 1); a hierarchical modified Jones model with industry-year-specific coefficients and three additional determinants, RoA, Lev, and SdRev (Model 2); a version of model 2 with an industry-year specific intercept instead of the inverse of total assets covariate (Model 3); a hierarchical modified Jones model with firm-specific coefficients (Model 4); a hierarchical modified Jones model with firm-specific coefficients (Model 4); a hierarchical modified Jones model with firm-specific coefficients (Model 4); a hierarchical modified Jones model with firm-specific coefficients (Model 4); a hierarchical modified Jones model with firm-specific intercept, instead of the inverse of total assets covariate (Model 5); a version of model 5 with a firm-specific intercept, instead of the inverse of total assets covariate (Model 6). The stacking weights are based on approximations of leave-one-out cross-validation performance of the models. Models with better out-of-sample performance receive more weight. The left plot of panel B presents the posterior predictive distributions (PPDs) of six models for an exemplary firm-year observation. The right part shows the combined, weighted posterior predictive distribution of all six models. The weights used are the stacking weights shown in panel A

## 4.2 Measuring abnormal accruals

**Specification** Our Bayesian approach allows for various ways of defining and measuring abnormal accruals (see Section 2.2). In this section, we compare point

Variable	Ν	Mean	StDev	Min	P05	P25	Median	P75	P95	Max
TA	139,241	0.00	1.00	-37.34	-1.62	-0.32	0.11	0.43	1.25	78.30
PredOLS	138,065	0.01	0.56	-21.65	-0.84	-0.19	0.07	0.26	0.71	47.82
PostMean	139,241	0.00	0.66	-32.02	-0.95	-0.18	0.08	0.27	0.71	70.40
PostSD	139,241	0.17	0.06	0.05	0.09	0.12	0.15	0.20	0.30	0.42

 Table 2
 Descriptive statistics of accruals measures

The table presents descriptive information on total accruals, OLS-based accruals predictions, and Bayesian accruals predictions (mean and standard deviation). TA is accruals, computed as earnings before extraordinary items minus operating cash flow scaled by lagged total assets and standardized to mean zero and standard deviation one. PredOLS is the residual from a standard industry-year modified Jones model with three coefficients (change in cash revenues, PP&E, and the inverse of lagged total assets). PostMean is the posterior mean of our predictive distribution that incorporates parameter and model uncertainty (i.e., is averaged over six different versions of the modified Jones model). PostSD is the standard deviation of the same predictive distribution

predictions between OLS and the Bayesian approach (Table 2). In line with the literature, we measure abnormal accruals  $(DA_{i,t})$  as the difference between observed accruals  $(TA_{i,t})$  and the expected (mean) normal accruals  $(NDA_{i,t} = TA_{i,t}|D)$ :

$$DA_{i,t} = TA_{i,t} - NDA_{i,t} \tag{18}$$

We descriptively assess the quality of the measurement of normal and abnormal accruals by correlating these measures, calculated using the literature's standard approach and our Bayesian approach, with the observed accruals levels. The OLS-based normal and abnormal accruals are estimated using the predictions of industry-year-level modified Jones MODEL, including performance, leverage, and volatility controls (model 2). The Bayes-based normal and abnormal accruals are estimated using the model-averaged predictive distribution.

	ТА	PostSD	PostMean	PredOLS
TA		-0.18	0.83	0.56
PostSD	-0.13		-0.23	-0.22
PostMean	0.73	-0.17		0.71
PredOLS	0.45	-0.17	0.62	

Table 3 Correlations between Accruals, OLS Fit, and Bayesian predictions

The table presents correlations between total accruals, Bayesian accruals predictions (standard deviation and mean), and OLS predictions. TA is accruals, computed as earnings before extraordinary items minus operating cash flow scaled by lagged total assets and standardized to mean zero and standard deviation one. PredOLS is the residual from the extended industry-year modified Jones model with six determinants (inverse of lagged total assets, change in cash revenues, PP&E, return on assets, leverage, and revenue volatility). PostMean is the posterior mean of our predictive distribution that incorporates parameter and model uncertainty (i.e., is averaged over six different versions of the modified Jones model). PostSD is the standard deviation of the same predictive distribution. Upper (lower) triangle: Pearson (Spearman)



**Fig. 4** Comparison of Prediction Accuracy. The figure plots total accruals against predicted accruals and abnormal accruals (residuals). The predictions (and residuals) on the left are based on the extended modified Jones model (including performance, leverage, and volatility controls) fitted separately by industry-year using OLS. The predictions (and residuals) on the right are based on the posterior means of our averaged Bayesian models. One observation (TA = 78.30) was omitted from the plots for the sake of presentation

**Results** Table 3 presents univariate correlations of observed accruals with OLS-based and Bayes-based predicted normal accruals and measured abnormal accruals. The correlation between observed accruals and OLS-based normal accruals is a modest 0.56.<sup>23</sup> By contrast, the correlation between observed accruals and Bayes-based normal accruals is 0.83. The reverse pattern (mechanically) emerges for the correlations of observed accruals with OLS- vis-à-vis Bayes-based abnormal accruals. Figure 4 illustrates the relation between actual, normal, and abnormal accruals using scatter plots. The upper graphs document that the Bayes-based normal accruals better predict observed accruals than the OLS-based normal accruals do. Importantly, the Bayes-based abnormal accruals in the lower graphs are widely unrelated to observed accruals, whereas the OLS-based abnormal accruals are strongly positively associated with observed accruals.

 $<sup>^{23}</sup>$ Without performance, leverage, and volatility determinants, the industry-year OLS model exhibits a correlation with observed accruals as low as 0.30.

Under the plausible assumption that variation in observed accruals is primarily driven by first-order economic factors (e.g., business model differences), instead of second-order reporting incentives, the above patterns highlight that our Bayesian estimates better capture the normal structure of firms' accruals and hence allow identifying abnormal deviations. Notably, this substantially better fit is achieved simply by using some basic variations of a weak and often criticized model.

#### 4.3 Testing for abnormal accruals

**Specification** We test for (opportunistic) abnormal accruals using a measurementerror model (see Section 2.3). Using simulations, we compare the power to detect abnormal accruals and the frequency of false positives of our proposed Bayesian approach with the standard approach.<sup>24</sup> In our simulations, we are particularly interested in the impact of model misspecification and uncertainty on test power and false positives.

To investigate the impact of misspecification, we simulate three distinct data generating processes (DGP) for normal accruals. The three processes generate total accruals following the modified Jones model, that is, as a linear function of the inverse of total assets (InvAt), change in cash revenues ( $\triangle CRev$ ), and property, plant, and equipment (PPE). They differ only in the level of coefficient heterogeneity. The first process assumes that coefficients only vary by industry-year (consistent with the standard approach in the literature). The second assumes that coefficients vary by both industry-year and firm. The third assumes that coefficients only vary by firm.

We use the entire sample of 139,241 observations to simulate the three series of total accruals resulting from the three distinct processes. We combine firms' observed modified Jones model determinants with the process-specific coefficients to generate three simulated series of total accruals. We obtain the industry-year- or firm-level coefficients for the modified Jones model determinants via random draws from multivariate normal distributions (calibrated using previous posterior coefficient distributions). Equipped with these coefficients, we generate the simulated accruals for the distinct industry-year- and the firm-level processes by multiplying the modified Jones model determinants (observed for a given firm in a given year) with the respective industry-year- or firm-level coefficients and adding normally distributed noise. To generate the mixed industry-year and firm-level process, we add one-half of the simulated accruals from the industry-year-level process to one-half of the simulated accruals from the firm-level process. To both models to arrive at the model-averaged process).

For each of the three accruals series, we fit an industry-year-level modified Jones model using OLS, resulting in three OLS predictions of normal accruals  $(NDA_{OLS,i,t}^{DGP}, where DGP)$  is either *Industry-Year*, *Industry-Year* + *Firm*, or *Firm*).

<sup>&</sup>lt;sup>24</sup>Simulations allow us to control the data generating process. As a result, we can seed opportunistic abnormal accruals to test for power to detect truly abnormal accruals. Such power tests could only imperfectly be done by using AAERs, restatements, or comment letters, for example, as these are imperfect proxies for opportunism.

For the industry-year-level process, the OLS model is correctly specified. For the mixed industry-year and firm-level process and the purely firm-level process, by contrast, the OLS model is increasingly misspecified. We define OLS abnormal accruals as the difference between observed and predicted accruals using the industry-year-level modified Jones model estimated via OLS.

For the Bayesian estimation, we essentially assume there are two distinct candidate models. An industry-year-level model and a firm-level model. We fit both hierarchical models to each of the three accruals series. We combine the respective posterior predictive distributions of the two candidate models using model averaging. We define Bayes abnormal accruals as the difference between observed and predicted (mean) accruals measured as the mean of the model-averaged posterior predictive distribution.<sup>25</sup>

To compare test power across the Bayesian and the standard approach, we randomly add earnings management to the simulated accruals series following the example of Kothari et al. (2005). We proceed as follows. We first randomly select 20,000 observations from the entire sample. We next randomly add a constant earnings management amount to 5% of the selected sample observations.<sup>26</sup> We calibrate the earnings management amount at 5% of the standard deviation of the simulated total accruals. We finally run regressions of abnormal accruals on a true earnings management indicator (taking the value of one for all firm-years with earnings management and zero otherwise). For the OLS version, we simply regress the OLS abnormal accruals on the earnings management indicator (*EM*):

$$DA_{OLS,i,t}^{DGP} = a_0^{DGP} + a_1^{DGP} EM_{i,t}^{DGP} + u_{i,t}^{DGP}$$
(19)

where  $DA_{OLS,i,t}^{DGP} = TA_{i,t}^{DGP} - NDA_{OLS,i,t}^{DGP}$ . For the Bayes version, we implement a measurement-error model where we regress the Bayes abnormal accruals as a noisy measure of observed abnormal accruals  $(DA_{BAY,i,t}^{*DGP})$  on the earnings management indicator. We measure the level of noise in the Bayes abnormal accruals as the standard deviation of the model-averaged posterior predictive distribution ( $\sigma_{DA,i,t}$ ):

$$DA_{BAY,i,t}^{*DGP} = c_0^{DGP} + c_1^{DGP} E M_{i,t}^{DGP} + e_{i,t}^{DGP}$$
$$DA_{BAY,i,t}^{DGP} \sim N \left( DA_{BAY,i,t}^{*DGP}, \sigma_{DA,i,t} \right)$$
(20)

<sup>&</sup>lt;sup>25</sup>The simulation is set up to show when accounting for a broader set of models and uncertainty in parameter estimates is important and that the Bayesian approach does so (to the extent relevant models are considered in the model averaging). In our setup, the literature's standard model (industry-year-level Jones model) is true in one case, while it is misspecified in the other two. The Bayesian approach, by contrast, considers any mixture of the industry-year and the firm-level model, which is the full set of true models. We choose this setup to focus on the key aspects (misspecification and uncertainty) while abstracting from other issues (e.g., all models being incorrect). The patterns documented by the simulation should carry over to exercises where the Bayesian approach does not include all true models. There is still going to be a benefit of considering multiple models. It will just not be perfect. See Table A5 of the online appendix.

<sup>&</sup>lt;sup>26</sup>We first estimate the accruals models and then add earnings management. We follow this sequence to avoid having to repeat the time-intensive accruals-model estimation for each of the 500 iterations of the simulation.

For both the OLS and Bayes regressions, we record whether the coefficient on the earnings management indicator is statistically significant at the 5% level (using the 95% confidence interval for the OLS version and the 95% credible interval for the Bayes version).

To compare false positives across the Bayesian and the standard approach, we slightly adjust the steps above. We keep the previously added earnings management (to have the same sample for both tests) and assign a new but random earnings management indicator. This random indicator falsely indicates earnings management.

We repeat the steps for the test power and false positives simulation 500 times. The fraction of rejections of the true earnings management indicator provides a measure of power of the OLS and Bayes approaches to detect true earnings management for the three different accruals processes. The fraction of rejections of the false earnings management indicator provides a measure of false positives generated by the OLS and Bayes approaches for the three different accruals processes.

To investigate the impact of uncertainty on power and false positives, we augment our simulations with subsample tests. We test for power and false positives separately in samples with high versus low uncertainty or measurement error:

$$DA_{OLS,i,t}^{DGP} = a_0^{DGP} + a_1^{DGP} EM_{i,t}^{DGP} \times LOW_{i,t} + a_2^{DGP} EM_{i,t}^{DGP} \times HIGH_{i,t} + u_{i,t}^{DGP} DA_{BAY,i,t}^{*DGP} = c_0^{DGP} + c_1^{DGP} EM_{i,t}^{DGP} \times LOW_{i,t} + c_2^{DGP} EM_{i,t}^{DGP} \times HIGH_{i,t} + e_{i,t}^{DGP} DA_{BAY,i,t}^{*DGP} \sim N \left( DA_{BAY,i,t}^{*DGP}, \sigma_{DA,i,t}^{DGP} \right)$$
(21)

We use two measures of uncertainty. Our first splits the sample into observations from firms with a short time series (*HIGH*) versus those with a long time series (*LOW*). Fifty percent of the firms in our sample have a time series of six years or less. We use this median as the cutoff for the split. We expect that predicted and abnormal accruals of firms with a short time series contain more measurement error, especially in the Bayesian firm-level version. Our second measure splits the sample into observations with high (*HIGH*) versus low (*LOW*) posterior uncertainty of the Bayesian predictions. We label observations with a posterior uncertainty ( $\sigma_{DA,i,t}^{DGP}$ ) in the fourth quartile of the  $\sigma_{DA,i,t}^{DGP}$  distribution as high uncertainty observations.<sup>27</sup>

**Results** Table 4 summarizes the rates at which the null of no earnings management is rejected for the OLS and Bayesian versions across the different accruals processes and sample splits using a two-tailed 95% confidence/credible interval. In Panel A, we observe that the OLS and Bayesian approaches exhibit about the same amount of power to detect true earnings management (59% versus 60%) for the industry-year processes. By contrast, we observe substantially higher power to detect

<sup>&</sup>lt;sup>27</sup>We use a two-step approach to accelerate simulations. We first estimate normal accruals and then regress abnormal accruals on an earnings-management indicator. Given the setup of our simulation, this two-step approach is unbiased, because the first-step estimation occurred on a sample of unmanaged accruals and the second-step earnings-management is randomly assigned (i.e., uncorrelated with the normal-accruals determinants, Chen et al. 2018).

	lation results			
Assignment	DGP	Term	% Rejected (OLS)	% Rejected (Bayes)
Panel A: Mai	n regression			
True	Industry-Year	EM	0.586	0.596
	Industry-Year + Firm	EM	0.434	0.606
	Firm	EM	0.342	0.922
False	Industry-Year	EM	0.044	0.048
	Industry-Year + Firm	EM	0.048	0.044
	Firm	EM	0.058	0.050
Panel B: Cros	ss-sectional split by long	(low) versus sh	ort (high) time series l	ength
True	Industry-Year	EM×Low	0.454	0.484
		EM×High	0.160	0.160
	Industry-Year + Firm	EM×Low	0.350	0.506
		EM×High	0.184	0.182
	Firm	EM×Low	0.252	0.838
		EM×High	0.188	0.316
False	Industry-Year	EM×Low	0.050	0.046
		EM×High	0.056	0.052
	Industry-Year + Firm	EM×Low	0.044	0.048
		EM×High	0.072	0.036
	Firm	EM×Low	0.038	0.040
		EM×High	0.132	0.024
Panel C: Cros	ss-sectional split by low v	ersus high pos	terior uncertainty	
True	Industry-Year	EM×Low	0.442	0.442
		EM×High	0.196	0.186
	Industry-Year + Firm	EM×Low	0.264	0.520
		EM×High	0.246	0.212
	Firm	EM×Low	0.202	0.832
		EM×High	0.210	0.266
False	Industry-Year	EM×Low	0.040	0.044
		EM×High	0.050	0.054
	Industry-Year + Firm	EM×Low	0.030	0.056
		$EM \times High$	0.108	0.040
	Firm	EM×Low	0.014	0.054
		$\mathrm{EM}{ imes}\mathrm{High}$	0.132	0.016

#### Table 4 Simulation results

The table presents rejection rates from earnings management tests using simulated data. The rates are based on 500 draws from the three data-generating processes: Industry-Year, Industry-Year + Firm, and Firm. Each draw samples 20,000 observations and randomly adds earnings management of 5% of the standard deviation of the simulated accruals (Industry-Year 0.063, Industry-Year + Firm 0.049, and Firm 0.073) to 5% of the observations. True denotes earnings management dummies (EM) that correctly identify managed earnings. False denotes earnings managements dummies that are randomly assigned to 5% of the samples. Panel B splits the samples into observations from firms with a time series longer than six years (Low) and observations from firms with shorter time series (High). Panel C splits the samples into observations with a posterior standard deviation higher than the third quartile (High) and observations with lower posterior standard deviation (Low)



% Rejected (Bayes) — % Rejected (OLS)

**Fig. 5** Main Simulation Results. The figure shows rejection rates for two tests of earnings management using simulated data. Statistics are based on 500 draws from the three data generating processes (DGP): industry-Year, averaged, and firm. Each draw samples 20,000 observations and randomly adds earnings management of 5% of the standard deviation of the simulated accruals to 5% of the observations. For each sample, the null of no earnings management is tested. The left figure shows tests using an earnings management dummy that correctly identifies managed earnings. The right figure shows tests using a randomly assigned dummy (for observations with a posterior standard deviation greater than the third quartile of sample)

earnings management for the Bayesian vis-à-vis the OLS approach for the averaged and the firm-level processes. Notably, power decreases for the OLS approach to 43% and 34% for the averaged process and firm-level process, respectively, whereas it increased for the Bayesian approach to 61% and 92% (Fig. 5).

These patterns suggest that model misspecification significantly reduces the power of the OLS approach, especially if the researcher's model is an industry-year model when the true model is a firm-level model. The Bayes approach does not suffer from a loss of power due to misspecification. It adapts its model weights across the different processes to best fit the true process. Table 5 shows that the Bayesian approach relies (almost) completely on the industry-year-level predictions for the industry-year

	Fitted Model	
DGP	Industry-Year Model	Firm Model
Industry-Year	1.000	0.000
Industry-Year + Firm	0.226	0.774
Firm	0.015	0.985

 Table 5
 Simulation stacking weights

The table presents the stacking weights computed after fitting an industry-year model and a firm model on each of the three data-generating processes. The stacking weights are computed following Yao et al. (2018)

process and on the firm-level predictions for the firm-level process. For the averaged process, it puts a weight of 23% on the industry-year-level predictions and 77% on the firm-level predictions. These weights differ from the underlying 50%-50% mix of the averaged process. The greater weight on the firm-level predictions likely reflects that the firm-level model already accounts for the across industry heterogeneity of the industry-year process. Thus only the year-to-year variation by industry remains to be explained by the industry-year model. Interestingly, the Bayesian approach not only does not suffer from loss of power, due to misspecification, but even increases its power, especially for the firm-level processes. One reason for this increase might be that there are more firms than industry-years in the sample. Accordingly, the parameters governing the across-firm distribution of coefficients can be estimated more precisely than the parameters governing the across-industry-year distribution of coefficients.

With respect to false positives, we observe that the OLS and Bayesian approaches exhibit about the same rate of false positives (close to the expected 5% level) for all processes in Panel A. Accordingly, misspecification per se does not, on average, appear to substantially increase false positives for the OLS approach. Once we partition observations in our simulated samples by low versus high uncertainty, however, we observe that model misspecification notably increases the rate of false positives. In Panel B, we observe that, for the subsample of firms with few firm-year observations, the false positive rate of the OLS approach increases from 5.6% to 7.2% and 13.2%, as we go from the industry-year process to the averaged and firm-level processes. By contrast, the false positive rate of the Bayesian approach decreases from 5.2% to 3.6% and 2.4%, as we go from the industry-year process to the averaged and firm-level processes. In Panel C, we observe the same pattern for the subsample of firms with the highest (posterior) uncertainty (Fig. 5).<sup>28</sup>

These patterns suggest that the combination of model misspecification and uncertainty can lead to substantially increased false positive rates for the OLS approach. By contrast, greater uncertainty leads the Bayesian approach to down-weight its evidence, resulting in fewer (false) rejections.

Taken together, our simulation results support the conceptual merits of the Bayesian approach over the standard approach. The former approach appears to exhibit more power and fewer false positives as a result of its inclusion of model and parameter uncertainty. Notably, our simulations support concerns (e.g., Ball 2013) that the standard approach may produce many false positives, due to model misspecification in combination with treatment indicators correlated with uncertainty. Our simulation documents that the power of the literature's standard approach to detecting earnings management is low, especially for firm-level accruals processes and subsamples characterized by high uncertainty. By contrast, the rate of false positives is highest in these cases. These patterns are noteworthy, as our Bayesian

<sup>&</sup>lt;sup>28</sup>The (rejection) magnitudes depend on the parameters of the simulation (e.g., degree of noise). They can be scaled up or down depending on the simulation parameters. Accordingly, our primary emphasis is on comparing the relative trends and patterns across the various data generating processes and sample splits for the literature's standard approach vis-à-vis the Bayesian one.

model-averaging approach suggests that observed accruals processes in the data indeed are better described by firm-level than industry-year-level heterogeneity. Moreover, earnings management (treatment) indicators used in the literature are often correlated with uncertainty (e.g., younger and volatile firms). This combination of model misspecification and selection on high uncertainty subsamples should be expected to favor false positives.

In the online appendix, we document that our inferences from the simulation are robust to using alternative forms of earnings management and data generating processes. We, for example, show that the Bayesian approach also provides greater power to detect earnings management and produces fewer false positives in subsamples with high uncertainty when modeling earnings management as a deterministic function of earnings (e.g., Einhorn and Ziv 2020; Bertomeu et al. 2021; McClure and Zakolyukina 2021), instead of a random constant. Similar patterns obtain when modeling earnings management as random noise (e.g., Fischer and Verrecchia 2000; Dye and Sridhar 2004; Beyer et al. 2019) and using absolute abnormal accruals as an outcome. Lastly, we show that, in subsamples with high uncertainty, the Bayesian approach produces fewer false positives, even if the underlying data generating process is not perfectly captured by its candidate models (or a mix thereof). These patterns highlight that the benefits of the Bayesian approach extend above and beyond the controlled environment of our stylized simulation.

# **5** Application

**Specification** In a final step, we apply our Bayesian approach to an extended set of models, including some of the most popular and recent accruals models. Following a review of the literature, we consider variants of the models developed by Jones (1991), Dechow et al. (1995), Kothari et al. (2005), Dechow and Dichev (2002), McNichols (2002), Ball and Shivakumar (2006), Collins et al. (2017), and Frankel and Sun (2018).<sup>29</sup> As before, we estimate industry-year and firm-level variants of these models. Equipped with the estimates of these model variants, we calculate the weights assigned to each of the variants and the corresponding averaged model.

**Results** Table 6 reports the individual model specifications and their weights. The basic modified Jones model receives almost no weight. This pattern is consistent

<sup>&</sup>lt;sup>29</sup>We anchor on these popular models (Fig. A1 in the online appendix) because they reflect information about the structure of accruals derived from theory and the properties of accounting. For ease of implementation, we restrict the set of models to those using (unweighted) OLS specification. Accordingly, we include models using flexible control functions to implicitly match on firm characteristics (e.g., Collins et al. 2017). By contrast, we omit models that rely on explicit matching approaches (such as entropy balancing; e.g., McMullin and Schonberger 2020). Explicit matching models tend to be motivated by concerns about nonlinearities in the cross-firm relation between accruals and firm characteristics. We expect the firm-level variants to help account for such nonlinearities by focusing on within-firm relations. The limited weight assigned to the Collins et al. (2017) model, which uses flexible controls to account for nonlinearities, provides support for this expectation.

#### Table 6 Application stacking weights

#### Panel A:

Fitted Model	Model Equation
ModJones	$NDA_{t} = \beta_{0}InvAt_{t-1} + \beta_{1} \triangle CRev_{t} + \beta_{2}PPE_{t}$
ModJonesROA	$NDA_{t} = \beta_{0}InvAt_{t-1} + \beta_{1} \triangle CRev_{t} + \beta_{2}PPE_{t} + \beta_{3}ROA_{t}$
AQ	$NDA_t = \beta_0 CFO_{t-1} + \beta_1 CFO_t$
ModJonesAQ	$NDA_{t} = \beta_{0}InvAt_{t-1} + \beta_{1} \triangle CRev_{t} + \beta_{2}PPE_{t} + \beta_{3}CFO_{t-1} + \beta_{4}CFO_{t}$
Collins	$NDA_t = \beta_0 + \beta_1 \triangle CRev_t + \sum_{k=2}^{5} \beta_{2,k} ROAQunitile_t^k$
	$+\sum_{k=2}^{5} \beta_{3,k} SGQunitile_t^k + \sum_{k=2}^{5} \beta_{4,k} BtoMQunitile_t^k$
BallShiva	$NDA_{t} = \beta_{0} + \beta_{1} \triangle CRev_{t} + \beta_{2}PPE_{t} + \beta_{3}ROA_{t} + \beta_{4}CFO_{t}$
	$+\beta_5 NegCFO_t + \beta_6 NegCFO_t \cdot CFO_t$
FrankelSun	$NDA_t = \beta_0 + \beta_1 \triangle CRev_t + \beta_2 OpCycle_t$
	$+\beta_3 \triangle CRev_t \cdot OpCycle_t + \beta_4 \triangle CFO_t + \beta_5 \triangle OCFSC_t$
	$+\beta_{6} \triangle CFO_{t} \cdot \triangle OCFSC_{t}$

#### Panel B:

Fitted Model	Weight
Industry-year-level r	nodel variants
ModJones	0.005
ModJonesROA	0.034
AQ	0.012
ModJonesAQ	0.000
Collins	0.053
BallShiva	0.004
FrankelSun	0.088
Firm-level model van	riants
ModJones	0.000
ModJonesROA	0.050
AQ	0.000
ModJonesAQ	0.243
Collins	0.097
BallShiva	0.065
FrankelSun	0.349

The table presents the stacking weights computed following Yao et al. (2018) and used to compute the averaged (ab)normal accruals used in the tests reported in Tables 7 to 10. Industry-year model variants are estimated with coefficients varying per industry-year. Firm model variants are estimated with coefficients varying per firm. To avoid lookahead bias, the models AQ and ModJonesAQ do not include  $CFO_{t+1}$  (following Frankel and Sun 2018).  $CFO_t$  is operating cash flows  $((cancf_t)/at_{t-1})$ .  $NegCFO_t$  is an indicator if operating cash flows is negative.  $OpCycle_t$  is the average percentile rank of operating cycle over the past three years  $(0.5*(rect_t + rect_{t-1})/sale_t + 0.5*(invt_t + invt_{t-1})/cogs_t - 0.5*(ap_t + ap_{t-1})/purchases_t)$ .  $OCFSC_t$  is the average percentile rank of OpCycle and  $(1 - net profit margin <math>ib_t/sale_t$ ).  $BtoM_t$  is the book-to-market ratio at the beginning of the fiscal period  $((ceq_{t-1}/mkvalt_{t-1}))$ .  $SG_t$  is revenue growth  $((sale_t/sale_{t-1})$ . All other variables are defined as in Table 1. Weights are computed for a sample of 62,022 observations from 1988-2017, remaining after the data cleaning steps described in Table 1 and requiring nonmissing values for all of the models' variables

with subsequent refinements (e.g., Kothari et al. 2005) not only improving but essentially subsuming the basic model. It echoes our earlier finding that an extended Jones model, which controls for firm characteristics, such as performance, dominates the basic Jones model. All other models receive a nonnegligible weight for at least one of their respective variants (industry-year or firm-level). This pattern highlights the usefulness of considering a broad set of models. The greatest weights are assigned to firm-level variants of the McNichols (2002) model, one of the most popular models, which combines the Jones (1991) and the Dechow and Dichev (2002) models, and the Frankel and Sun (2018) model, the most recent one. Consistent with our earlier results, these patterns reiterate the need to account for firm-level heterogeneity. They also document that the combination of two seminal models (Jones 1991; Dechow et al. 1995) helps in modelling accruals and that recent work by Frankel and Sun (2018) substantially improves this modelling.

We assess the usefulness of the averaged model by examining its ability to explain future performance and earnings management indicators (AAERs, restatements, and comment letters). Following Frankel and Sun (2018), Table 7 reports the results from regressions of future performance on current performance, constructed by combining current cash flow from operations and normal accruals predicted by the respective accruals models. We find that the current performance measure based on our averaged model's (mean) normal accruals better explains future performance than current performance measures constructed using the normal accruals produced by the individual models.<sup>30</sup> The averaged model achieves an  $R^2$  of 54% (column 1), whereas the individual models exhibit explanatory power ranging from 45% to 48% (columns 2 to 8). The results suggest the averaged model significantly improves over and above the individual models' ability to capture the structure of normal accruals.

Table 8 reports the results from regressions of indicators for (income-increasing) AAERs on abnormal accruals. We find that abnormal accruals are positively and statistically significantly associated with AAERs, irrespective of the accruals model used to estimate abnormal accruals. Compared to the abnormal accruals produced by the individual models, the ones produced by the averaged model exhibit a higher coefficient magnitude. The averaged model produces a slope coefficient of 0.035, suggesting a one standard-deviation increase of abnormal accruals is associated with a 3.5% increase in the likelihood of a financial reporting enforcement action. The slope coefficients of the individual models, by contrast, range from 0.018 to 0.031. These results are consistent with the averaged model producing less noisy measures of abnormal accruals.<sup>31</sup>

<sup>&</sup>lt;sup>30</sup>In keeping with the literature, the normal accruals of the individual models are estimated at the industryyear level using OLS. To avoid a mechanical improvement of the prediction ability, we exclude controls for future cash flows in estimating the individual models and the averaged model.

<sup>&</sup>lt;sup>31</sup>Noise due to measurement error can be expected to bias the relation between AAERs and abnormal accruals toward zero.

	Averaged (1)	ModJones (2)	ModJonesROA (3)	AQ (4)	ModJonesAQ (5)	Collins (6)	BallShiva (7)	FrankelSun (8)
Panel A: Coefficient	ts and R <sup>2</sup> of future	e earnings regress	ions					
(Intercept)	$-0.005^{***}$	$-0.005^{***}$	$-0.005^{***}$	-0.005***	$-0.005^{***}$	$-0.005^{***}$	$-0.005^{***}$	$-0.005^{***}$
	[-10.687]	[-9.364]	[-10.533]	[-9.234]	[-9.472]	[-10.046]	[-9.455]	[-9.984]
$EARN_{NDA}$	$0.849^{***}$	0.795***	$0.720^{***}$	$0.857^{***}$	$0.850^{***}$	$0.739^{***}$	$0.851^{***}$	$0.777^{***}$
	[259.896]	[216.688]	[227.235]	[221.556]	[226.270]	[225.991]	[221.255]	[231.594]
$\mathbb{R}^2$	0.542	0.451	0.475	0.462	0.473	0.472	0.461	0.484
N.Obs.	57,128	57,128	57,128	57,128	57,128	57,128	57,128	57,128
Panel B: Vuong z-s	tatistics for R <sup>2</sup> con	mparison						
Averaged		8.85***	$6.61^{***}$	7.78***	6.79***	6.89***	7.83***	5.70***
ModJones	-8.85		-20.54	-6.67	-14.98	-14.92	-7.36	-18.60
ModJonesROA	-6.61	20.54***		$7.17^{***}$	$1.30^{*}$	2.05**	8.44***	-5.50
AQ	-7.78	6.67***	-7.17		-9.25	-5.35	0.41	-14.36
ModJonesAQ	-6.79	$14.98^{***}$	-1.30	9.25***		0.35	7.37***	-8.34
Collins	-6.89	$14.92^{***}$	-2.05	5.35***	-0.35		$6.20^{***}$	-7.34
BallShiva	-7.83	7.36***	-8.44	-0.41	-7.37	-6.20		-11.81
FrankelSun	-5.70	$18.60^{***}$	5.50***	$14.36^{***}$	8.34***	7.34***	$11.81^{***}$	
Table 7, Panel A sho $CFO_t + NDA_t$ for 2018). For the Bayes earnings and all moc row model's $\mathbb{R}^2$ bein respectively	ws simple OLS reg each model using th sian averaged mode lel variables, such a g significantly larg	pressions of the for the model's predict al (Averaged), we a as beginning book for than the column	m $EARN_{t+1} = a_0 +$ ed $NDA_t$ . To avoid le use the posterior meau -to-market ratios, redu 1 model. ***, **, * in	$a_1 E ARN_{NDA,I} +$ ookahead bias, the 1 ons as our point preduce the sample to 5 dicate statistical si	$u_{i+1}$ . $EARN_{NDA,i}$ nodels AQ and Moc liction of $NDA_i$ . A 7,128 observations) gnificance at the 1 p	is a measure of no llonesAQ do not ir dditional variable . Panel B shows te bercent, 5 percent,	ormal earnings and nclude $CFO_{i+1}$ (F requirements (e.g., est statistics of Vuc and 10 percent con	is computed as rankel and Sun such as future ong tests of the nfidence levels,

Table 7 Future performance

	Averaged (1)	ModJones (2)	ModJonesROA (3)	AQ (4)	ModJonesAQ (5)	Collins (6)	BallShiva (7)	FrankelSun (8)
DA	0.032**	0.023**	0.025***	0.029***	0.022**	0.025***	0.016	0.025**
	[2.269]	[2.461]	[2.676]	[2.846]	[2.176]	[2.646]	[1.612]	[2.462]
RoA	$0.023^{***}$	$0.023^{***}$	$0.028^{***}$	$0.020^{**}$	$0.022^{***}$	$0.025^{***}$	$0.022^{***}$	$0.022^{***}$
	[2.823]	[2.777]	[3.185]	[2.487]	[2.705]	[3.008]	[2.782]	[2.677]
Size	0.006***	0.006***	0.006***	0.006***	0.006***	$0.006^{***}$	$0.006^{***}$	$0.006^{***}$
	[7.143]	[7.154]	[7.145]	[7.155]	[7.150]	[7.147]	[7.142]	[7.148]
Lev	$-0.011^{**}$	$-0.011^{**}$	$-0.011^{**}$	$-0.011^{**}$	$-0.011^{**}$	$-0.011^{**}$	$-0.011^{**}$	$-0.011^{**}$
	[-2.461]	[-2.479]	[-2.456]	[-2.429]	[-2.460]	[-2.348]	[-2.384]	[-2.460]
MPM	-0.009**	$-0.010^{**}$	$-0.010^{**}$	$-0.010^{**}$	$-0.010^{**}$	$-0.010^{**}$	$-0.010^{**}$	$-0.007^{*}$
	[-2.193]	[-2.449]	[-2.438]	[-2.408]	[-2.392]	[-2.327]	[-2.388]	[-1.837]
StdSale	$0.014^{***}$	$0.014^{***}$	$0.014^{***}$	$0.013^{***}$	$0.014^{***}$	$0.013^{***}$	$0.014^{***}$	$0.014^{***}$
	[4.130]	[4.094]	[4.061]	[3.969]	[4.146]	[4.017]	[4.156]	[4.049]

AAERs
œ
Ð
9
Ъ

	Averaged (1)	ModJones (2)	ModJonesROA (3)	AQ (4)	ModJonesAQ (5)	Collins (6)	BallShiva (7)	FrankelSun (8)
OpCycle	0.014*** [3.589]	0.014*** [3.454]	0.014*** [3.449]	$0.014^{***}$ $[3.430]$	0.014*** [3.481]	$0.014^{***}$ $[3.442]$	0.014*** [3.539]	0.015*** [3.597]
Industry FE Year FE	Yes Yes	Yes Yes	Yes Yes	Yes Yes	Yes Yes	Yes Yes	Yes Yes	Yes Yes
P.R <sup>2</sup>	0.102	0.102	0.103	0.103	0.102	0.103	0.101	0.102
N.Obs.	31,709	31,709	31,709	31,709	31,709	31,709	31,709	31,709
Table 8 shows the $a_7 OpCycle_t + u_t$ non-missing value a firm is labeled b Table 1. The avera	e average partial The sample cor cs for each model y the SEC as emp se narrial effects	effects of the follo hists of all two-dig 's variables. To avo ploying upward ear are hias-corrected	wing logistic regression it-SIC-year subsamples bid lookahead bias, the n nings management in ye remaindeVal and Wei	$I_{AAER_i} = a_i + \epsilon$ in which at least c nodels AQ and Mc art and 0 otherwii dner 2016) PR <sup>2</sup> i	$u_t + a_1 DA_t + a_2 RoA$ ine firm had received a odJonesAQ do not incli se (Dechow et al. 2011)	$a_{1} + a_{3}Size + a_{4}I$ in AAER. We furth ude $CFO_{i+1}$ (Fraude $CF$ ). The remaining $v$	$Lev_t + a_5 NPM_t +$ her require all obser nkel and Sun 2018) ariable definitions	$a_6StdSale_t +$ rvations to have $A_{AAER_t}$ is 1 if can be found in

Table 8 (continued)

model is estimated instead. \*\*\*, \*\*, \* indicate statistical significance at the 1 percent, 5 percent, and 10 percent confidence levels, respectively

	Averaged (1)	ModJones (2)	ModJonesROA (3)	AQ (4)	ModJonesAQ (5)	Collins (6)	BallShiva (7)	FrankelSun (8)
ADA	0.074**	0.014	0.025	0.004	0.012	0.014	-0.011	0.021
	[2.925]	[0.937]	[1.549]	[0.269]	[0.706]	[0.830]	[-0.625]	[1.208]
RoA	$0.056^{**}$	0.056**	0.056**	0.056**	$0.056^{**}$	$0.056^{**}$	0.055**	$0.056^{**}$
	[4.714]	[4.722]	[4.714]	[4.671]	[4.714]	[4.703]	[4.587]	[4.733]
Size	$0.009^{**}$	$0.009^{**}$	0.009**	0.009**	0.009**	$0.009^{**}$	0.009**	$0.009^{**}$
	[8:098]	[8.066]	[8.082]	[8.054]	[8.065]	[8.063]	[8.037]	[8.072]
Lev	0.002	0.001	0.002	0.001	0.001	0.002	0.001	0.002
	[0.338]	[0.238]	[0.261]	[0.233]	[0.239]	[0.245]	[0.223]	[0.254]
MPM	$-0.040^{**}$	$-0.042^{**}$	$-0.042^{**}$	$-0.042^{**}$	$-0.042^{**}$	$-0.042^{**}$	$-0.042^{**}$	$-0.041^{**}$
	[-5.452]	[-5.690]	[-5.655]	[-5.697]	[-5.681]	[-5.681]	[-5.714]	[-5.627]
StdSale	$0.022^{**}$	$0.022^{**}$	$0.021^{**}$	$0.022^{**}$	$0.022^{**}$	$0.022^{**}$	$0.023^{**}$	$0.022^{**}$
	[2.995]	[2.986]	[2.946]	[3.041]	[3.022]	[3.008]	[3.132]	[2.994]

	Averaged (1)	ModJones (2)	ModJonesROA (3)	AQ (4)	ModJonesAQ (5)	Collins (6)	BallShiva (7)	FrankelSun (8)
OpCycle	0.027** [4.840]	0.027** [4.845]	0.027*** [4.855]	0.027** [4.843]	0.027** [4.850]	0.027** [4.850]	$0.027^{**}$ $[4.841]$	0.028** [4.867]
Industry FE Year FE	Yes Yes	Yes Yes	Yes Yes	Yes Yes	Yes Yes	Yes Yes	Yes Yes	Yes Yes
P.R <sup>2</sup> N.Obs.	0.077 55,664	0.076 55,664	0.076 55,664	0.076 55,664	0.076 55,664	0.076 55,664	0.076 55,664	0.076 55,664
Table 9 shows th a <sub>7</sub> $OPCycle_t + u_t$ to Audit Analytic to Audit Analytic do not include C (Hennes et al. 20) the McFadden Ps and 10 percent co	e average partial e e average partial e S. The sample consi s. We further requ <i>FO</i> <sub>+1</sub> (Frankel a <i>FO</i> <sub>+1</sub> (Frankel a <i>B</i> <sub>0</sub> . The remaining eudo R <sup>2</sup> . Inference mfidence levels, re	effects of the follow lists of all two-digit-tists in all observations and Sun 2018). $I_{RE}$ g variable definition gs are unchanged if spectively.	ing logistic regression $I_l$ SIC-year subsamples in w to have nonmissing value $s_{T_l}$ is 1 if a firm restated s can be found in Table a linear probability mode	$\alpha_{EST_{i}} = a_{i} + a_{i}$ thich at least one les for each moc d its financials t l. The average p el is estimated in	$+ a_1 ADA_i + a_2 RoA_i$ firm had a restatement, i el's variables. To avoid ecause of a nonclerical artial effects are bias-co stead. ***, **, * indicat	$+ a_3 Size + a_4$ , due to accountin, lookahead bias, l issue in the acc orrected (Fernánc te statistical sign	$Lev_t + a_5 N P M_t + Berline A polication far grule application far the models AQ and the models AQ and you the invert a southing in year t a counting in year t and Weidme fiftcance at the 1 per fifteance at the 1 per fifteanc$	$a_6StdSale_1 +$ ilure, according 1 ModJonesAQ nd 0 otherwise r 2016). P.R <sup>2</sup> is cent, 5 percent,

Table 9 (continued)

	Averaged (1)	ModJones (2)	ModJonesROA (3)	AQ (4)	ModJonesAQ (5)	Collins (6)	BallShiva (7)	FrankelSun (8)
ADA	$0.101^{*}$	0.035	0.035	0.023	0.034	0.032	0.022	0.045
	[1.934]	[1.049]	[0.976]	[0.663]	[0.942]	[0.862]	[0.625]	[1.155]
RoA	$0.091^{**}$	$0.092^{**}$	$0.091^{**}$	$0.091^{**}$	$0.092^{**}$	$0.091^{**}$	$0.091^{**}$	$0.092^{**}$
	[4.266]	[4.292]	[4.259]	[4.261]	[4.287]	[4.268]	[4.252]	[4.294]
Size	0.045**	0.045**	0.045**	$0.045^{**}$	$0.045^{**}$	$0.045^{**}$	$0.045^{**}$	$0.045^{**}$
	[12.267]	[12.256]	[12.257]	[12.259]	[12.259]	[12.257]	[12.257]	[12.260]
Lev	0.003	0.002	0.003	0.003	0.003	0.003	0.003	0.003
	[0.233]	[0.205]	[0.215]	[0.206]	[0.209]	[0.214]	[0.211]	[0.220]
MPM	$-0.041^{**}$	$-0.043^{**}$	$-0.043^{**}$	$-0.043^{**}$	$-0.043^{**}$	$-0.043^{**}$	$-0.043^{**}$	$-0.042^{**}$
	[-3.703]	[-3.950]	[-3.929]	[-3.954]	[-3.938]	[-3.943]	[-3.952]	[-3.873]
StdSale	$0.036^{**}$	$0.035^{**}$	$0.035^{**}$	$0.036^{**}$	$0.036^{**}$	$0.036^{**}$	$0.036^{**}$	$0.035^{**}$
	[2.364]	[2.311]	[2.330]	[2.335]	[2.344]	[2.343]	[2.368]	[2.340]

Table 10 Comment letters

	Averaged (1)	ModJones (2)	ModJonesROA (3)	AQ (4)	ModJonesAQ (5)	Collins (6)	BallShiva (7)	FrankelSun (8)
OpCycle	0.044**	0.044**	0.045**	0.044**	0.045**	0.045**	0.044**	0.045**
	[4.492]	[4.502]	[4.512]	[4.498]	[4.508]	[4.511]	[4.504]	[4.535]
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
P.R <sup>2</sup>	0.142	0.142	0.142	0.142	0.142	0.142	0.142	0.142
N.Obs.	44,250	44,250	44,250	44,250	44,250	44,250	44,250	44,250
Table 10 shows th $a_7 OpCycle_t + u_t$ . require all observal and Sun 2018). $IC$ variable definitions are unchanged if a	e average partial ef The sample consis tions to have nomm $L_i$ , is 1 if a firm h s can be found in T8 linear probability	ffects of the followi sts of all two-digit-S issing values for eac as received a comm able 1. The average model is estimated	ng logistic regression <i>Ic</i> IC-year subsamples in w th model's variables. To nent letter by the SEC p partial effects are bias-co instead. ***, **, * indic	$L_i = a_i + a_r + c_i$ thich at least one f avoid lookahead b bertaining to a 10- prected (Fernándo cate statistical sign	$i_1 A D A_i + a_2 R o A_i + irm received a commential received a commential, the models AQ and K or 10-Q in year t and Weidner 2011 inficance at the 1 perceived and weidner at the 1 perceived and 1 perceived and the 1 perceived and $	$a_3Size + a_4Lev_1$ at letter according 1 ModJonesAQ dc nd 0 otherwise (I (6). P.R <sup>2</sup> is the Mo	+ $a_5NPM_t$ + $a_6$ , to Audit Analytics. to not include $CFO_t$ bu et al. 2020). The Fadden Pseudo $\mathbb{R}^2$	StdSale <sub>t</sub> + . We further $_{++1}$ (Frankel $_{-}$ remaining $_{-}$ Inferences ence levels,

Table 10 (continued)

respectively

Tables 9 and 10 report the results from regressions of indicators for restatements (related to accounting rule misapplications) and comment letters on absolute abnormal accruals. Following Dechow et al. (2011) and Du et al. (2020), we use *absolute* abnormal accruals to explain restatements and comment letters, because these incidences, unlike the income-increasing AAERs, capture both income-increasing and -decreasing reporting abnormalities. We find that the absolute abnormal accruals produced by our averaged model are positively and statistically significantly associated with both restatements (Table 9) and comment letters (Table 10). The absolute abnormal accruals produced by the individual models, by contrast, exhibit economically lower and statistically insignificant associations. These results are consistent with the averaged model exhibiting greater power to detect earnings management than any of the individual models.<sup>32</sup> Notably, the averaged model appears to increase power, especially for absolute abnormal accruals. Absolute abnormal accruals measures tend to suffer particularly from errors in the estimation of abnormal accruals (Hribar and Nichols 2007).

## 6 Conclusion

We provide an applied introduction to key methods of Bayesian estimation of relevance to empirical accounting research. To illustrate the methods, we apply them to an issue of central concern to the literature and practice: the detection of accruals-based earnings management.

Our proposed Bayesian approach explicitly incorporates uncertainty about parameters and models in the estimation of normal (or unmanaged) accruals and tests for abnormal (or suspect) accruals. We document that the Bayesian approach significantly reduces the extent of accruals that can confidently be classified as abnormal and lowers the incidence of false positives in tests for earnings management. Our approach even increases the power to detect earnings management (i.e., reduce false negatives). This improvement is due to the fact that it allows for firm-level heterogeneity in accruals processes via hierarchical modelling. Firm-level heterogeneity appears to be key for improving the modelling of normal accruals and detection of abnormal accruals.

We illustrate the Bayesian approach with the help of the popular and familiar modified Jones model to ease comparison with the literature's standard estimation approach. Our proposed approach, however, is *not* model specific. To the contrary, it can combine several models, such as the modified Jones model and other popular and more recent models (e.g., Dechow and Dichev 2002; Frankel and Sun 2018), to obtain superior normal and abnormal accruals estimates. Even more so, it can be applied to the problem of estimating normal levels of a host of variables of interest (e.g., productivity, investment, tone, or readability), not just accruals. Given its broad

<sup>&</sup>lt;sup>32</sup>As each individual model produces a lower association than the averaged model, the results in Tables 9 and 10 illustrate that inferences drawn from the average association across multiple models are weaker than those drawn from the association of an averaged model.

applicability, conceptual appeal, and increasing computational feasibility, we expect the Bayesian tools explained here to be widely adopted and to contribute to more credible inferences in the accounting literature.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s11142-021-09654-0.

Acknowledgements We thank the editor, Peter Easton, and two anonymous referees for many insightful comments and suggestions. We gratefully acknowledge helpful comments from Ray Ball, Joseph Gerakos, Trevor Harris, Katharina Hombach, Martin Jacob, Nico Lehmann, Christian Leuz, Valeri Nikolaev, Jiri Novak, Stephen Penman, Thomas Ruchti (discussant), Catherine Schrand, Christoph Sextroh, Robert Stoumbos, David Windisch, David Veenman, and seminar participants at Columbia University, Tilburg University, the 2019 EASYS, the 2020 FARS Midyear Meeting, and the 2021 EAA Annual Meeting. We provide sample data and code here: github.com/hschuett/AccForUncertaintyCode.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

## References

- Ball, R. 2013. Accounting informs investors and earnings management is rife: two questionable beliefs. Accounting Horizons 27(4): 847–853. https://doi.org/10.2308/acch-10366.
- Ball, R., and L. Shivakumar. 2006. The role of accruals in asymmetrically timely gain and loss recognition. *Journal of Accounting Research* 44(2): 207–242. https://doi.org/10.1111/j.1475-679X.2006.00198.x.
- Barker, R., S. Penman, T.J. Linsmeier, and S. Cooper. 2020. Moving the conceptual framework forward: accounting for uncertainty. *Contemporary Accounting Research* 37(1): 322–357. https://doi.org/10.1111/1911-3846.12585.
- Belloni, A., V. Chernozhukov, and C. Hansen. 2013. Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies* 81(2): 608–650. https://doi.org/10.1093/restud/rdt044.
- Bertomeu, J., E. Cheynel, E.X. Li, and Y. Liang. 2021. How pervasive is earnings management? evidence from a structural model. *Management Science* 67(8): 5145–5162. https://doi.org/10.1287/mnsc.2020.3717.
- Betancourt, M. 2018. A conceptual introduction to Hamiltonian Monte Carlo. Working Paper. Available at: arXiv:1701.02434.
- Beyer, A., I. Guttman, and I. Marinovic. 2019. Earnings management and earnings quality: theory and evidence. *The Accounting Review* 94: 77–101. https://doi.org/10.2308/accr-52282.
- Biddle, G.C., and G. Hilary. 2006. Accounting quality and firm–level capital investment. *The Accounting Review* 81(5): 963–982. https://doi.org/10.2308/accr.2006.81.5.963.
- Biddle, G.C., G. Hilary, and R.S. Verdi. 2009. How does financial reporting quality relate to investment efficiency? *Journal of Accounting and Economics* 48(2): 112–131. https://doi.org/10.1016/j.jacceco.2009.09.001.
- Bloomfield, M.J., J. Gerakos, and A. Kovrijnykh. 2017. Accrual reversals and cash conversion. Working Paper. Available at: https://ssrn.com/abstract\_id=24 95610.
- Bushee, B.J., I.D. Gow, and D.J. Taylor. 2018. Linguistic complexity in firm disclosures: obfuscation or information? *Journal of Accounting Research* 56(1): 85–121. https://doi.org/10.1111/1475-679X.12179.

- Carpenter, B., A. Gelman, M.D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. 2017. Stan: a probabilistic programming language. *Journal of Statistical Software* 76(1): 1–32. https://doi.org/10.18637/jss.v076.i01.
- Chen, W., P. Hribar, and S. Melessa. 2018. Incorrect inferences when using residuals as dependent variables. *Journal of Accounting Research* 56(3): 751–796. https://doi.org/10.1111/1475-679X.12195.
- Collins, D.W., R.S. Pungaliya, and A.M. Vijh. 2017. The effects of firm growth and model specification choices on tests of earnings management in quarterly settings. *The Accounting Review* 92(2): 69–100. https://doi.org/10.2308/accr-51551.
- Dechow, P., W. Ge, and C. Schrand. 2010. Understanding earnings quality: a review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics* 50(2): 344–401. https://doi.org/10.1016/j.jacceco.2010.09.001.
- Dechow, P.M., and I.D. Dichev. 2002. The quality of accruals and earnings: The role of accrual estimation errors. *The Accounting Review* 77: 35–59.
- Dechow, P.M., W. Ge, C.R. Larson, and R.G. Sloan. 2011. Predicting material accounting misstatements. Contemporary Accounting Research 28(1): 17 - 82.https://doi.org/10.1111/j.1911-3846.2010.01041.x.
- Dechow, P.M., R.G. Sloan, and A. Sweeney. 1995. Detecting earnings management. *The Accounting Review* 70: 193–225.
- DeFond, M.L., and J. Jiambalvo. 1994. Debt covenant violation and manipulation of accruals. Journal of Accounting and Economics 17(1-2): 145–176. https://doi.org/10.1016/0165-4101(94)90008-6.
- Du, K., S. Huddart, L. Xue, and Y. Zhang. 2020. Using a hidden Markov model to measure earnings quality. *Journal of Accounting and Economics* 69(2-3): 1–27. https://doi.org/10.1016/j.jacceco.2019.101281.
- Dye, R.A., and S.S. Sridhar. 2004. Reliability-relevance trade-offs and the efficiency of aggregation. Journal of Accounting Research 42(1): 51–88. https://doi.org/10.1111/j.1475-679X.2004.00129.x.
- Einhorn, E., and A. Ziv. 2020. Biased voluntary disclosure. *Review of Accounting Studies* 17(2): 420–442. https://doi.org/10.1007/s11142-011-9177-0.
- Fernández-Val, I., and M. Weidner. 2016. Individual and time effects in nonlinear panel models with large N, T. Journal of Econometrics 192(1): 291–312. https://doi.org/10.1016/j.jeconom.2015.12.014.
- Fischer, P.E., and R.E. Verrecchia. 2000. Reporting bias. *The Accounting Review* 75(2): 229–245. https://doi.org/10.2308/accr.2000.75.2.229.
- Frankel, R.M., and Y. Sun. 2018. Predicting accruals based on cash-flow properties. *The Accounting Review* 93(5): 165–186. https://doi.org/10.2308/accr-52001.
- Gelman, A., and D.B. Rubin. 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7(4): 457–472. https://doi.org/10.1214/ss/1177011136.
- Gelman, A., H.S. Stern, J.B. Carlin, D.B. Dunson, A. Vehtari, and D.B. Rubin. 2013. Bayesian data analysis. Boca Raton: Chapman and Hall/CRC.
- Gerakos, J., and A. Kovrijnykh. 2013. Performance shocks and misreporting. *Journal of Accounting and Economics* 56(1): 57–72. https://doi.org/10.1016/j.jacceco.2013.04.001.
- Glaeser, S., and W.R. Guay. 2017. Identification and generalizability in accounting research: a discussion of Christensen, Floyd, Liu, and Maffett (2017). *Journal of Accounting and Economics* 64(2): 305–312. https://doi.org/10.1016/j.jacceco.2017.08.003.
- Healy, P.M., and J.M. Wahlen. 1999. A review of the earnings management literature and its implications for standard setting. Accounting Horizons 13(4): 365–383. https://doi.org/10.2308/acch.1999.13.4.365.
- Hennes, K.M., A.J. Leone, and B.P. Miller. 2008. The importance of distinguishing errors from irregularities in restatement research: The case of restatements and CEO/CFO turnover. *The Accounting Review* 83(6): 1487–1519. https://doi.org/10.2308/accr.2008.83.6.1487.
- Hribar, P., and D.W. Collins. 2002. Errors in estimating accruals: implications for empirical research. *Journal of Accounting Research* 40(1): 105–134. https://doi.org/10.1111/1475-679X.00041.
- Hribar, P., and D.C. Nichols. 2007. The use of unsigned earnings quality measures in tests of earnings management. *Journal of Accounting Research* 45(5): 1017–1053. https://doi.org/10.1111/j.1475-679X.2007.00259.x.
- Huang, X., S.H. Teoh, and Y. Zhang. 2014. Tone management. *The Accounting Review* 89(3): 1083–1113. https://doi.org/10.2308/accr-50684.
- Jaynes, E.T. 2003. Probability Theory. The Logic of Science. Cambridge University Press.
- Johnstone, D. 2018. Accounting theory as a Bayesian discipline. Foundations and Trends in Accounting 13(1-2): 1–266. https://doi.org/10.1561/1400000056.

- Jones, J.J. 1991. Earnings management during import relief investigations. Journal of Accounting Research 29: 193–228. https://doi.org/10.2307/2491047.
- Kothari, S.P., A.J. Leone, and C.E. Wasley. 2005. Performance matched discretionary accrual measures. Journal of Accounting and Economics 39(1): 163–197. https://doi.org/10.1016/j.jacceco.2004.11.002.
- Kothari, S.P., N. Mizik, and S. Roychowdhury. 2015. Managing for the moment: the role of earnings management via real activities versus accruals in SEO valuation. *The Accounting Review* 91(2): 559– 586. https://doi.org/10.2308/accr-51153.
- Leone, A.J., M. Minutti-Meza, and C.E. Wasley. 2019. Influential observations and inference in accounting research. *The Accounting Review* 94(6): 337–364. https://doi.org/10.2308/accr-52396.
- Leuz, C., D. Nanda, and P.D. Wysocki. 2003. Earnings management and investor protection: an international comparison. *Journal of Financial Economics* 69(3): 505–527. https://doi.org/10.1016/S0304-405X(03)00121-1.
- Lewandowski, D., D. Kurowicka, and H. Joe. 2009. Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis* 100(9): 1989–2001. https://doi.org/10.1016/j.jmva.2009.04.008.
- Lewis, C.M. 2012. Risk modeling at the SEC: The accounting quality model. https://www.sec.gov/news/speech/2012-spch121312cmlhtm (accessed: 01/20/2020).
- McClure, C., and A.A. Zakolyukina. 2021. Non-GAAP reporting and investment. Working Paper. Available at: https://ssrn.com/abstract\_id=35 07069.
- McElreath, R. 2020. *Statistical rethinking: a Bayesian course with examples in R and Stan*. Boca Raton: Chapman and Hall/CRC.
- McMullin, J.L., and B. Schonberger. 2020. Entropy-balanced accruals. *Review of Accounting Studies* 25: 84–119. https://doi.org/10.1007/s11142-019-09525-9.
- McNichols, M.F. 2002. Discussion of the quality of accruals and earnings: the role of accrual estimation errors. *The Accounting Review* 77(s-1): 61–69. https://doi.org/10.2308/accr.2002.77.s-1.61.
- Nikolaev, V.V. 2018. Identifying accounting quality. Working Paper. Available at: https://ssrn.com/abstract\_id=24 84958.
- Owens, E.L., J.S. Wu, and J. Zimmerman. 2017. Idiosyncratic shocks to firm underlying economics and abnormal accruals. *The Accounting Review* 92(2): 183–219. https://doi.org/10.2308/accr-51523.
- Piironen, J., and A. Vehtari. 2017. Comparison of Bayesian predictive methods for model selection. Statistics and Computing 27(3): 711–735. https://doi.org/10.1007/s11222-016-9649-y.
- Roychowdhury, S. 2006. Earnings management through real activities manipulation. Journal of Accounting and Economics 42(3): 335–370. https://doi.org/10.1016/j.jacceco.2006.01.002.
- Sletten, E., Y. Ertimur, J. Sunder, and J. Weber. 2018. When and why do IPO firms manage earnings? *Review of Accounting Studies* 23: 872–906. https://doi.org/10.1007/s11142-012-9194-7.
- Stan Development Team. 2018. Stan modeling language users guide and reference manual, version 2.18.0. http://mc-stan.org.
- Steel, M.F.J. 2020. Model averaging and its use in economics. Journal of Economic Literature 58(3): 644–719. https://doi.org/10.1257/jel.20191385.
- Syverson, C. 2011. What determines productivity? *Journal of Economic Literature* 49(2): 326–65. https://doi.org/10.1257/jel.49.2.326.
- van de Schoot, R., S. Depaoli, R. King, B. Kramer, K. Märtens, M.G. Tadesse, M. Vannucci, A. Gelman, D. Veen, J. Willemsen, and C. Yau. 2021. Bayesian statistics and modelling. *Nature Reviews Methods Primers* 1(1): 1–26. https://doi.org/10.1038/s43586-020-00001-2.
- Yao, Y., A. Vehtari, D. Simpson, and A. Gelman. 2018. Using stacking to average Bayesian predictive distributions (with discussion). *Bayesian Analysis* 13(3): 917–1003. https://doi.org/10.1214/17-BA1091.
- Zhou, F. 2021. Disclosure dynamics and investor learning. *Management Science* 67(6): 3429–3446. https://doi.org/10.1287/mnsc.2020.3638.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.