Macroeconometrics with High-dimensional Data

A Bayesian Model Averaging Approach

Stefan Zeugner

ECARES – Solvay Business School of Economics and Management

Université libre de Bruxelles

Dissertation présentée en vue de l'obtention du grade

Docteur en Sciences économiques et de gestion

2012 September

Supervisor:	Robert Kollmann
Jury members:	Christine De Mol
	Gernot Doppelhofer
	Domenico Giannone
	David Veredas

I am grateful to the Université libre de Bruxelles for financing this thesis via a MiniARC scholarship

To Katharina

Acknowledgements

It has taken me almost five years to produce 140 pages of thesis. In total, I spent more than three years of these in exploring solutions to the issues in this thesis that did not work out in the end. During that period, hardly a day did pass without me working on, thinking of, or having bad conscience due to, my thesis – only briefly interrupted by quick spells of euphoria after having finally found a result. But when things seemed the worst, they suddenly brightened up, and accelerated into a high-intensity final sprint to see hard work and deep engagement finally paying off.

Yet it truly was the engagement of others that made this thesis possible: I am deeply grateful to the Université libre de Bruxelles for providing me with the scholarship that financed my full-time devotion to doctoral work. My gratitude also belongs to my co-authors Robert K., Martin F., and Martin W. – not only for their stimulating and engaged work on the articles that form my thesis chapters,[1] but also for their patience. In both aspects, Robert deserves particular mention in his role as thesis supervisor. I also thank Martin F. for sharing countless days with me working on innumerable lines of computer code.[2] The members of the thesis jury, Christine D., Gernot D., Domenico G., and David V. deserve my gratefulness for their constructive feedback that greatly helped to improve and focus the thesis chapters. Moreover, I owe my results to fruitful discussions with colleagues, seminar and conference participants around the world (who are so many that they are mentioned in the article versions of the individual chapters). All of them have helped me in focusing my thesis on, and contributing to, the field of model averaging in macroeconometrics.

I would also thank those who made my life in Brussels livable: The helpful and uncomplicated staff and faculty in Ecares, as well as my fellows in the 'doctoral situation', who helped me tackle the zeniths and nadirs of my thesis career. Chatting with them (and most with Claudio K.) has helped me come around difficult corners in thesis- and thesis-dominated life several times. I am also indebted to my friends in Brussels and 'VWLers' in Vienna who have provided me with an important hook to life beyond the thesis.

Finally, I am obliged to my parents and family whom I depended on for emotional support during difficult times. But, first and foremost, I would like to thank my wife Katharina, who supported me throughout this long spell of research, setbacks and successes while bearing the brunt of their impact on my and our life.

g'scheng is!

[1] `http://www.zeugner.eu/studies/thesis`

[2] A finite amount of these code lines survived to create the 'BMS' software package this thesis is based on, and that Martin and me provide on `bms.zeugner.eu`.

Contents

List of Figures

List of Tables

Thesis Summary

This thesis is composed of four essays on theory and applications of Bayesian Model Averaging (BMA), based on joint work with Martin Feldkircher (first and second chapter), my supervisor Robert Kollmann (third chapter), Martin Wagner (fourth chapter).

BMA addresses the econometric problem of selecting the few out of many potential explanatory variables which should be included in the 'true' model for an outcome variable. BMA tackles such model uncertainty by assigning prior model probabilities to each variable combination and updating them to data-based posterior model probabilities. The purpose of the latter is twofold: First, they allow for inference about the true model and its constituent variables. Second, they enable the construction of aggregate posterior statistics as a weighted average of individual model estimators. The soundness and robustness of BMA results, however, hinges on the set-up of its priors, which therefore plays a crucial part in all chapters of this thesis.

The *first chapter* addresses the robustness problems arising from a default prior framework that is widely used in BMA. For computing the model-specific results that provide the basis for model averaging, most BMA studies rely on natural-conjugate Bayesian regression with Zellner's g (Zellner, 1986). While this framework has vital computational advantages, it requires the choice of a 'shrinkage' hyperparameter that expresses how tightly coefficient priors are centered at zero. Since this parameter can exert striking influence on posterior results, several articles have suggested fixing it to a 'default' scalar (e.g., Foster and George, 1994, Fernandez et al., 2001) – a recommendation that is widely followed in applied research. However, empirical results under this default prior framework have been shown to suffer from a lack of robustness (Eicher et al. 2008, Ciccone and Jarociński, 2011).

In response, we analytically demonstrate that the value of a fixed shrinkage parameter directly affects the concentration of model weights, which in turn is crucial for robustness. Under this viewpoint, the value of the shrinkage parameter should theoretically vary with the error variance, which is unknown. We therefore propose to forgo fixed shrinkage altogether and instead focus on an endogenous shrinkage parameter that adjusts to the data quality of individual models. To this end, we focus on a hyper-prior such as the one proposed by Liang et al. (2008). A substantial part of the chapter is dedicated to algebraically refining the incomplete analytical setup of this hyper-prior in order to render its numerical implementation feasible. Moreover, we demonstrate some properties of posterior results under a hyper-prior and establish a relationship with common OLS diagnostics. Finally, simulations and an application to the empirics of economic growth demonstrate the merits of the hyper-prior both in terms of prior robustness and predictive performance.

The *second chapter* discusses the shrinkage parameter in the light of a particular robustness problem highlighted by Ciccone and Jarociński (2011). Their study demonstrates

that the importance BMA attributes to potential determinants of economic growth varies tremendously over different revisions of growth data. The authors conclude that default priors appear too sensible for this strand of growth empirics. In response, we show that the found instability owes mainly to the fixed default prior setting relied on in this exercise. We demonstrate that applying a hyper-prior to the original data yields a marked reduction in the instability of posterior results. The improvements in robustness come at a price, though: Results from the hyper-prior set-up show that the conclusions to be drawn from such data are weaker than what has been implied by previous studies (e.g., Sala-i-Martin et al., 2004).

The financial crisis of 2007-09 led to the consideration of balance sheet leverage as a potential determinant of real economic activity (e.g., Kollmann et al., 2011). The *third chapter* therefore explores whether there is empirical evidence for a link between leverage and real activity in the US. To this end, we assess forecasts for real activity based on leverage of the financial sector, households and non-financial businesses, while controlling for a large set of macro/financial variables commonly used by forecasters. We apply BMA to compare the predictive performance of leverage to that of individual controls. In addition to in-sample results, we construct out-of-sample forecasts, requiring the estimation of over two billion models. Moreover, we adjust model priors to avoid coincidental significance of alternative predictors due to their groupwise consideration. The econometric findings on the importance of leverage are complemented and corroborated by parsimonious least squares models.

The results document that leverage is negatively related to the future growth of real activity, and positively linked to the conditional volatility of future real activity and of equity returns. The joint information in sectoral leverage series is more relevant for predicting future real activity than the information contained in any individual leverage series. Moreover, we find that the predictive power of leverage is roughly comparable to that of macro and financial variables. Nonetheless, leverage information would not have allowed to predict the 'Great Recession' of 2008-2009 any better than conventional macro/financial predictors.

The *fourth chapter* assesses the empirical importance of recently proposed 'financial' predictors for US business cycles, while controlling for a wide range of established potential explanatory variables (such as interest rates, industrial production, etc.). Addressing this proto-typical econometric exercise with standard BMA methods runs two major risks: First, in a small sample with many controls (potentially more variables than observations), the results can be very sensitive to the averaging set-up (even under elaborate priors). Second, under small sample with many controls, a researcher faces a trade-off between risking over-fitting on the one hand, and omitted variable bias for the variables in question. Such a situation usally requires elaborate – and there specification of controls risks to err either on the side of over-fitting

To address both issues, we propose Principal-Component-augmented Model Averaging. This approach subsumes the control variables under a handful of factors, allowing to concentrate model averaging on the small set of remaining 'focus' variables (financial indicators, in our case) as well as over the number of included factors. In contrast to established similar methods, this setup provides for an endogenous updating of the relative importance between 'focus' variables and controls, and thus softens the over-fitting vs. omitted variable bias trade-off. We demonstrate that our approach is asymptotically equivalent to outright model averaging but improves robustness in small samples.

For implementation, we provide a framework with carefully adjusted 'default' priors, in particular on the number of factors to be included. We subsequently assess the link between financial predictors and real activity both with Bayesian and Frequentist Principal-Component-augmented model averaging. We find that results are indeed more robust in comparison with outright BMA. In addition, we show that the predictive performance of our approach is superior to outright BMA and closer to that of factor models (cf. Stock and Watson, 2006). On the empirical side, we find that evidence for a linear relationship between our financial variables and economy activity since 2000 is limited at most.

1

Benchmark Priors Revisited: On Adaptive Shrinkage and the Supermodel Effect in Bayesian Model Averaging

Co-authored with:
Martin Feldkircher, *Oesterreichische Nationalbank*

Abstract
The default g-priors predominant in Bayesian Model Averaging tend to over-concentrate posterior mass on a tiny set of models – a feature we denote as 'supermodel effect'. To address it, we propose a 'hyper-g' prior specification, whose data-dependent shrinkage adapts posterior model distributions to data quality. We demonstrate the asymptotic consistency of the hyper-g prior, and its interpretation as a goodness-of-fit indicator. Moreover, we highlight the similarities between hyper-g and 'Empirical Bayes' priors, and introduce closed-form expressions essential to computationally feasibility. The robustness of the hyper-g prior is demonstrated via simulation analysis, and by comparing four vintages of economic growth data.

Keywords: Bayesian model averaging, hyper-g prior, shrinkage factor, Zellner's g prior, model uncertainty.

JEL Classifications: C11, C15, C21, C52, O50.

1.1 Introduction

Statistical inference that neglects model uncertainty leads to overstated confidence in statistical estimates, as has been amply demonstrated since the seminal contributions by Raftery (1995) and Hoeting et al. (1999). *Bayesian Model Averaging* (BMA) tackles such model uncertainty directly by basing inference on a weighted average of all potential covariate combinations, or 'models'. In a Bayesian setting, these weights arise naturally as posterior model probabilities that correspond to the classical likelihood concept. Relying on this framework, numerous authors (e.g., Raftery, 1995; Fernández et al., 2001a; Liang et al., 2008) have demonstrated that BMA outperforms other strategies in terms of predictive ability. Virtually all of them have so far concentrated on linear models with model-specific inference based on the 'Normal-Gamma' coefficient prior with *Zellner's g* (Zellner, 1986). This prior structure has proven popular in BMA, since it leads to simple closed-form expressions of posterior statistics and because it reduces prior elicitation to the choice of a single scalar hyperparameter g. This *shrinkage* parameter determines how far a model's coefficients are shrunk toward zero: High values for g are meant to embody weak prior knowledge and correspond to model estimators that are close to least squares results. In contrast, low values for g imply less reliance on the data and posterior coefficients closer to their prior values (zero). Crucially, the parameter g shapes the weights of models in BMA. The exact specification of g is subject to intense debate,[1] but the use of a constant hyperparameter g as such has been less frequently criticized.

This paper demonstrates that the practical advantages of Zellner's g come at a serious cost: g exerts non-negligible influence on posterior inference since it governs how posterior mass is spread over the models. For given data, high values of g concentrate posterior mass on few models, which runs the risk of overfitting. In contrast, small values of g spread posterior model probabilities (PMPs) more evenly among the models (irrespective of model sizes and size penalty terms). Posterior statistics, in particular PMPs and the covariates' posterior inclusion probabilities are thus notoriously sensitive to the value of the g prior. In other words, the researcher's prior on g determines how much posterior mass is attributed to the few best-performing models – regardless of whether these have been generating the data.[2] In this paper, we establish the conditions for this g-induced concentration of posterior mass (which we will henceforth refer to as the *supermodel effect*). While crucial in terms of prior sensitivity, this feature went more or less unnoticed in previous simulation studies that focused on 'asymptotic consistency': In order to uncover a single 'true' model in Monte Carlo simulations with a weak noise component, such exercises profit from a large value for g that induces posterior mass to concentrate on the best-performing model.

The proper Bayesian approach to address this problem is to introduce a non-degenerate *hyperprior distribution on* g, and thus 'let the data choose'. Such a flexible prior allows for shrinking the estimated coefficients more toward zero under models with a large noise component,[3] i.e., inducing data-dependent shrinkage. Only few papers have applied such hyperpriors in BMA so far: Among them are Strachan and van Dijk (2004), Cui and George (2008), Liang et al. (2008), and Ley

[1] The literature on the optimal choice of g (e.g., Liang et al., 2008; Ley and Steel, 2011b; Hoeting et al., 1999; Fernández et al., 2001a; Eicher et al., 2011) has concentrated on two theoretical considerations: First, asymptotic consistency, i.e. the choice of g such that BMA asymptotically uncovers 'the true model'. However, from a Bayesian viewpoint, many models might be 'true', in the sense that they are generating the data examined. Second, the specification of g was studied in terms of its virtues as a model size penalty term to favor parsimonious models. From a Bayesian perspective, however, such preferences on parameter size should rather be considered in the formulation of model priors, which constitute a crucial component of BMA.

[2] Note that this effect not only poses a risk for model averaging, but also for model selection. For instance, a too large value for g might lead to underestimating the model uncertainty related to selecting a particular model. Moreover, it might increase the risk of selecting the 'wrong' model, cf. discussion in section 1.3.

[3] I.e. by up- or downweighting the prior beliefs on coefficients β.

and Steel (2011b). These authors have proposed various hyperprior specifications in response to theoretical issues other than the problem described above,[1] but none of them has focused on the expected properties of flexible priors under small samples. Most of these hyperpriors have in common that their respective statistics are not available in closed form, thus forcing the researcher to resort to MCMC sampling. The contribution by Liang et al. (2008) differs in providing a closed-form solution – which is so computationally demanding that they implement it via analytical approximations. In this paper, we provide algebraic transformations of the Liang et al. (2008) prior that allow for a sound and accurate numerical application at minimal computational cost. We use this augmented hyper-g prior to show that the model-specific advantages of the hyper-g prior also extend to inference under model uncertainty, as it is not exposed to the supermodel effect a priori. The hyper-g prior adjusts the distribution of posterior mass in dependence of the information provided by the data. Thus if noise dominates the data, PMPs under the hyper-g prior will be distributed more evenly, whereas in the case of minor noise, posterior mass will be concentrated even more than under fixed settings with large values for g.

Based on the above considerations, the contribution of our paper is fivefold: First, we show that fixed coefficient priors may introduce too much or too few shrinkage into individual models, but also have an even stronger impact on the concentration of model probabilities in BMA.[2] We demonstrate the supermodel effect analytically, in a simulation exercise and an empirical application. Second, we propose a particular prior framework that reconciles the Liang et al. (2008) hyperprior with asymptotic consistency, and provide closed-form representations for important posterior quantities. Third, we show further properties of the hyper-g prior: We demonstrate how its posterior statistics are analytically related to the strand of 'Empirical Bayes' priors, and why their results hardly differ under data-sets with a weak noise component. Moreover, we show the relationship of the hyper-g to the familiar OLS F-statistic as a measure of goodness-of-fit. Fourth, we show how the superior robustness of the hyper-g prior addresses the supermodel effect under a simulation exercise with both a simple and a complex data-generating process. Our results show that under noisy data the hyper-g prior dilutes posterior mass among models whereas the popular fixed priors incorrectly favor one (wrong) model. We examine the forecasting properties of various settings for g by means of a simulation study, which points to superior predictive ability of the hyper-g prior under varying signal-to-noise ratios. Finally, our empirical exercise illustrates this advantage in the context of growth regressions. We show how far the BMA parameter instability over revisions of growth data (as found by Ciccone and Jarociński (2010)) is due the supermodel effect, and that the hyper-g prior can reduce it to a great extent.

The remainder of this study is organized as follows: the next section briefly reiterates the concept of Bayesian model averaging and its most popular prior settings. Section 1.3 sketches the reasons underlying the supermodel effect and provides formal conditions for its presence. Section 1.4 introduces the hyper-g prior, outlines further posterior statistics and properties, and introduces an implementation strategy of practical relevance. Section 1.5 presents a simulation exercise that examines the supermodel effect inherent to traditional priors and highlights the predictive performance of flexible priors. The following section demonstrates the sensitivity of posterior results to the choice of g by means of an empirical application to a prominent growth data set. Section 1.7 concludes the paper.

[1] See footnote 1 for a discussion of these papers' motivations.

[2] The detrimental effect of fixed priors on robustness and the advantage of hyperpriors have not only been noted in Bayesian regression-type models, but also other Bayesian frameworks (see, e.g., Giannone et al. (2012) for a similar motivation in the case of Bayesian VARs). The impact on model averaging, however, is novel, to the best of our knowledge.

1.2 Bayesian Model Averaging under Zellner's g prior

This section summarizes the popular set-up of Bayesian model averaging (BMA) under the natural conjugate framework with Zellner's g prior and reviews the prior settings that have resurfaced most often in the literature so far. Consider the canonical regression problem of sample size N with the dependent variable in the $N \times 1$ vector y, X_s an $N \times k_s$ design matrix of covariates, and ε an N-dimensional vector of residuals in the following, linear model M_s:

$$y = \mathbf{1}\alpha_s + X_s\beta_s + \varepsilon$$

Here α_s denotes the (scalar) intercept, and β_s the the $k_s \times 1$-vector of unrestricted regression coefficients. The residuals are assumed to be normally IID with variance σ^2, i.e. $\varepsilon \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. Note that X_s can be assumed to be centered ($X_s'\mathbf{1} = \mathbf{0}$) without loss of generality, as this will only affect the posterior distribution of the constant α_s. Bayesian Model Averaging deals with uncertainty about the model M_s by drawing on the model-specific inference presented above. In the generic linear BMA problem, model uncertainty focuses on the choice of covariates X_s, which may be drawn from a set of K potential regressors. This induces 2^K unique covariate combinations, as represented by the model candidate space $\mathcal{M} = \{M_1, M_2, \dots, M_{2^K}\}$ (cf. Hoeting et al., 1999, for a more detailed account).

The Bayesian framework calls for specifying a prior distribution on the model's parameters α, β_s, and σ^2. The bulk of the BMA literature (e.g., Raftery, 1995; Chipman et al., 2001), favors the natural-conjugate approach, or its variant outlined by Fernández et al. (2001a): In order to represent lack of information over constant and variance, place improper priors on constant $p(\alpha) \propto 1$ and variance $p(\sigma) \propto \sigma^{-1}$.[1] The prior on coefficients β is assumed to be normal and potentially allows for model-specific elicitation of prior expected value and coefficient covariance. However, the explicit formulation of these hyperparameters is difficult to perform given the many combinations possible in model selection problems. Virtually all linear BMA applications have thus opted for a common uninformative prior centered at zero, with the variance structure given by *Zellner's g prior* (Zellner, 1986):

$$\beta_s|\sigma^2, M_s, g \;\sim\; N\big(0, \sigma^2 g (X_s'X_s)^{-1}\big)$$

This prior assumes the coefficient covariance to be proportional to the posterior covariance expression $(X_s'X_s)^{-1}$ that arises from the sample, with the scalar g determining how certain the researcher is in centering the prior coefficient distribution at zero. Apart from offering computational efficiency, Zellner's g thus reduces the elicitation of the covariance structure to choosing the scalar g. Employing Bayes' theorem via

$$p(\beta_s|y, X_s, M_s) = \int_0^\infty p(\beta_s|M_s, y, X_s, \sigma^2)\, dp(\sigma^2)$$

yields the posterior coefficient distribution as k_s-variate student-t[2] with expected value $E(\beta_s|y, X, M_s, g) = \frac{g}{1+g}\hat{\beta}_s$, where $\hat{\beta}_s$ denotes the standard OLS estimator for M_s.[3] Note that the posterior expected

[1] Note that the specification for α and σ departs from earlier tradition which typically elicited proper priors for the two parameters. However, both choices do not affect the crucial posterior statistics: The improper prior on the constant allows for an easy disentanglement of the constant with respect to the other coefficients. In contrast to the traditional Gamma-priors, the improper prior on σ offers the advantage of being invariant under scale transformations (Fernández et al., 2001a, p. 391)

[2] Note that this posterior distribution requires $N > 2$.

[3] The posterior variance of β_s is $\frac{\breve{y}'\breve{y}}{N-2}\left(1 - \frac{g}{1+g}R_s^2\right)\frac{g}{1+g}(X_s'X_s)^{-1}$, where $\breve{y} = y - \mathbf{1}\bar{y}$ denotes the centered response vector.

value is a convex combination of its OLS estimator and the prior expected value (zero) weighted by the shrinkage factor $\frac{g}{1+g}$. The larger the shrinkage factor, the more importance is attributed to sample data rather than to prior information.

For its use in BMA, the main advantage of Zellner's g is that it yields a closed-form expression for the marginal likelihood of M_s:[1]

$$p(y|M_s, g) \propto (1+g)^{-\frac{k_s}{2}} \left(1 - \tfrac{g}{1+g} R_s^2\right)^{-\frac{N-1}{2}} \tag{1.1}$$

with k_s denoting the number of covariates included in model M_s and R_s^2 its OLS R-squared. This marginal likelihood is crucial in determining the posterior model probability that arises from Bayes' theorem $p(M_s|y,X,g) \propto p(y|M_s,X,g)p(M_s)$ as an update of a prior model probability $p(M_s)$:

$$p(M_s|y,X,g) = \frac{p(y|M_s,X,g)p(M_s)}{\sum_{j=1}^{2^K} p(y|M_j,X,g)p(M_j)} \tag{1.2}$$

Multiplied with a normalization constant, these posterior model probabilities serve as model weights in Bayesian model averaging. In this vein, the marginal posterior distribution of any statistic Θ may be obtained as a mixture over posterior model probabilities:[2]

$$p(\Theta|y,X,g) = \sum_{j=1}^{2^K} p(\Theta|y,X,M_j)p(M_j|y,X,g)$$

This property is particularly useful in computing the posterior moments of the coefficient vector β as a weighted average over all models.[3] Likewise, posterior inclusion probabilities (PIPs), used for assessing the importance of single covariates, are obtained as the sum of probabilities for all models in which the covariate is included.

In view of equation (1.2), BMA inference hinges on posterior model probabilities and, in turn, on two important prior specifications: the model priors $p(M_s)$ and Zellner's g prior for the coefficients: The Bayesian framework calls for defining prior model probabilities $p(M_j)$ for all models contained in the model space $j \in \{1, 2, \ldots, 2^K\}$. While advocates of purism may call for subjective prior specification of $p(M_s)$, the number of model candidates renders this virtually infeasible. Consequently, most authors have relied on the uniform model prior $p(M_s) = 2^{-K}$, whereas several (Brown et al., 1998; Sala-i-Martin et al., 2004; Ley and Steel, 2009) have proposed to specify model priors in dependence of average model size k_s, typically in such a way that prior elicitation is reduced to choosing the prior expected model size.

In addition to model priors, the choice of Zellner's g prior crucially affects marginal likelihoods $p(y|M_s,X,g)$ and thus PMPs. Its discussion so far has focused on two considerations:

[1] Note that although the term $((y-\bar{y})'(y-\bar{y}))^{-\frac{N-1}{2}}$ is constant over models, it is frequently included in the marginal likelihood expression, such as in Fernández et al. (2001a) – while others, such as Liang et al. (2008) omit it.

[2] Note that the concept lined out in equation (1.2), and its implication for posterior statistics, is of course not limited to linear models alone. However, the bulk of the empirical BMA literature focuses on linear models using the g-prior described in this section. The purpose of this paper is to discuss the g-prior's consequences for linear models, and thus it refrains from discussing the implications of similar priors for more complex models.

[3] Note that we have retained the improper priors for α and σ as common to all models.

- Consistency: The choice of g such that posterior model probabilities asymptotically uncover 'the true model' M_T, i.e. $p(M_T|y, X, g) \to 1$ as $N \to \infty$
- The importance of g as a penalty term enforcing parameter parsimony (the factor $(1+g)^{\frac{k_j-k_s}{2}}$ in (1.2))

Both issues have been reviewed by Fernández et al. (2001a): With respect to consistency, they prove that a choice of $g = w(N)$ such that $\lim_{N\to\infty} w(N) = \infty$ and $\lim_{N\to\infty} \frac{w'(N)}{w(N)} = 0$ ensures consistency as it was mentioned above. Still, consistency leaves open the exact specification of g. Over the course of more than a decade, various 'automatic' or 'default' specifications have been put forward (e.g., Fernández et al., 2001a; Eicher et al., 2011) that typically specify g according to sample size N. Note that the bulk of the literature concentrates on priors *fixing* g in such a way that the penalty term $(1+g)^{-\frac{k_s}{2}}$ in (1.2) asymptotically mimics popular information criteria.[1] In particular, two settings for g resurface steadily in the literature: The Unit Information Prior (g-UIP) corresponds to $g = N$. Through its dependence on sample size it is a consistent prior and draws on the notion that the 'amount of information' contained in the prior equals the amount of information in one observation (Kass and Wasserman, 1995). Fernández et al. (2001a, p.424) demonstrate that as $N \to \infty$ the log of the Bayes factor for two models approaches the ratio of their Bayesian information criteria. Secondly, setting $g = K^2$ (g-RIC) calibrates the posterior model probability to asymptotically match the risk inflation criterion proposed by Foster and George (1994). Based on an extensive study of various specifications for g, Fernández et al. (2001a) recommend the 'benchmark' prior which bridges the g-UIP and the g-RIC by setting $g = \max(N, K^2)$.

1.3 The Supermodel Effect

The advantages of Zellner's g prior have fostered its widespread use in BMA, even though its exact specification is still subject to debate (as highlighted by the previous section). In general, g determines the tightness of the prior distribution on coefficients β around their prior expected value zero: Large g implies a diffuse prior distribution, i.e. the researcher is very uncertain about the prior expected value and relies heavily on the data. Small g means a prior that is more tightly centered at zero,[2] and leaves less scope to the data to determine the coefficients. In this sense, it is evident why most g specification schemes aim to set this hyperparameter according to data quality: A high signal-to-noise ratio warrants strong reliance on the data and thus a high g, while an important noise component should be met with a low g.

In practice, the choice of g can have considerable consequences for the robustness of BMA results. For instance, with very noisy data, a large g could attribute too much weight to results that are mainly driven by a particular realization of the error term. Such a case may lead to situations as in Ciccone and Jarociński (2010), who show that BMA under the 'benchmark' specification from Fernández et al. (2001a) produces results that differ strikingly over small revisions to the

[1] Information criteria are a widely used approach for model selection and are conceptually similar to the marginal likelihoods that arise from incorporating model uncertainty in a Bayesian framework (Indeed, several popular information criteria (IC) can be derived from such a setting, such as Schwarz (1978)). Drawing on this similarity, 'frequentist' model averaging techniques rely on IC in order to obtain 'posterior' model weights (see Claeskens and Hjort, 2008, for an overview). Due to their numerical connections, empirical results under IC-based linear model averaging are usually quite similar to BMA with g-priors that mimick IC. In this paper, we therefore forgo the explicit discussion of IC-based model averaging techniques, and concentrate on their Bayesian analogues.

[2] In general, small g implies centering at the prior (expected values of coefficients). Here, we follow the bulk of the literature in presupposing coefficient priors to be centered at zero.

response variable. Such robustness problems can arise from a too loose g that focuses posterior model mass on too few 'supermodels'. In this section, we demonstrate that g is positively linked with the concentration of posterior model probabilities – and that this *supermodel effect* matters to empirical practice. A look at the role of the shrinkage factor $\frac{g}{1+g}$ in the marginal likelihood from (1.1) provides some intuition:

$$p(y|M_s, g) \propto \underbrace{\left(1 - \tfrac{g}{1+g}\right)^{\frac{k_s}{2}}}_{A} \underbrace{\left(1 - \tfrac{g}{1+g} R_s^2\right)^{-\frac{N-1}{2}}}_{B} \tag{1.3}$$

The shrinkage factor $\frac{g}{1+g}$ affects marginal likelihood (and thus posterior model probability) via a size penalty term (A) and a model fit term (B). The term (A) shapes the distribution of posterior mass between different model sizes, while term (B) determines the concentration of PMP within models of the same size k_s. Among models of the same size k_s, larger $\frac{g}{1+g}$ will increase the relative posterior weight of the models with the largest R_s^2. The term (B) thus implies a direct positive link between the shrinkage factor and relative PMP concentration among models of the same size k_s.

The term (A) has stimulated most of the debate on the g prior through its virtues as a size penalty term that could mimic well-founded information criteria. But from a Bayesian viewpoint, size penalty represents prior preferences on model size that should be fused into the formulation of the model prior rather than a coefficient prior. Instead, one should consider that the term (A) can reinforce the link between g and PMP concentration: Increasing the shrinkage factor $\frac{g}{1+g}$ strengthens the size penalty and skews the posterior model size distribution to smaller (more parsimonious) models. When most of the posterior mass focuses on models below the size of $K/2$ (which typically applies to empirical exercises), strengthening the size penalty means concentrating mass on model sizes that comprise fewer models to choose from. In this sense, the term (A) contributes to a positive link between g and PMP concentration.

The interplay of posterior model size distribution (from term (A)) and relative PMP concentration among models of a given size (from term (B)) is not straightforward, and can depend on data set characteristics and the particular value of g. Proposition 1 therefore formally pins down the conditions for the supermodel effect:

PROPOSITION 1 *For linear BMA with a fixed common Zellner's g prior, a given realization of* (y, X)*, and any model prior that does not depend of g, the following holds: The cumulative posterior probability of the best r models have a non-negative derivative with respect to g if* $E_r(k|y, X) + \eta < E(k|y, X)$*, where* $E(k|y, X)$ *represents posterior model size and* $E_r(k|y, X)$ *expected posterior model size of the best r models; with* $\eta > 0$ *vanishing as* $N \to \infty$ *or* $g \to \infty$*.*

Thus an increase in g will increase the concentration of PMP on the most important models (by PMP), as long as their average model size is somewhat smaller than the overall posterior model size. Conversely, the most important models could only lose PMP with increasing g if their model sizes are relatively large. In any case, the cumulative PMP of the most important models will increase almost sure over the domain of g.[1] Proposition 1 also holds implications for the special case of model selection (as opposed to model averaging): Increasing g attributes increasingly (perceived) posterior importance to the model which the largest PMP (if its model size is smaller than average model size). However, at some point the conditions of Proposition 1 will not be met locally, and the PMP of this best-performing model will be overtaken by the PMP of a model with fewer parameters.

[1] This result obtains trivially from the fact that the PMP for the null model tends to 1 as g approaches infinity, and that the null model will be the most important model for all g greater than some finite $\bar{g}$.

Thus, model selection with high g values under fixed priors risk selecting a too small model (that is possibly not even nested in the ' true' model), and underestimating the model uncertainty around it.

In order to illustrate the supermodel effect, consider the BMA results from the growth data set as in section 1.6, for different values of g and under uniform model priors.[1] Figure 1.1 (top panel) exhibits the posterior model size as well as the cumulative posterior probability of the best 2,000 models for varying values of g. The results show that increasing g leads to a marked reduction in posterior model size, while increasing the concentration of posterior model probabilities. The rate of increase in the PMP of the best 2,000 models certainly depends on the characteristics of the data set and might vary locally, but the exercise illustrates that the concentration of PMP broadly intensifies with increasing g. In particular, this effect applies to ranges of g that are typically used in the empirics of economic growth.

The previous discussion has shown that term (B) in equation (1.3) directly links g positively with PMP concentration. In contrast, term (A) might have an ambiguous effect, which also underlies the qualifying conditions in Proposition 1. In order to disentangle the two effects in the empirical exercise, we neutralize the term (A) by the following model prior:

$$p(M_s) = \frac{(1+g)^{\frac{k_s}{2}}}{(\sqrt{1+g}+1)^K}$$

Such a model prior exactly cancels the size penalty term (A) in the marginal likelihood (1.3). The impact for differing values of g is shown in Figure 1.1 (bottom panel). When size penalty is neutralized, higher g leads to increased posterior model size, as posterior model mass then concentrates more on the least parsimonious model which have the largest R^2. As expected, the cumulative PMP of the best 2,000 models increases with g, although at a considerably lower pace than under the uniform model priors discussed above. We therefore conclude that for the growth data examined in this paper, the supermodel effect has a sizable impact, and is mainly driven by the size penalty term $(1 - \frac{g}{1+g})^{\frac{k_s}{2}}$.

Note that the supermodel effect not only has implications for the skewness of the PMP distribution, but also for the regressors' posterior inclusion probabilities (PIPs). The term (B) in (1.3) establishes a direct positive link between g and the concentration of PIPs, as larger g leads to more concentration in the relative PMPs for each model size. The term (A) can reinforce this effect: Note that the variance of R^2 for models of size k is by definition (weakly) greater than the variance of R^2 for models of size $K - k$. This implies that relative PMP concentration among models of the same size k is more intense for smaller k. Increasing g therefore tends to skew the relative distribution of the PIPs. In view of Proposition 1 and the above illustration we therefore conclude that g is positively linked not only with the concentration of PMPs, but also of posterior inclusion probabilities.

1.4 Flexible priors: The Hyper-g Prior

The previous section has demonstrated the problems with fixed g priors: First, it might be insensitive to assume a common parameter for all models considered, and second, eliciting the right

[1] Results are from BMA estimations with uniform model priors of the Sala-i-Martin et al. (2004) data set with growth and initial income according to the Penn World Tables revision 6.3. Figure 1.1 displays the result for 24 different values of g, each estimated from MCMC sampling with 200,000 burn-in draws and 2,000,000 subsequent iterations.

parameter value strongly risks over- or under-identification with respect to posterior model and inclusion probabilities. Introducing a flexible hyper-prior on g, in contrast, would allow to update prior beliefs according to data quality, and thus mitigate the risks from choosing a prior value for g.

In principle, the concerns raised in the previous section might be addressed by virtually any flexible hyper-prior setting for g. Based on different motivations, recent contributions have introduced several candidates that could potentially be used to that end (cf. Ley and Steel, 2011b, for an overview). However, most of them do not yield closed form solutions for posterior statistics of interests. Instead, those have to be obtained by numerical sampling techniques, which complicates the computationally demanding task of evaluating models in BMA. Among the proposed candidates g, only the hyper-g prior by Liang et al. (2008) stands out as fairly flexible prior distribution that allows for closed-form posterior statistics. We therefore concentrate on the latter approach to demonstrate the properties of hyper-priors vs. fixed priors.

Liang et al. (2008) introduce two priors motivated on theoretical grounds,[1] among them the closed-form hyper-g prior. While they ingeniously outline the basic features of the hyper-g prior, their posterior expressions involve ratios of hypergeometric functions, which are difficult to evaluate computationally. For feasible computational implementation, the authors thus resort to Laplace approximations – an approach that risks numerical inaccuracies, in particular with respect to the mentioned ratios. This section therefore introduces algebraic transformations of these expressions that yield accurate statistics at low computational cost. Moreover it complements Liang et al. (2008) by establishing additional, common posterior expressions, in particular with respect to second moments of posterior parameters and predictive distributions. Finally we proceed to show some properties of the hyper-g prior, in particular how it be may reconciled with consistency in the sense of Fernández et al. (2001a), its asymptotic equivalence to the Empirical Bayes (EBL) prior, and the relationship between its posterior statistics and the OLS F-statistic.

1.4.1 The hyper-g prior and its posterior statistics

The hyper-g prior for g translates into a Beta prior on the shrinkage factor $\frac{g}{1+g}$ that is common to all models (Liang et al., 2008, p. 415):

$$\frac{g}{1+g} \sim \text{Beta}\left(1, \frac{a}{2} - 1\right)$$

i.e. $\frac{g}{1+g}$ is Beta distributed with $E(\frac{g}{1+g}) = \frac{2}{a}$.[2] The elicitation of g is therefore supplanted by the choice of the hyperparameter $a \in (2, \infty)$: $a = 4$ renders the prior distribution of $\frac{g}{1+g}$ uniform, while moving a close to 2 concentrates the prior mass on the shrinkage factor close to 1. Conversely, any $a > 4$ tends to concentrate prior mass near 0. Liang et al. (2008) therefore omit those cases and concentrate on $a \in (2, 4]$ – a strategy we will follow in this study.

[1] Liang et al. (2008) motivate their paper with two 'paradoxes' that arise with constant g. First, they raise a BMA formulation of 'Bartlett's paradox' stating that if $g \to \infty$ for fixed N and K, the Bayes Factor $B(M_s : M_0)$ of any model with respect to the null model eventually goes to zero. Second, they refer to an 'information paradox' stating that for fixed N and K, if the R-squared of model M_s converges to unity, its Bayes factor with respect to any other fit-wise inferior model does not go to infinity. Moreover, both arguments bite only in the case when N and K are kept constant: Bartlett's paradox in this case may be less relevant as typical specifications for g require it to rise in line with N. The 'information paradox' does not contradict the standard consistency argument that requires the respective Bayes Factor to converge to infinity only when N tends likewise to infinity. See the comment by Zellner (2008) for a more detailed discussion.

[2] Note that this is equivalent to putting the following prior on g: $p(g) = \frac{a-2}{2}(1+g)^{-\frac{a}{2}}$.

An integral representation for the Gaussian hypergeometric function ${}_2F_1(a, b, c, z)$ allows for straightforwardly establishing the model-specific posterior distribution of the shrinkage factor (Abramowitz and Stegun, 1972, p.563).

$$p(\tfrac{g}{1+g}|y, X_s, M_s) = \frac{k_s + a - 2}{2\,{}_2F_1\left(\frac{N-1}{2}, 1, \frac{k_s+a}{2}, R_s^2\right)} \left(1 - \tfrac{g}{1+g}\right)^{\frac{k_s+a-4}{2}} \left(1 - \tfrac{g}{1+g} R_s^2\right)^{-\frac{N-1}{2}} \tag{1.4}$$

The kernel of the shrinkage factor's posterior distribution mimics the expression for marginal model likelihood (1.1).[1] It thus skews the posterior density towards values close to one as the parameters N or R_S^2 increase. In contrast, the shrinkage factor density concentrates closer to zero with increasing parameter size k_s or the hyperparameter a. In other words, the posterior density adapts to a model's marginal likelihood, rewarding good fit with increasing the shrinkage factor towards one (i.e., emulating maximum -likelihood estimates), while punishing parameter size with shrinking posterior estimates towards zero. In this sense, the posterior density of $\frac{g}{1+g}|y, X_s, M_s$ follows a behavior similar to information criteria or the OLS F-statistic.

The integration constant of (1.4) is a Gaussian hypergeometric function, which consequently also turns up in the expression for marginal likelihood of model M_s (cf. Liang et al., 2008, equation (17)):

$$p(y|X_s, M_s) \propto (\breve{y}'\breve{y})^{-\frac{N-1}{2}} \frac{a-2}{k_s+a-2}\, {}_2F_1\left(\tfrac{N-1}{2}, 1, \tfrac{k_s+a}{2}, R_s^2\right) \tag{1.5}$$

While this expression differs from the expression for marginal likelihood under fixed g (1.1), it displays similar behaviour with respect to parameters. Its partial derivatives with respect to parameters correspond to the ones in (1.1). Building on equation (1.5), Liang et al. (2008, equation (19)) proceed by expressing the posterior expected value of the shrinkage factor $\mathrm{E}\left(\frac{g}{1+g}\middle| y, X_s, M_s\right)$ as a ratio of two hypergeometric functions. The expression is primarily relevant for the expected value of the response:[2]

$$E(y|X_s, M_s) = \mathbf{1}E(\alpha_s|X_s, M_s) + \mathrm{E}\left(\tfrac{g}{1+g}\middle| y, X_s, M_s\right) X_s\hat{\beta}_s \tag{1.6}$$

with $\hat{\beta}_s$ denoting the estimated OLS coefficient for model M_s. Equation (1.6) highlights the importance of the shrinkage factor, as the hyper-g prior allows for model-specific, data-adaptive shrinkage as opposed to fixing the value for the shrinkage factor a priori.

The posterior statistics outlined so far suffice for the analysis in Liang et al. (2008). However, fully Bayesian inference requires several more expressions, notably with respect to second moments. Therefore, we introduce in equations (1.7)-(A.2) the moments of the shrinkage factor and the coefficients, as well as the posterior distribution of coefficients (For completeness, the posterior predictive distribution is provided in the appendix). Note that straightforward integration characterizes all of these posterior moments as fractions of differing hypergeometric functions. However, they may all be expressed as functions of a single scalar $F_s^* \equiv {}_2F_1(\frac{N-1}{2}, 1, \frac{k_s+a}{2}, R_s^2)$ using Gauss' relations for contiguous hypergeometric functions (Abramowitz and Stegun, 1972, p.563). Let $\bar{N} \equiv N - 3$

[1] Note that due to this feature, the mode of density (1.4) is the local Empirical Bayes prior (EBL) from section 1.4.2.

[2] Note that $\mathrm{E}\left(\beta_s|y, X_s, M_s\right) = \mathrm{E}\left(\frac{g}{1+g}\middle| y, X_s, M_s\right)\hat{\beta}_s$.

and $\bar{\theta}_s \equiv k_s + a - 2$ represent collected terms. Tedious, but straightforward algebra then yields the following results for posterior moments (as long as $R_s^2 \in (0,1)$):[1]

$$\mathrm{E}\left(\frac{g}{1+g}\middle| y, X_s, M_s\right) = \frac{1}{R_s^2(\bar{N}-\bar{\theta}_s)}\left(\frac{\bar{\theta}_s}{F_s^*} - \bar{\theta}_s + \bar{N}R_s^2\right) \tag{1.7}$$

$$\begin{aligned}\mathrm{Cov}(\beta|y, X_s, M_s) &= \frac{\breve{y}'\breve{y}}{N-2}(X'X)^{-1}\ \frac{\bar{N}}{(\bar{N}-\bar{\theta}_s-1)^2-1}\frac{1-R_s^2}{R_s^2}\times \\ &\times\left(\left(1+\frac{2}{\bar{N}}\frac{R_s^2}{1-R_s^2}\right)\frac{\bar{\theta}_s}{F_s^*} + ((\bar{N}-2)R_s^2-\bar{\theta}_s)\right)\end{aligned} \tag{1.8}$$

Note that the equations above all contain the term $\bar{\theta}_s/F_s^*$.[2] So for each model's statistics, a hypergeometric function (or its Laplace approximation) has to be computed only once, which benefits numerical implementation in terms of computational burden.[3] In the BMA implementation used in the next section, the speed loss of hyper-g vs. a fixed-g setting is around 30%. Similarly cumbersome algebra also establishes the higher moments of the shrinkage factor and the moments of the predictive distribution as simple transformations of $\bar{\theta}_s/F_s^*$. Section 1.A.1 in the appendix presents these terms for reference, as well as a closed-form expression for the posterior coefficient density.

1.4.2 Properties of the hyper-g prior

The hyperparameter a can be trimmed to capture prior beliefs on the shrinkage factor in the associated Beta distribution. It is straightforward, for instance, to specify the prior beliefs such that the expected shrinkage factor matches the expressions of popular fixed g priors. In general, most popular settings for g can thus be emulated by $a = 2 + 2/w(N)$, with $w(N) > 0$, $w'(N) > 0$ and $\lim_{N\to\infty} w(N) = \infty$, thus positioning the prior expected value at $E(\frac{g}{1+g}) = \frac{w(N)}{1+w(N)}$. Setting a in dependence of sample size has the appealing virtue of ensuring 'consistency' in the sense of Fernández et al. (2001a, p.6).[4] By the same mechanism as in the corresponding fixed settings, the weight of the prior vanishes with increasing sample size and thus lets the posterior probability of a 'true' model $p(M_T|y)$ tend to to unity. Note that this applies to any true model; a proof is provided in section 1.A.2 in the appendix.
In this light, we concentrate on the following specifications for adaptive shrinkage priors:

- *HG-UIP*: $a = 2+\frac{2}{N}$ corresponds to the 'g-UIP'-shrinkage factor with $E\left(\frac{g}{1+g}\right) = \frac{N}{1+N}$. Then 95% of the prior mass on the shrinkage factor is contained in the interval $[1-0.95^N, 1]$.

[1] In case $R_s^2 = 0$ (in particular for the null model), the respective quantities are $\mathrm{E}\left(\frac{g}{1+g}\middle| y, X_s, M_s\right) = \frac{2}{k_s+a}$, and $\mathrm{Cov}\left(\beta_s|y, X_s, M_s\right) = \frac{2}{k_s+a}\frac{\breve{y}'\breve{y}}{N-2}(X'X)^{-1}$

[2] Note that $\bar{\theta}_s/F_s^*$ is just $2/(a-2)$ times the integration constant of $p(g|y, X_s, M_s)$ or $\bar{\theta}_s/F_s^* = \frac{a-2}{BF(M_s:M_0)}$ where $BF(M_s : M_0)$ is the null-based Bayes Factor for model M_s.

[3] Note that with respect to equations (1.7) and (A.1) it is straightforward to derive the corresponding expressions for $E(g|y, X_s, M_s)$ and $E(g^2|y, X_s, M_s)$. However, $E(g|y, X_s, M_s)$ will only be finite for $k_s+a > 4$ and $E(g^2|y, X_s, M_s)$ only for $k_s + a > 6$. We therefore concentrate on the posterior moments of the shrinkage factor.

[4] Consistency does not directly apply to the g-RIC prior outlined below. However, throughout the following sections, g-RIC is in practice identical with the g-BRIC prior (as always $K^2 > N$). Since the latter qualifies for consistency, the notion may be extended to g-RIC, at least in our case.

- *HG-RIC*: $a = 2+\frac{2}{K^2}$ corresponds to 'g-RIC'-shrinkage with $E\left(\frac{g}{1+g}\right) = \frac{K^2}{1+K^2}$. In this case 95% of the prior mass is contained in the interval $[1-0.95^{K^2}, 1]$. Akin to Fernández et al. (2001a), such a setting will be asymptotically consistent by choosing $w(N) = \max(N, K^2)$.

- *Empirical Bayes – Local (EBL):* $g_s = \arg\max_g p(y|M_s, X, g)$. Authors such as George and Foster (2000) or Hansen and Yu (2001) advocate an 'Empirical Bayes' approach by using information contained in the data (y, X) to elicit g. The latter provide a theoretical underpinning for doing so locally, i.e. separately for each model. In the formulation given in Liang et al. (2008), this corresponds to $g_s = \max(0, F_s - 1)$ where F_s is the standard F-statistic for M_s, with $F_s = \frac{R_s^2(N-1-k_s)}{(1-R_s^2)k_s}$. Note that this formulation frequently raises objections, since it is not necessarily consistent and the data-dependency of g runs counter the intuition of a prior.

Similarly, other specifications akin to 'classic' g formulations could be implemented – as long as they depend on N as defined above, in order to retain asymptotic consistency. Henceforth, we will refer to such elicited hyper-g priors as 'consistent hyper-g priors'. However, as posterior expressions are quite insensitive to the value of a, and most of these formulations will lead to a close to 2, the resulting posterior statistics will be virtually identical. We therefore limit our attention to the two specifications above.

Equations (1.7)-(1.8) reveal a certain resemblance to the respective posterior statistics under the 'Empirical Bayes - Local' (EBL) approach, whose posterior statistics depend on the OLS F-statistic. This feature is not surprising, as many Bayesian posterior statistics under a well-defined, non-degenerate prior asymptotically converge to their maximum-likelihood equivalent. Section 1.A.3 in the appendix shows that also the posterior model probabilities under EBL and consistent hyper-g priors (1.5) converge asymptotically, given that the 'true' model is not the null model (with zero covariates). But the similarities between EBL and hyper-g priors also extend to small samples: The main difference between their posterior expressions is the term $\bar{\theta}_s/F_s^*$, which guarantees non-negativity for the hyper-g statistics. Considering that the models associated with very low $\bar{\theta}_s/F_s^*$ (and thus high PMP) are disproportionally weighted into model averaging, this term thus virtually disappears from model-averaged statistics, if the data is not completely dominated by noise.[1] Consequently, the posterior statistics under both types of flexible priors will be very similar under any sample with a decent signal-to-noise ratio. However, compared to hyper-g, the EBL setting has two major drawbacks: First, it is not a prior in the classical sense, as it draws on the dependent variable. Second, it cannot be established whether the EBL setting is consistent if the true model is the null model. Nonetheless, due to its computational simplicity, the EBL prior can serve as a reasonable approximation to (and shares its asymptotic properties with) the hyper-g prior under data with a small noise component.

In view of the resemblance between the hyper-g prior results and the OLS- F-statistic, the posterior distribution of the shrinkage factor $\frac{g}{1+g}$ could be interpreted in terms of goodness-of-fit: Equation (1.7) presents its model-specific expected value as close to $1-1/\hat{F}_s$, where $\hat{F}_s$ represents an adjusted OLS F-statistic for the model M_s: $\hat{F}_s = \frac{R_s^2(\bar{N}-\bar{\theta}_s)}{(1-R_s^2)\bar{\theta}_s}$. Larger values of the shrinkage factor hence correspond to more variance explained by the model M_s. The model-averaged expected value of the shrinkage factor $E(\frac{g}{1+g}|y, X)$ may be interpreted likewise. If $K + a < N + 1$,the following inequality will hold asymptotically under a consistent hyper-g prior as N tends to infinity – in

[1]Note that ${}_2F_1(\frac{N-1}{2}, 1, \frac{k_s+a}{2}, R_s^2)$ increases rapidly as R_s^2 increases. The term $\bar{\theta}_s/F_s^*$ could thus noticeably affect model-averaged posterior moments only in case the data examined offers a very low signal-to-noise ratio.

small samples, it will hold as well except in cases of very low data quality (cf. section 1.A.4).

$$\frac{(1-p(M_0|y,X))^2}{1-E(\frac{g}{1+g}|y,X)} \leq E\left(\frac{\bar{N}-\bar{\theta}}{\bar{\theta}}\frac{R_s^2}{1-R_s^2}\middle|y,X\right) \equiv E(\hat{F}|y,X) \tag{1.9}$$

The model-weighted average of adjusted F-statistics thus establishes an upper bound for the posterior shrinkage factor. Consequently, the shrinkage factor can be related to goodness-of-fit in the data (y,X). The term involving the probability of the null model $p(M_0|y,X)$ is necessary, since the F-statistic of a model M_s will in general move in line with $E(\frac{g}{1+g}|y,X,M_s)$, except in case of the null model (there, the posterior conforms to the prior $\frac{2}{a}$).

A similar argumentation allows for relating posterior shrinkage to the F-statistic of the full data sample: If $N\to\infty$ under a consistent hyper-g prior, then the F-statistic of the full model with K regressors will form an upper bound for a function of posterior shrinkage (inequality (1.10)). Note that this inequality will also hold in small samples as long as there are some posterior model probabilities considerably larger than the one of the null model.[1]

$$\frac{1}{1-E(\frac{g}{1+g}|y,X)} \leq \frac{R_F^2}{(1-R_F^2)}\frac{(N-E(k|y,X)-a-1)}{(E(k|y,X)+a-2)} \tag{1.10}$$

Here, R_F^2 denotes the OLS R-squared of the full model, and $E(k|y,X)$ is the expected posterior model size. The right-hand side thus constitutes an adjusted F-statistic that relates R_F^2 with 'effective parameter size' $E(k|y,X)+a-2$. Note that this adjusted F-statistic is (almost sure) larger than the F-statistic of the full model, which illustrates the estimation advantage of shrinkage methods versus OLS. It is thus straightforward to express shrinkage as a function bounded by the unadjusted OLS F-statistic, which allows for applying likelihood-ratio significance tests in a classic sense. The relationship between the F-test and and information criteria thus implies that the posterior expected shrinkage factor is an indicator for goodness-of-fit of the model average that behaves similar to information criteria.

1.5 A simulation exercise

In this section we carry out a simulation study that empirically investigates the supermodel effect and assesses the predictive performance of selected prior structures. We group this broadly into *fixed* prior settings, as discussed in section 1.2, as opposed to model-specific adaptive *flexible g* priors (as in section 1.4.1). In the following, we concentrate on the 8 prior structures given in Table 1.1.

The first two fixed settings correspond to what Fernández et al. (2001a) coined the 'benchmark' prior and is widely used in applied work.[2] In macroeconomic studies such as (e.g., Fernández et al., 2001b), their recommendation usually results in the g-RIC prior. The implied (large) value for g under g-RIC is expected to have two consequences: first g-RIC will favor parsimonious models, and second posterior mass will be concentrated on a small set of models.[3] The unit information prior

[1] Even though this inequality will hold in virtually all relevant cases for small samples, it may not hold in case the dependent-covariate correlation is less than expected under a null hypothesis of no relation. As a rule of thumb, $R_F^2 > \frac{K+a-2}{N-3}$ is sufficient for (1.10) to hold in any case. Please refer to section 1.A.4 in the appendix for further details.

[2] See, for instance, Fernández et al. (2001b), Masanjala and Papageorgiou (2008) as well as Koop and Potter (2003).

[3] Note that this feature facilitates quick convergence of stochastic search algorithms such as the MC^3 to the target distribution.

Fixed Prior Settings	
g-RIC	Risk inflation criterion, $g = K^2$.
g-UIP	Unit information prior, $g = N$.
g-E($\frac{g}{1+g}\|y$)	$\frac{g}{1+g}$ is set to the posterior mean under the HG-4 prior (i.e. $E(\frac{g}{1+g}\|y)$).

Flexible Prior Settings	
EBL	Local empirical Bayes estimate of g.
HG-3	Hyper-g prior with $a = 3$.
HG-4	Hyper-g prior with $a = 4$.
HG-RIC	Hyper-g prior with $a = 2 + 2/K^2$.
HG-UIP	Hyper-g prior with $a = 2 + 2/N$.

Table 1.1: Definition of Prior Settings.

and the g-E($\frac{g}{1+g}|y$) complete the set up for fixed prior structures on g. For the latter we impose $\frac{g}{1+g}$ a priori to equate the (model weighted) posterior mean of ($\frac{g}{1+g}|y$) under the HG-4 setting (the hyperprior with $a = 4$). We have chosen this particular prior structure to exemplify the impact of adaptive shrinkage: both the HG-4 and g-E($\frac{g}{1+g}|y$) priors share the same average shrinkage factor and thus should yield a similar posterior model size distribution. However, posterior results are expected to seriously differ regarding the relative concentration of PMPs. In keeping g constant, the g-E($\frac{g}{1+g}|y$) setting will favor models which have a comparably small posterior support under the HG-4 prior. The differing results between g-E($\frac{g}{1+g}|y$) and HG-4 thus illustrate the impact of model-specific shrinkage as opposed to adapting the aggregate shrinkage factor to the data.

The flexible prior structures with model specific, data-dependent shrinkage divide into local empirical Bayes (EBL) estimates and the hyper-g prior corresponding to a fully Bayesian approach. One strength in placing a prior on g lies in the fact that we can incorporate our prior beliefs following the rules of Bayesian statistics[1] via the hyperparameter a. For the simulation study, we devise four different values for a: HG-3 (a=3) corresponds to a prior expected shrinkage factor of $\frac{2}{3}$, whereas HG-4 (a=4) corresponds to a flat prior over the shrinkage factor. We contrast these two settings with two consistent priors that are calibrated to match the g-RIC and g-UIP prior structure (HG-RIC, HG-UIP). I.e., the prior expected value of the shrinkage factor $E(\frac{g}{1+g})$ conforms to the shrinkage factors induced by g-RIC ($g = K^2$) or g-UIP ($g = N$).

Data-wise, we employ two different settings, where the first set-up 'A' follows Fernández et al. (2001a). Each Monte Carlo run draws ten potential explanatory variables ($\mathbf{x}_1, \ldots, \mathbf{x}_{10}$) with $N = 100$ observations from a standard normal distribution for each covariate. Five more variables are generated by multiplying the first five regressors with the vector $(0.3, 0.5, 0.7, 0.9, 1.1)$ in order to induce a correlation structure among the covariates. Note that this correlation structure hampers uncovering the data-generating model under short samples.

The second set-up 'B' is more demanding since the data-generating process cannot be traced back to a single model. This is more in line with the Bayesian model averaging approach, whose question is not whether the preferred model is perfectly true, but whether under the assumed model(s) the observed data is a plausible outcome.[2] The data-generating process is composed of five partially nested models with unequal model weights imposed. This creates a 'hierarchy' of models, with y_4

[1] See Laud and Ibrahim (1995) for a model selection approach designing information criteria that allow for the input of prior knowledge.

[2] See, for instance, Gelman et al. (1995).

and y_5 dominating the remaining models in terms of explained variation.[1]

Setup 'A': $y = 4 + 2\mathbf{x}_1 - \mathbf{x}_5 + 1.5\mathbf{x}_7 + \mathbf{x}_{11} + 0.5\mathbf{x}_{13} + \sigma\varepsilon$

Setup 'B':
$$\begin{aligned}
y &= 0.2y_1 + 0.1y_2 + 0.1y_3 + 0.3y_4 + 0.3y_5\\
y_1 &= 4 + 2\mathbf{x}_1 - \mathbf{x}_5 + 1.5\mathbf{x}_7 + \mathbf{x}_{11} + 0.5\mathbf{x}_{13} + \sigma\varepsilon\\
y_2 &= 4 + 4\mathbf{x}_1 - \mathbf{x}_5 + 1.5\mathbf{x}_2 + \mathbf{x}_8 + 0.5\mathbf{x}_{11} + \sigma\varepsilon\\
y_3 &= 4 + 1\mathbf{x}_5 - \mathbf{x}_7 + 1.5\mathbf{x}_3 + \mathbf{x}_9 + 0.5\mathbf{x}_6 + \sigma\varepsilon\\
y_4 &= 4 + 2\mathbf{x}_1 - \mathbf{x}_2 + 1.5\mathbf{x}_4 + \mathbf{x}_7 + 0.5\mathbf{x}_6 + \sigma\varepsilon\\
y_5 &= 4 + 2\mathbf{x}_1 - \mathbf{x}_{10} + 1.5\mathbf{x}_{11} + \mathbf{x}_{12} + 0.5\mathbf{x}_{13} - 2\mathbf{x}_{14} + \sigma\varepsilon
\end{aligned}$$

Posterior inference under the different prior structures will be examined with varying signal-to-noise ratios. In particular we conduct the simulation study for four levels of noise:[2] $\sigma = 1/2, \sigma = 1, \sigma = 2.5, \sigma = 5$. For each value of σ and each setting, results are computed as averages from 50 Monte Carlo draws. The relatively low number of covariates $K = 15$ allows for easily enumerating the full model space of 2^K models. This guarantees that the differences of results for the competing priors do not arise from variation due to stochastic search.

Empirical research frequently focuses on the posterior inclusion probabilities (PIPs) of the variables entering the analysis and the posterior moments of the related coefficients. Table 1.2 and 1.3 highlight PIPs for setting 'A': Under a small degree of noise ($\sigma = 1/2$ and $\sigma = 1$) results do not differ considerably between fixed and data-dependent priors for g. Under the $\sigma = 2.5$ setting, the PIPs of the coefficients from the the data-generating model exhibit differences in magnitude but still lead to the same interpretation. Results change when looking at the $\sigma = 5$ case. Posterior mass under the flexible priors is spread more evenly than under fixed g priors. The g-RIC prior shows strong support for the first variable, with a large PIP for β_1 of approximately 0.8. The remaining variables receive negligible posterior support, tempting the researcher to believe that the data-generating process is solely driven by the first variable. In contrast, flexible priors still 'identify' all variables. As expected, mass is spread more evenly, and over larger models,[3] which results in a high share of covariates with PIP close to 0.5 and thus reflects the serious degree of noise in the data. Note, however, that under high noise, the g-E($\frac{g}{1+g}|y$) prior (which is data-adaptive but not model-specific) is more prone to misidentifying variables (by PIP) than the hyper-g priors. This result suggests that while adapting shrinkage to data quality is crucial, it is also important to allow for model-specific adjustment of the shrinkage factor.

In addition to PIPs, the posterior model probability of the data-generating model can be of interest to examine consistency properties in the sense of Fernández et al. (2001a). Tables 1.6 and 1.7 show summary statistics for its posterior model probability (under setting 'A'). In line with asymptotic consistency, more information in the data leads the hyperprior to uncover the data-generating process with highest precision, whereas increasing noise deteriorates the selection ability of BMA for all settings. The ratio of the posterior model probability for the data-generating process to the one with highest PMP is given in Table 1.7. The results show that in situations described by higher degrees of noise in the data all specifications favor a model different from the one generating the data.

[1] Note that setup 'B' is observationally equivalent to generating data from a single, complicated model.

[2] Note that in this setting, varying σ has an effect that is similar to varying the number of observations N. We therefore leave N constant.

[3] Note that the sum of PIPs equals posterior model size. Therefore, if posterior mass concentrates on larger models due to noise, the PIPs will not discriminate much among covariates but will exhibit high absolute values. It is therefore more insightful to compare the relative differences in PIPs rather than their absolute values.

While flexible priors fail to uncover the data-generating model (as do the fixed priors) the assigned PMP for the best (and wrong) model is considerably smaller than under fixed priors. Hence, the degree of uncertainty is reflected in the evenness of posterior mass distribution. Figures 1.2 and 1.3 exemplify the differences in PMP concentration for the 8 priors. The first figure shows the cumulative posterior mass of the 100 best models under the four signal-to-noise settings. From these figures and Table 1.6, it becomes evident that flexible priors uncover the data-generating model with highest precision *and* concentrate most mass on this model(s) in situations characterized by a high degree of information in the data. This means that under the flexible priors, posterior expected shrinkage $E(\frac{g}{1+g}|y,X)$ is larger than the constant factors $\frac{g}{1+g}$ under the fixed priors. As noise increases, the flexible priors distribute mass more evenly among explanatory variables which reflects the rise in uncertainty. In contrast, fixed g priors are not capable of adjusting posterior mass distribution to uncertainty inherent in the data. This inability to adapt limits the merits of Bayesian model averaging under fixed g priors with respect to robust inference and predictive ability. Figure 1.3 uses a QQ-plot to compare for PMPs under the various prior settings with the g-RIC prior. For all data-dependent priors, differences to fixed priors increase with noise as expected.[1]

Under setting 'B', the employed models should rather be understood as approximations, while uncovering a 'true' model is of minor importance. The results exemplify once again the supermodel effect behavior of fixed prior settings illustrated in Figures 1.4 and 1.5. Small degrees of noise trigger a concentration of posterior mass under the hyper-g prior and the empirical Bayes approach. An increase in noise is reflected in a wider spread of posterior mass among models under flexible priors, whereas fixed priors still concentrate on a small number of models. Moreover, note how close the hyper-g results are to the Empirical Bayes prior under both settings 'A' and 'B', which are particularly striking when noise is small. This illustrates that not only both priors converge to the same posterior results asymptotically,[2] but that also lead to similar conclusions under small samples (that are characterized by a noise component that is not too large).

Finally, we examine the robustness of the various g-prior settings via their performance in out-of-sample prediction. Results from a prediction exercise are expected to vary considerably between fixed and flexible prior settings, since the latter incorporate data adaptive shrinkage. Akin to Liang et al. (2008) we randomly split the data from settings 'A' and 'B' into 70 estimation and 30 out-of-sample observations. We then calculate the root mean squared error (RMSE) of the forecasts for the 30 out-of-sample data points, averaged over 50 Monte Carlo steps. The RMSE statistics shown in Table 1.8 are normalized with respect to forecasting results under the g-RIC prior. Thus values below 1 indicate better predictive accuracy of the respective prior structure than under the g-RIC prior. The top panel of Table 1.8 shows mixed results for setting 'A'. As expected, the g-RIC prior with its focus on parsimonious models excels in nearly all signal-to-noise settings, concentrating on a single (and luckily the correct data-generating) model. In the $\sigma = 1/2$ case, however, the flexible priors concentrate mass even more tightly than does the g-RIC and consequently yield better predictions in terms of RMSE. At intermediate noise levels, g-RIC outperforms the other priors by greater margins by exploiting the comparative advantage that the data-generating process is composed of a single model. This shows *how the supermodel effect could be exploited* to achieve superior predictive performance under the following conditions: If a researcher has prior knowledge that the data is generated by a simple model with few covariates, and if the noise component is neither too weak such that flexible priors would outperform any fixed priors, nor too strong such

[1] We have omitted results from the HG-3 setting, since results are very similar to that of HG-4.

[2] Note that under this simulation set-up, decreasing the variance of the noise component is equivalent to increasing the sample size at constant variance. The striking similarities of results under Empirical Bayes and hyper-g priors therefore illustrate the asymptotic equivalence between both set-ups (cf. section 1.4.2).

that a fixed prior setting would hardly identify the one 'true' model as the best-performing one, then a high g-parameter will concentrate more mass on the 'correct' model than flexible prior settings. In view of this conditions, the supermodel effect can be exploited best in a typical simulation set-up. However, such an approach can be dangerous under more complex datasets, and in any case when such peculiar prior knowledge is not given. Simulation setting 'B' illustrates the merits of flexible priors when the lack of ideal conditions turns the supermodel effect into a disadvantage for fixed g-priors: The predictive performance of flexible priors dominates throughout nearly all signal-to-noise settings. Especially the HG-3 prior and the empirical Bayes approach demonstrate superior predictive abilities with the latter one outperforming the g-RIC prior for all signal-to-noise setups. This contrasts with typical simulation exercises in the literature, as their data-generating processes emanates from a parsimonious single models, which plays in favor of (large) fixed priors because of the supermodel effect. Given their focus on large models, it is surprising how well hyper-g priors perform in this prediction exercise. The results suggest that imposing a model prior that favors parsimonious models yields even better predictive performance under a flexible prior.[1]

1.6 (Un)stable growth determinants

Emanating from the pioneering work of Sala-i-Martin et al. (2004),[2] numerous studies have employed model averaging techniques in the empirics of economic growth. For comparability, we follow this approach and investigate the effect of the prior settings from section 1.5 on inference in a cross-country growth data set. A vast part of the empirical studies use international income data provided by the Penn World Tables (PWT). This data base publishes GDP data adjusted for purchasing power parities (PPP), which is essential for conducting country comparison studies. The core purpose of the PWT is collecting prices for the same or similar goods in different countries. Gathering prices is carried out on an irregular basis with each 'generation' of the table complying with a different round of price collection (Johnson et al., 2009). The methodologies employed by PWT to derive PPP-adjusted GDP data - and in particular growth rates thereof - have been frequently criticized as being plagued by considerable measurement error. Johnson et al. (2009) carry out a replication exercise by re-estimating selected growth equations employed in the literature using different versions of PWT. They show that estimates vary markedly across different versions of the PWT. In particular, growth studies using high frequency data (e.g. annual as opposed to long run averages) are especially prone to the measurement error inherent to the PWT methodology. Johnson et al. (2009) conclude that in order to make significant policy conclusions empirical results should be robust to PWT revisions.

Ciccone and Jarociński (2010) investigated the impact of PWT revisions on Bayesian model averaging results. They use three different versions of the Penn World Table income data (PWT 6.0, PWT 6.1 and PWT 6.2) and show that the identification of 'robust' determinants - as measured by the respective PIP - varies tremendously among data revisions. Feldkircher and Zeugner (2012) argue that this instability is partially rooted in the prior setup chosen by Ciccone and Jarociński (2010): a uniform prior on the model space is coupled with the g-RIC prior, with the latter being characterized by the 'supermodel effect'.

In what follows we build on this example and estimate cross-country growth regressions for an extended set of four PWT revisions:

[1] Tentative results available from the authors on request.

[2] See also Crespo Cuaresma and Doppelhofer (2007) and Sala-i-Martin (1997) for traditional approaches to model averaging as opposed to Fernández et al. (2001b) and Eicher et al. (2011) for Bayesian strategies.

$$\Delta y^j = \alpha + \gamma y_0^j + \vec{\beta}_s X_s + \varepsilon, \tag{1.11}$$

Here, Δy^j denotes the average annual growth of income per capita over the period from 1960 to 1996 for $N = 75$ countries, α the intercept, ε the error term, and $\mathbf{X}_s = (\mathbf{x}_1 \dots \mathbf{x}_s)$ a matrix whose columns represent a subset s of explanatory variables. Initial income is denoted by y_0^j and is the only explanatory variable that changes with PWT vintages. The potential growth determinants (whose combinations are represented by $\mathbf{X}_s$) are the ones originally put forward in Sala-i-Martin et al. (2004) and employed in Ciccone and Jarociński (2010). These 66 variables comprise measures for factor accumulation and convergence (as implied by the Solow growth model), human capital, institutional environment, and socio-geographical determinants. The estimation is carried out for each of the four considered PWT revisions, indexed by $j \in \{\text{PWT 6.0}, \text{PWT 6.1}, \text{PWT 6.2 and PWT 6.3}\}$. Note that all of them stem from the same PWT generation and are thus based on the same raw price data. In this setting, we examine the effect of small perturbations to income data on posterior results under different g priors. The correlation of the dependent variable between vintages ranges from 0.92 to 0.97 and thus the revisions can be considered as reasonably small perturbations. In order to represent loose prior expectations about model size, all estimations are based on the Ley and Steel (2009) binomial-beta model prior anchored at an expected model size of $K/2$ variables.[1]

We compare the stability of the posterior inclusion probabilities over revisions by the ratio of maximum over minimum PIP for a variable: $\text{Max/Min}_k = \max(PIP(X_k^j))/\min(PIP(X_k^j))$, with $k = 1, \dots, 67$ and $j = 1, \dots, 4$ denoting the four PWT data sets. Note that for all PWT vintages, the g-RIC prior complies with the 'benchmark' prior put forward in Fernández et al. (2001a). Moreover, note that in Ciccone and Jarociński (2010) the country sample - and thus the number of observations - changes from revision to revision. Feldkircher and Zeugner (2012) show that conditioning on the same observations throughout the revisions reduces the instability of posterior results. We therefore opt for holding samples constant since we are interested in the part of PIP instability that is caused by employing the fixed g-RIC prior (the benchmark prior). Table 1.10 summarizes the results. Employing a flexible prior decreases PIP variation to a great extent, regardless of which PWT vintages are considered: The overall Max/Min ratios under a flexible prior are 67% smaller than under the fixed g-RIC prior.

This suggests that posterior results of the fixed and flexible priors will also differ considerably in qualitative terms. To further examine the robustness of growth determinants, Table 1.11 lists the posterior inclusion probabilities per variable under the g-RIC and the hyper-g (UIP) prior for all four PWT vintages. Explanatory variables are labeled 'robust' when they exceed a PIP of 0.50. This threshold can be motivated from a predictive stance (Barbieri and Berger, 2003) as well as from an intuitive perspective. Under the g-RIC prior, only a proxy for human capital (primary schooling in 1960) can be considered as robust in all four PWT vintages. Results are unstable for regional dummy variables such as a dummy for East Asia, Tropical Area, and Latin America. Moreover, no clear-cut results emerge on the effect of initial GDP on international differences in income. This is particularly worrisome from the viewpoint of economic theory and casts doubts on the results under the g-RIC setting. The hyper-g (UIP) prior, in contrast, identifies several variables as robust. Both initial GDP and primary schooling display impressive posterior support ($PIP \geq 90\%$) and their posterior coefficients are well in line with economic theory.[2] Furthermore, the regional dummy for Africa, and variables for the share of Confucian population receive robust posterior support over all PWT revisions, which provides strong empirical evidence in the vein of

[1] Note that Ciccone and Jarociński (2010) have used a uniform prior on the model space. Results under the uniform prior are available from the authors upon request. See also Feldkircher and Zeugner (2012).

[2] Results are available from the authors upon request.

Johnson et al. (2009). Finally, the proxies for fertility and for Buddhism show considerable posterior support in three out of the four PWT vintages. The robustness of regional dummy variables points to heterogeneous growth dynamics in the examined country set.

The difference in posterior results between flexible and fixed priors is rooted in how posterior mass is distributed among covariates and models. For the PWT data set, this can be seen best when comparing the posterior expected model size. While a model for economic growth should be expected to contain 3 to 5 explanatory variables according to the g-RIC prior, the posterior model size for flexible priors lies in the range of 12 to 14 regressors. Accordingly, the corresponding shrinkage factors of the flexible priors are considerably lower (across all revisions) than the value implied by the fixed g-RIC prior (see the bottom of Table 1.11). These differences become even more striking when considering the implied posterior mean for g, which is 4448 under the g-RIC compared to 23 - 26 under the hyper-g (UIP) prior. The smaller shrinkage factors (and implied values for g) under the flexible priors indicate noise to prevail more strongly under PWT data sets than inherently assumed under g-RIC. With that in mind, it comes as no surprise that the flexible priors push the PIPs of most variables towards the 50% threshold, which points to empirical evidence neither in favor nor against their inclusion in models of economic growth. The inconclusiveness that plagues the empirical growth literature is thus more correctly mirrored in the posterior results of the flexible priors, whereas the benchmark prior is prone to the risk of over-fitting the PWT data sets.

Figure 1.6 (upper left panel) illustrates the problems arising from fixed g priors in the (PWT 6.1) data set. The concentration of posterior model probabilities under flexible priors is considerably lower than under the g-UIP setting, and is far less than under the g-RIC setting. In view of the discussion in section 1.3, this points to the supermodel effect severely affecting the posterior statistics in this popular data set. This characteristic is also mirrored in the corresponding posterior inclusion probabilities (Figure 1.6, upper right panel). Under fixed-g settings (particularly g-RIC), the PIPs are more skewed than under flexible priors, which again illustrates that fixed-g priors risk discriminating more among covariates than seems justified by the data at hand. Moreover, figure 1.6 displays virtually identical results for the various hyper-g settings, which illustrates the feature that their differing prior expectations do not have too much impact on posterior statistics. Finally, their results are indistinguishable from the Empirical Bayes prior, which further illustrates how quickly hyper-g and EBL prior results converge in small samples.

The above results illustrate the advantages of flexible g priors in BMA inference. However, it is not clear whether these advantages carry over to robustness in predictive performance: On the one hand, the instability of of fixed priors such as the g-RIC and g-UIP should deteriorate forecast performance. On the other hand, it is a stylized fact in the forecasting literature that parsimonious models outperform saturated models in terms of forecasting quality. It is thus not clear a priori which prior setting excels in forecasting economic growth. We therefore conduct a forecast evaluation in which we randomly split the observations into a training sample of 56 countries and a 'hold-out' sample of 19 countries. The training sample is used to estimate the models for forecasting the remaining observations of the hold-out sample. The corresponding root mean square error (RMSE) is calculated over 30 random sample splits. Table 1.8 summarizes the results: All flexible prior settings outperform the fixed priors g-RIC and g-UIP. In particular, the predictive performance of the hyper-g (UIP) prior is superior to the other priors in most of the PWT revisions.

Finally, note that the posterior expected shrinkage factor, an indicator for goodness-of-fit, did not steadily increase from PWT 6.0 to PWT 6.3. We notice that goodness-of-fit and data quality need not to go hand in hand.[1] However, we think a goodness-of-fit measure might be a reasonable indicator to get a first impression about how the data revision progressed. The sharp drop of the shrinkage factor from PWT 6.1 to PWT 6.2 might further stimulate the debate whether newer is always better in the context of PWT revisions.[2]

1.7 Concluding remarks

The widespread use of Zellner's g prior in linear BMA rests on two convenient features: it provides closed-form solutions and reduces the complexity of prior elicitation to the scalar g. Consequently, theoretical considerations have mostly focused on the choice of g, in particular in view of its virtues as a penalty term for model size. This study departs from earlier literature in bringing forward two arguments that have been overlooked so far: First, model size considerations should be disentangled from the primary purpose of g (which is scaling coefficient covariance) and rather be fused into the formulation of model priors. The elicitation of g should thus not interfere with prior desiderata on model size. Second, we demonstrate that fixing g to arbitrary values may have unintended consequences on posterior model probabilities: The higher g, the more tightly posterior mass will concentrate on the few best-performing 'supermodels' – regardless of model sizes, number of observations, or signal-to-noise ratios. Ultimately, a large value for g will favor a single model, thereby emulating model selection rather than model averaging. As previous studies predominantly have assessed BMA performance on simulated data generated by a single model, they tended to favor g-specifications with large values of g that effectively *select* the right model. However, in empirical practice a large g runs the risk of putting to much posterior weight on a single model. We demonstrate that the popular prior suggested by Fernández et al. (2001a) is particularly prone to this behavior.

As it is virtually impossible to specify the 'right' value for g under unknown variance, we propose to put a prior distribution on this parameter instead: Such a hyperprior allows for data-adaptive and model-specific shrinkage, thus adjusting the impact of prior beliefs to data quality. In discriminating among models only as far as data quality allows, a prior on g thus offers a remedy for the supermodel effect. In this manner, we focus on a particular *hyper-g* prior, whose formulation offers three main advantages: First, it admits closed form solutions for almost any quantity of interest, thereby facilitating implementation. Second, its hyperparameter allows for formulating prior beliefs on coefficient variance, but without incurring the risk of unintended consequences on posterior model mass. Third, we demonstrate that the hyper-g prior can be reconciled with BMA consistency. We complement the existing literature on the hyper-g prior by providing additional posterior expressions that allow for fully Bayesian inference, as well as for sound numerical implementation. Moreover, we demonstrate that its posterior statistics can be considered as a goodness-of-fit indicator, and show why its results are closely related to those of the Empirical Bayes g-prior.

A simulation exercise contrasts various formulations of fixed and hyper-g priors. The fixed g-priors perform well when the data-generating process rests on a single model that is part of the candidate model space – but so does the hyper-g prior. The virtues of flexible prior structures become evident in more complex settings: Flexible priors outperform fixed g settings in terms of forecasting accuracy and exhibit a more stable structure of posterior model and inclusion probabilities over varying degrees of noise in the data.

[1] An increase in goodness-of-fit might be driven by stronger correlation of measurement error in both, explanatory variables and dependent variable.

[2] See also Johnson et al. (2009).

The final section illustrates these conclusions by investigating fixed and flexible priors under different revisions of an economic growth data set. The results demonstrate that fixing g has a detrimental effect on the stability of posterior results. While fixed g-priors list initial income among the most unstable growth determinants, the estimates from the flexible priors are well in line with economic theory: Both conditional income convergence and human capital are identified as robustly related to income growth, along with a handful of regional dummies. In contrast to fixed settings, these results are insensitive to data revisions. Concluding, the hyper-g prior offers a sound, fully Bayesian approach that features the virtues of prior input and predictive gains without incurring the risk of mis-specification.

1.A Technical appendix

1.A.1 Posterior statistics of the hyper-g prior: Further results

Joining tedious algebra with Gauss' contiguous relations for hypergeometric functions(Abramowitz and Stegun, 1972, p.563) allows to establish important posterior expressions of the hyper-g prior in closed form, on top of the ones provided in section 1.4.1. It is important to note that the posterior moments of coefficients, the shrinkage factor, and the predictive distribution all arise as simple transformations of the posterior model likelihood in equation (1.5). Their implementation thus bears virtually no computational cost. The following equations present the most important posterior statistics (keeping the notation from section 1.4.1). Equation (A.1) expresses the posterior second moment of the shrinkage factor if ($R_s^2 \in (0,1)$):[1]

$$\begin{aligned} \mathrm{E}\left(\left(\frac{g}{1+g}\right)^2 \middle| y, X_s, M_s\right) &= \frac{1}{(R_s^2)^2(\bar{N}-\bar{\theta}_s)(\bar{N}-(\bar{\theta}_s+2))}\times \\ \times \left(\left((\bar{N}-2)R_s^2-(\bar{\theta}_s+2)\right)\frac{\bar{\theta}_s}{F_s^*} + \left(\bar{N}R_s^2-\bar{\theta}_s\right)^2 - 2\left(\bar{N}(R_s^2)^2-\bar{\theta}_s\right)\right) \end{aligned} \tag{A.1}$$

In addition to posterior moments, the posterior distribution of coefficients $\beta_s|y, M_s$ can also be established in closed form, but as ratio of two hypergeometric functions:

$$\begin{aligned} p(\beta_s|y,X_s,M_s) &= \int_0^\infty p(\beta_s|y,X_s,M_s,g)p(g|y,X_s,M_s)dg = \\ = \frac{\Gamma\left(\frac{N-1+k_s}{2}\right)\Gamma\left(\frac{k_s+a}{2}\right)\frac{N-1}{2}}{\Gamma\left(\frac{N-1+k+a}{2}\right)} &\frac{\sqrt{|X_s'X_s|}}{\pi^{\frac{k_s}{2}}}(\breve{y}'\breve{y})^{\frac{N-1}{2}}(\beta_s'X_s'X_s\beta_s)^{-\frac{N-1+k_s}{2}}\times \\ \times \frac{{}_2F_1\left(\frac{N-1+k_s}{2}, \frac{N-1}{2}+1, \frac{N-1+k_s+a}{2}, 1-\frac{(y-X_s\beta_s)'(y-X_s\beta_s)}{\beta_s'X_s'X_s\beta_s}\right)}{{}_2F_1\left(\frac{N-1}{2},1,\frac{k_s+a}{2},R_s^2\right)} \end{aligned} \tag{A.2}$$

Note that this expression is of close, though not perfect resemblance to a hypergeometric function distribution of type II.[2] The first two moments of this distribution are provided in section 1.4.1.

Similar algebra also allows for simplifying the predictive distribution and yields its moments in closed form: Consider using the data (X, y) to forecast the dependent variable $\hat{y}$ conditional on 'prediction' covariates $\hat{X}$. Let X be an $N \times k$ matrix, y be $N \times 1$, while $\hat{y}$ is $l \times 1$ and $\hat{X}$ $l \times k$. The posterior predictive distribution of $\hat{y}$ is then given as a a multivariate t-distribution of dimension l (Eklund and Karlsson, 2007, equation (A.15))[3]

$$\begin{aligned} \hat{y}|\hat{X},X,y,g,M_s &\sim t_l(\bar{y}+s\hat{X}\hat{\beta},\Sigma,N-1) \\ \text{where } \Sigma &= \left(I_l + s\hat{X}(X'X)^{-1}\hat{X}'\right)\frac{\breve{y}'\breve{y}}{N-1}(1-sR_s^2) \end{aligned}$$

[1] In case $R_s^2 = 0$ (which relates in particular to the null model), the expression for (A.1) is $\mathrm{E}\left(\left(\frac{g}{1+g}\right)^2 \middle| y, X_s, M_s\right) = \frac{8}{(k_s+a)(k_s+a+2)}$

[2] See Guptar and Nagar (2000) for the exact definition of the type II hypergeometric distribution.

[3] The slight differences with respect to Eklund and Karlsson (2007) are due to the fact that we employ an improper prior on beta variance σ and the constant.

Here, s denotes the shrinkage factor $s = \frac{g}{1+g}$, and R_s^2 the (centered) R-squared of y on X. $\bar{y}$ denotes an N-dimensional vector whose elements are the arithmetic mean of y, and $\breve{y} \equiv y - \bar{y}$ the centered response variable. Integrating the density function of $\hat{y}|\hat{X}, X, y, g, M_s$ with respect to the shrinkage factor yields the integrand of the following equation (after some rearrangement):

$$f(\hat{y}|\hat{X}, X, y, M_s) = \frac{\Gamma\left(\frac{N-1+l}{2}\right)}{\Gamma\left(\frac{N-1}{2}\right)\pi^{\frac{l}{2}}} \overbrace{\frac{(k+a-2)((\breve{y}'\breve{y})^{\frac{N-1}{2}})}{2F_s^*}}^{\frac{a-2}{2}\frac{1}{p(y|X)}} \times$$
$$\times \int_0^1 \left| I_l + s\hat{X}(X'X)^{-1}\hat{X}' \right|^{-\frac{1}{2}} (1-s)^{\frac{k+a-4}{2}} \times$$
$$\times \left(\breve{y}'\breve{y}(1-sR_s^2) + (\hat{y} - \bar{y} - s\hat{X}\hat{\beta})' \left(I_l + s\hat{X}(X'X)^{-1}\hat{X}' \right)^{-1} (\hat{y} - \bar{y} - s\hat{X}\hat{\beta}) \right)^{-\frac{N-1+l}{2}} ds$$

To our knowledge, there is no closed-form solution to to the integral above, neither to its Laplace approximation. We therefore recommend to resort to numerical integration. Nonetheless, it is possible to obtain the predictive variance, i.e. the squared predictive standard error, as:

$$Var(\hat{y}|y, X, \hat{X}, M_s) = \hat{X} Var(\beta|y, M_s)\hat{X}' +$$
$$+ \frac{\breve{y}'\breve{y}}{N-3}\frac{N+l}{N} I_l \left(\frac{(N-3)(1-R_s^2)}{N-1-k-a} - \frac{k+a-2}{N-1-a-k} / F_s^* \right)$$

1.A.2 Consistency of the hyper-g prior

Fernández et al. (2001a) define asymptotic 'consistency' as follows: Consider that only Model M_s is true, while all other models $M_j \neq M_s$ are not true. Consistency then requires:

$$\operatorname*{plim}_{n\to\infty} p(M_s|y, X_s) = 1 \quad \text{and} \quad \operatorname*{plim}_{n\to\infty} p(M_j|y, X_s) = 0 \quad \forall M_j \neq M_s$$

Liang et al. (2008, Appendix B) have proven the above for the hyper-g prior except for the case where the true model M_s is the null model M_0. They stop short their proof because in this case the Bayes factor $B(M_j : M_0)$ is (Liang et al., 2008, p.423):

$$\frac{p(M_j|y, X_s)}{p(M_0|y)} \geq \int_0^\infty (1+g)^{-\frac{k_j}{2}} p(g) dg \tag{A.3}$$

Moreover they state that if the above integral vanishes as $N \to \infty$, then consistency is ensured. Applying the hyper-g setting transforms the right-hand side in (A.3) into the following (by $a > 2$):

$$\int_0^\infty (1+g)^{-\frac{k_j}{2}} p(g) dg = \frac{a-2}{2} \int_0^1 (1+g)^{-\frac{k_j+a}{2}} dg = \frac{a-2}{k_j+a-2}$$

If $a = 2 + w(N)$ with $w(N) > 0$ and $\lim_{N\to\infty} w(N) = 0$, then the integral vanishes and thus concludes the proof.

1.A.3 Relationship between hyper-g and Empirical Bayes prior

It is well established in Bayesian statistics that under any non-degenerate prior, Bayesian regression results asymptotically approach their maximum-likelihood equivalent with increasing sample

size. Against this backdrop it is not surprising that the results under the Empirical Bayes prior in section 1.5 are close to the hyper-g settings. This sections outlines why posterior model probabilities under consistent hyper-g priors and the Empirical Bayes are close in small samples and converge asymptotically. As a byproduct, this section demonstrates the asymptotic consistency of the Empirical Bayes prior if the ' true' model is not the null model. The results in this section are based on Laplace approximations[1], cf. Gelfand and Dey (1994) for their asymptotic properties in the context of Bayesian model selection.
Consider the familiar form of the Laplace approximation, where $\hat{\theta}$ is the maximizer of the integrand's logarithm $h(\theta)$:

$$\int_{\Theta} \exp(h(\theta))d\theta \approx \sqrt{\frac{2\pi}{-h''(\hat{\theta})}} \exp h(\hat{\theta})$$

Consider in turn the Bayes Factor for the hyper-g prior formulation as in (1.5), between a model with k covariates, and the null model:

$$BF_h = \frac{a-2}{2} \int_0^{\infty} (1+g)^{\frac{N-1-k-a}{2}} (1+g(1-R^2))^{-\frac{N-1}{2}} dg$$

Letting

$$h(g) = \frac{1}{2}\big((N-1-k-a)\log(1+g) - (N-1)\log(1+(1-R^2)g)\big)$$

yields the maximizer:

$$\hat{g} = \max(\frac{R^2(N-1-k-a)}{(1-R^2)(k+a)} - 1, 0)$$

where $\hat{g} = 0$ if and only if $k + a \geq R^2(N-1)$. Liang et al. (2008, p.421) note the similarity to the local Empirical Bayes (EBL) estimator of g, but abstain from further investigating the issue.
The second derivative of $h(g)$ is given as

$$h''(g) = \frac{1}{2}\left(-\frac{N-1-k-a}{(1+g)^2} + \frac{(N-1)(1-R^2)^2}{(1+(1-R^2)g)^2}\right)$$

The Bayes factor under a hyper-g prior is thus approximately equal to:

$$BF_h \approx (a-2)\sqrt{\frac{\pi}{\frac{N-1-k-a}{(1+\hat{g})^2} - \frac{(N-1)(1-z)^2}{(1+(1-z)\hat{g})^2}}} (1+\hat{g})^{\frac{N-1-k-a}{2}} (1+\hat{g}(1-R^2))^{-\frac{N-1}{2}}$$

In case we have $\hat{g} > 0$, then algebraic manipulation of the expression above yields:

$$BF_h \approx (a-2)\sqrt{\pi}\sqrt{\frac{N-1}{(N-1-k-a)(k+a)}} \left(\frac{R^2}{(1-R^2)}\frac{N-1-k-a}{k+a}\right)^{-\frac{k+a-2}{2}} \left(\frac{(1-R^2)(N-1)}{N-1-k-a}\right)^{-\frac{N-1}{2}}$$

Now consider the equivalent null-based model Bayes Factor for the EBL approach which is:

$$BF_{EBL} = \left(\frac{R^2}{1-R^2}\frac{N-1-k}{k}\right)^{-\frac{k}{2}} \left(\frac{(1-R^2)(N-1)}{(N-1-k)}\right)^{-\frac{N-1}{2}}$$

in case if $k \leq R^2(N-1)$

[1]Note that due to perceived numerical difficulties, Liang et al. (2008) propose the use of a Laplace approximation for the posterior model likelihood under the hyper-g distribution (Liang et al. (2008, equation (17))). Depending on the data, Laplace approximations can be prone to substantial numerical inaccuracies in small samples. However, they may be useful for the purpose of this section which is mainly interested in asymptotic results.

Therefore, if $k + a \leq R^2(N-1)$:

$$BF_h \approx (a-2)\sqrt{\pi}\sqrt{\tfrac{N-1}{(k+a)(N-1-k-a)}}\left(\tfrac{(1-z)}{z}\tfrac{k+a}{N-1-k-a}\right)^{\frac{a-2}{2}}\left(\tfrac{k+a}{k}\right)^{\frac{k}{2}}\left(\tfrac{N-1-k-a}{N-1-k}\right)^{\frac{N-1-k}{2}} BF_{EBL}$$

So if $a \to 2$,[1] the hyper-g Bayes Factor is approximately equivalent to an EBL Bayes factor times a k-based model prior (that does not depend on R_s^2). Moreover, this model prior is bounded in a relatively narrow range: Note that

$$(1+a)^{-a/2} \leq \left(\frac{k+a}{k}\right)^{\frac{k}{2}}\left(\frac{N-1-k-a}{N-1-k}\right)^{\frac{N-1-k}{2}} < 1$$

The upper bound follows from the fact that $((k+a)/k)^{k/2} = (1+\frac{a}{2}/\frac{k}{2})^{\frac{k}{2}} < \exp(\frac{a}{2})$. Similarly $((N-1-k-a)/(N-1-k))^{\frac{N-1-k}{2}} < \exp(-\frac{a}{2})$. Setting $k=1$ and letting $N-1 \geq k+a+1$ performs the lower bound.[2] The effect of the term in square roots actually counters the impact of the latter term, as $\sqrt{\frac{4}{N-1}} \leq \sqrt{\frac{N-1}{(N-1-k-a)(k+a)}} \leq 1$ for $k+a < N-1$. The 'model prior' thus results in an upweighting of models with few or many coefficients, while intermediate model sizes are downweighted (a feature very similar to the model prior of Ley and Steel (2009)). Since the model prior does not depend on the level of N, it will lose importance as $N \to \infty$. In the limit, therefore, both EBL and consistent hyper-g will approach the same Bayes Factors between any model except the null model. If the true model is not the null model, then the posterior model probabilities under EBL will therefore approach those under a consistent hyper-g – which establishes the asymptotic consistency of the EBL prior in the sense of Fernández et al. (2001a), provided the true model is not the null model.

Moreover, note that even in small samples, the impact of the k-based 'model prior' is virtually negligible with respect to the importance of BF_{EBL}. Thus, at least as long as $R^2(N-1) > k+a$, BF_h is quite close to BF_{EBL}. And as long as the signal-to-noise ratio in the data is not too small, BMA posterior statistics will be disproportionally based on models with large PMPs (i.e., models with $(N-1)R^2 \gg k+a$). Any models with large differences between BF_h and BF_{EBL} will thus hardly affect posterior model probabilities. Finally, note that in this case models with high PMPs will display very hypergeometric terms $F_s^* \gg 1$, which renders the posterior moments from sectio 1.4.1 very close to their EBL equivalents. This effect explains why this paper's results under hyper-g and EBL are so close.

1.A.4 The shrinkage factor and goodness-of-fit

In order to demonstrate inequalities (1.9) and (1.10), consider a reformulation of the posterior expected value of the shrinkage factor (1.7), where p_s is shorthand for posterior model probability $p(M_s|y,X)$ of model M_s (and p_0 denotes the PMP of the null model).[3]

$$\mathrm{E}\left(\frac{g}{1+g}\middle| y, X\right) = \epsilon + \sum_{j=1}^{2^K} p_s \frac{\bar{N}R_s^2 - \bar{\theta}_s}{R_s^2(\bar{N}-\bar{\theta}_s)} + p_0\frac{2}{a} \tag{A.4}$$

$$\text{where} \quad \epsilon = \sum_{j=1}^{2^K} p_s \frac{\bar{\theta}_s}{F_s^* R_s^2(\bar{N}-\bar{\theta}_s)}$$

[1] Recall that any consistent hyper-g prior requires $a \to 2$ for $N \to \infty$.

[2] Note that if $k=0$, $BF_{EBL} = BF_h = 1$.

[3] Note that this formulation assumes R_s^2 for all models other than the null model to be strictly larger than zero – a notion we will follow throughout this section. However, the following inequalities can be easily generalized to the case of $R_s^2 = 0$ as long as the the full model has $R_F^2 > 0$.

Equation (A.4) can be transformed into the following:

$$1 - \mathrm{E}\left(\tfrac{g}{1+g}\Big| y, X\right) + \epsilon - \tfrac{a-2}{a} p_0 = \sum_{j=1}^{2^K} p_s \frac{\bar{\theta}_s (1 - R_s^2)}{(\bar{N} - \bar{\theta}_s) R_s^2} \tag{A.5}$$

The ϵ term is based on the expression $\bar{\theta}_s / F_s^*$ in (1.7), whose only role is to keep $E(\frac{g}{1+g}|X, y)$ non-negative in case of a 'bad' model, whereas it rapidly vanishes for models with higher signal-to-noise ratios. For any setting of the hyperparameter a, ϵ vanishes as $N \to \infty$ for fixed K as long as the null model is not the single 'true' model – but even in small samples, ϵ tends rapidly towards zero as data quality increases. Using a consistent hyper-g prior such as HG-UIP, ensures that ϵ will asymptotically vanish even when the true model is the null model, as in this case R_s^2 will vanish with order of magnitude $O_T(\bar{N})$: Since $F_s^* \geq 1$, the model-specific expression is bounded from below while consistency will drive p_s to zero.

Likewise, under a consistent prior the requirement that $a \to 2$ as $\bar{N} \to \infty$ will induce the term $p_0(a-2)$ to vanish asymptotically.

That said, in small samples with any viable signal-to-noise ratio, any models with very low PMP will hardly affect posterior results, and hence the expression ϵ will vanish as soon as there exist some models with posterior model probabilities considerably larger than p_0 (which therefore must have their $F_s^* \gg 1$).

Omitting the expression $\epsilon - \frac{a-2}{a} p_0$ from (A.5) directly leads to inequality (1.9): Define $\hat{F}_s \equiv \frac{(\bar{N} - \bar{\theta}_s) R_s^2}{\bar{\theta}_s (1 - R_s^2)}$, then (1.9) follows from applying Jensen's inequality to (A.5).

A similar argument applies to the demonstration of inequality (1.10). As long as $K + a < N + 1$, the following will hold by Jensen's inequality:

$$\sum_{s=1}^{K^2} p_s \left(\tfrac{1}{\bar{N} - \bar{\theta}_s}\right) \geq \frac{(1 - p_0)^2}{\sum_{s=1}^{K^2} p_s \left(\bar{N} - \bar{\theta}_s\right)}$$

where $\sum_{s=1}^{K^2} p_s$ denotes the model probability-weighted sum over all models except the null model. Multiply with $\bar{N}$ and subtract $(1 - p_0)$ to obtain

$$\sum_{s=1}^{K^2} p_s \left(\tfrac{\theta_s}{\bar{N} - \bar{\theta}_s}\right) \geq (1 - p_0) \frac{\sum_{s=1}^{K^2} p_s \bar{\theta}_s}{\sum_{s=1}^{K^2} p_s \left(\bar{N} - \bar{\theta}_s\right)}$$

Moreover, since any nested model's R-squared R_s^2 cannot exceed the R-squared of the full model R_F^2, we have that $\frac{1 - R_s^2}{R_s^2} \geq \frac{1 - R_F^2}{R_F^2}$ and therefore:

$$\sum_{s=1}^{K^2} p_s \left(\tfrac{1 - R_s^2}{R_s^2} \tfrac{\theta_s}{\bar{N} - \bar{\theta}_s}\right) \geq (1 - p_0) \frac{\sum_{s=1}^{K^2} p_s \bar{\theta}_s}{\sum_{s=1}^{K^2} p_s \left(\bar{N} - \bar{\theta}_s\right)} \frac{1 - R_F^2}{R_F^2}$$

Now note that $\sum_{s=1}^{K^2} p_s \bar{\theta}_s = E(\bar{\theta}|y, X) - p_0(a-2)$. Consistency lets $p_0(a-2)$ rapidly vanish with $N \to \infty$, which ensures that the right-hand side of this inequality is weakly larger than its version in expected value terms $E(\bar{\theta}_s|y, X) \equiv \sum_{s=0}^{K^2} p_s \bar{\theta}_s$:

$$(1-p_0)\frac{\sum_{s=1}^{K^2} p_s\bar{\theta}_s}{\sum_{s=1}^{K^2} p_s\big(\bar{N}-\bar{\theta}_s\big)}\frac{1-R_F^2}{R_F^2} \geq \frac{E\big(\bar{\theta}_s\big|y,X\big)}{E\big(\bar{N}-\bar{\theta}_s\big|y,X\big)}\frac{1-R_F^2}{R_F^2}$$

This establishes an upper bound for the left-hand side of (A.5) (recall that $\bar{N} \equiv N-3$ and $E(\bar{\theta}_s|y,X) = E(k_s|y,X)+a-2$):

$$\mathrm{E}\left(\tfrac{g}{1+g}\Big|y,X\right) - \epsilon + \tfrac{a-2}{a}p_0 \leq 1 - \frac{E\big(\bar{\theta}_s\big|y,X\big)}{E\big(\bar{N}-\bar{\theta}_s\big|y,X\big)}\frac{1-R_F^2}{R_F^2}$$

Under a consistent hyper-g prior, the term $\epsilon - \frac{a-2}{a}p_0$ will vanish asymptotically as $N \to \infty$, which establishes inequality (1.10). How close $E\big(\frac{g}{1+g}\big|y,X\big)$ comes to this upper bound is mainly determined by the posterior variance of model size (the less variance, the closer), and by parsimoniousness of the model priors.Note that the term ϵ on the left-hand side might break the inequality (1.10) in peculiar small samples. However, this term tends to be very small: Numerical simulations of a null hypothesis with varying N, K, a and standard deviations have yielded no single instance in which $R_F^2 > \frac{K+a-2}{\bar{N}}$ and $E(\frac{g}{1+g}|y,X)$ larger than the right-hand side above. Therefore, if $R_F^2 > \frac{K+a}{N}$[1], then the ϵ term can be safely omitted from the inequality above.

1.A.5 Proof of Proposition 1

Given (y,X), define log-likelihood of a model M_γ as $\ell_\gamma \equiv \frac{k}{2}\log(1-s) - \frac{N-1}{2}\log(1-sz_\gamma)$ where $s = \frac{g}{1+g}$ is shorthand for the shrinkage factor, and z_γ denotes the R-squared of model M_γ. Moreover, index models according to their rank at a given s. Denoting the model the prior model probability of model M_γ by m_γ, the PMP of the best r models is given as:

$$PMP_r^* = \sum_{i=1}^{r}\exp(l_i)m_i / \sum_{i=1}^{2^K+1}\exp(l_i)m_i$$

Let $E_r(\theta)$ denote the posterior average of a statistic θ_i weighted according to r posterior model probabilities:

$$E_r(\theta) = \sum_{i=1}^{r}\exp(l_i)m_i\theta_i / \sum_{i=1}^{r}\exp(l_i)m_i$$

Moreover, write the posterior average over all models as $E(\theta) = E_{2^K+1}(\theta)$. Then, to prove proposition 1, it suffices to find the conditions under which the sign of the following derivative is positive:

$$\frac{d\log PMP_r^*}{ds} = E_r\left(\frac{\partial\ell}{\partial s}\right) - E\left(\frac{\partial\ell}{\partial s}\right)$$

LEMMA 1 *The posterior probability of the best model PMP_1^* has a non-negative derivative with respect to s if $k^*+\eta < E(k|y,X)$, where $E(k|y,X)$ represents posterior parameter size, and k^* the parameter size the best model; with $\eta > 0$ vanishing as $N \to \infty$, $s \to \infty$.*

[1] Note that this threshold is just slightly higher than the expected value of R_F^2 under the classic null hypothesis of no significant variance explanation by a regression model. As a rule of thumb, if the standard F-statistic for the full model is 'significant' by at least 20%, then the inequality above is guaranteed to hold.

PROOF Let $\frac{dPMP_1^*}{ds} < 0$, which implies

$$E(k) - k^* < (N-1)\left(\frac{1-z^*}{1-sz^*} - E\left(\frac{1-z}{1-sz}\right)\right)$$

The right-hand side could be positive or negative – although $E(k) > k*$ will in may cases be associated with a positive right-hand side. However, for constant N and $s \to \infty$ the right hand side vanishes rapidly which already proves one part of lemma 1.

Now that since $\exp(\ell_r^*)m_r^* \geq \exp(\ell_i)m_i \; \forall \; i \geq r$ we have that

$$\frac{sz_i}{1-sz_i} \leq \frac{(1-s)^{\frac{k_r-k_i}{N-1}}(\frac{m_r}{m_i})^{\frac{2}{N-1}}}{1-sz_r} - 1 \qquad \forall \; i \geq r \tag{A.6}$$

Therefore, $\frac{dPMP_1^*}{ds} < 0$ implies the following inequality:

$$E(k) - k^* < (N-1)\left(\frac{1-s}{s(1-sz^*)}\left(E((1-s)^{\frac{k^*-k}{N-1}}(\frac{m^*}{m})^{\frac{2}{N-1}} - 1)\right)\right)$$

Whether the right-hand side is positive or negative, depends on the actual distribution of parameter size k. Nonetheless, if $N \to \infty$ under constant s, the $N-1$ term in the exponent will dominate the $N-1$ factor in front, and the right-hand side will go to zero. Moreover, if s is a monotonically increasing function of N, the right-hand side will vanish even more rapidly. Defining η as the right-hand side of inequality (A.6) concludes the proof of lemma 1.

PROOF of proposition 1

Let be $(N-1)\left(E_{r-1}(\frac{1-z}{1-sz}) - E(\frac{1-z}{1-sz})\right)$ be bounded by η_{r-1}. We will have $\frac{dPMP_r^*}{ds} < 0$ if and only if

$$E(k) - E_r(k^*) < (N-1)\left(E_r\left(\frac{1-z^*}{1-sz^*}\right) - E\left(\frac{1-z}{1-sz}\right)\right) \tag{A.7}$$

Denote posterior model probability of model M_i by $p_i = \exp(\ell_i)m_i / \sum_{j=1}^{2^K+1}\exp(\ell_j)m_j$. If $\frac{1-z_r}{1-sz_r} \leq E_{r-1}(\frac{1-z}{1-sz})$, then the right hand side in (A.7) will be bounded by η_{r-1}

If $\frac{1-z_r}{1-sz_r} > E_{r-1}(\frac{1-z}{1-sz})$ then the bound η_{r-1} will not hold necessarily, but:

$$(N-1)\left(E_r\left(\frac{1-z}{1-sz}\right) - E\left(\frac{1-z}{1-sz}\right)\right) < (N-1)(1-s)\sum_{i=r+1}^{2^K+1} p_i\left(\frac{z_i}{1-sz_i} - \frac{z_r}{1-sz_r}\right)$$

By inequality (A.6) the right hand-side is dominated by the right-hand side of the following inequality:

$$(N-1)\left(E_r\left(\frac{1-z}{1-sz}\right) - E\left(\frac{1-z}{1-sz}\right)\right) < (N-1)\frac{1-s}{s(1-sz_r)}\sum_{i=r+1}^{2^K+1} p_i\left((1-s)^{\frac{k_r-k_i}{N-1}}(\frac{m_r}{m_i})^{\frac{2}{N-1}} - 1\right)$$

A similar argument as before is applied: if $s \to \infty$ and/or $N \to \infty$, then the right-hand side will vanish. In case this right-hand side is greater than η_{r-1}, η_r might be redefined appropriately.

1.B Charts and Tables

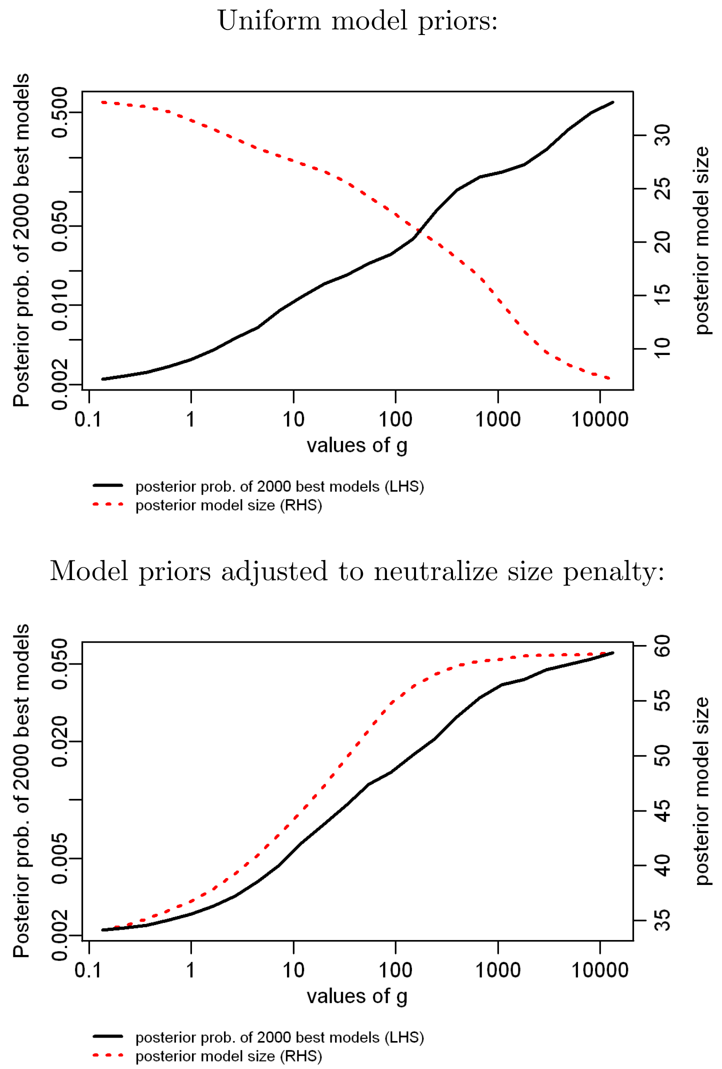


Figure 1.1: Illustration of the supermodel effect: sum of posterior model probabilities (PMP) for the best 2000 models (by PMP) and posterior model size for the PWT 6.3 growth data set (see section 1.6). Top panel shows results under uniform model priors. Bottom panel shows results under model priors that neutralize the size penalty from the g-prior (cf. section 1.3).

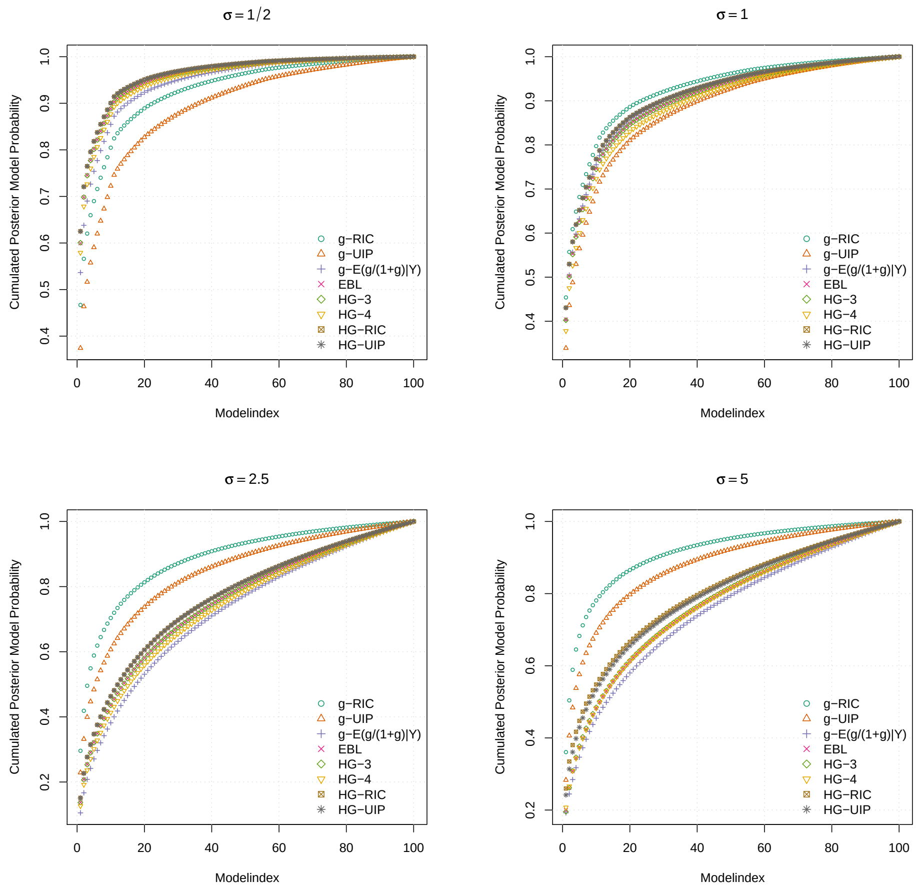


Figure 1.2: Cumulative posterior model probabilities for the best 100 models under Setting A. Top panel corresponds to a noise level of $\sigma = 1/2$ (left) and $\sigma = 1$ (right). Bottom panel corresponds to a ratio of $\sigma = 2.5$ (left) and $\sigma = 5$ (right).

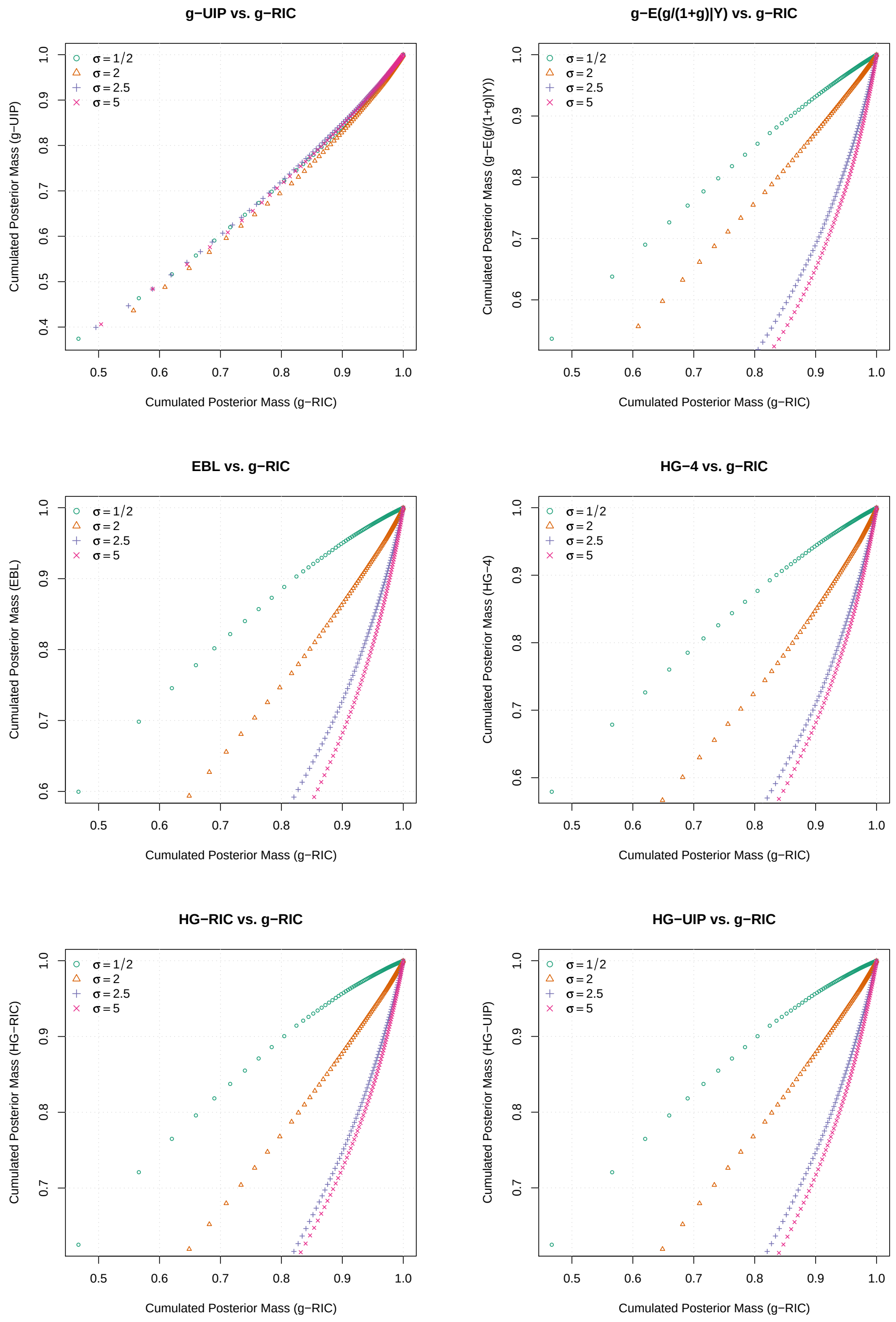


Figure 1.3: QQ-plot of cumulative posterior mass for different g-priors versus that of the g-RIC setting (Setting A, based on 50 Monte Carlo draws).

1

	signal-to-noise ratio of $\sigma = 1/2$								signal-to-noise ratio of $\sigma = 1$							
	g-RIC	g-UIP	g-E($\frac{g}{1+g}\|y$)	EBL	HG-3	HG-4	HG-RIC	HG-UIP	g-RIC	g-UIP	g-E($\frac{g}{1+g}\|y$)	EBL	HG-3	HG-4	HG-RIC	HG-UIP
$\boldsymbol{\beta_1}$	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
β_2	0.085	0.114	0.068	0.055	0.055	0.059	0.050	0.050	0.094	0.132	0.109	0.115	0.116	0.124	0.107	0.107
	(0.072)	(0.069)	(0.067)	(0.058)	(0.055)	(0.058)	(0.051)	(0.051)	(0.125)	(0.128)	(0.127)	(0.128)	(0.127)	(0.128)	(0.127)	(0.127)
β_3	0.083	0.112	0.066	0.054	0.052	0.056	0.047	0.047	0.085	0.125	0.101	0.108	0.109	0.117	0.099	0.099
	(0.067)	(0.061)	(0.064)	(0.055)	(0.046)	(0.049)	(0.043)	(0.043)	(0.073)	(0.084)	(0.078)	(0.081)	(0.080)	(0.083)	(0.077)	(0.077)
β_4	0.077	0.106	0.061	0.050	0.050	0.054	0.046	0.046	0.078	0.117	0.093	0.100	0.101	0.110	0.092	0.092
	(0.065)	(0.064)	(0.062)	(0.060)	(0.059)	(0.060)	(0.057)	(0.057)	(0.057)	(0.068)	(0.062)	(0.063)	(0.063)	(0.065)	(0.060)	(0.060)
$\boldsymbol{\beta_5}$	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
β_6	0.086	0.116	0.068	0.055	0.055	0.060	0.050	0.050	0.067	0.105	0.082	0.089	0.090	0.098	0.081	0.081
	(0.059)	(0.061)	(0.052)	(0.046)	(0.045)	(0.048)	(0.042)	(0.042)	(0.021)	(0.031)	(0.025)	(0.029)	(0.029)	(0.032)	(0.027)	(0.027)
$\boldsymbol{\beta_7}$	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
β_8	0.113	0.136	0.098	0.085	0.082	0.086	0.077	0.077	0.087	0.129	0.104	0.111	0.111	0.121	0.101	0.102
	(0.153)	(0.132)	(0.156)	(0.149)	(0.138)	(0.141)	(0.135)	(0.135)	(0.061)	(0.075)	(0.068)	(0.069)	(0.068)	(0.071)	(0.065)	(0.065)
β_9	0.084	0.114	0.066	0.054	0.054	0.058	0.049	0.049	0.083	0.125	0.099	0.107	0.108	0.117	0.098	0.098
	(0.066)	(0.070)	(0.058)	(0.055)	(0.054)	(0.057)	(0.051)	(0.051)	(0.049)	(0.066)	(0.057)	(0.065)	(0.064)	(0.068)	(0.060)	(0.060)
β_{10}	0.086	0.115	0.068	0.056	0.056	0.060	0.051	0.051	0.100	0.145	0.118	0.126	0.127	0.136	0.116	0.116
	(0.082)	(0.078)	(0.077)	(0.069)	(0.066)	(0.069)	(0.062)	(0.062)	(0.086)	(0.106)	(0.095)	(0.105)	(0.103)	(0.107)	(0.099)	(0.099)
$\boldsymbol{\beta_{11}}$	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
β_{12}	0.085	0.115	0.066	0.054	0.054	0.059	0.049	0.050	0.109	0.151	0.126	0.133	0.133	0.142	0.123	0.123
	(0.054)	(0.056)	(0.048)	(0.043)	(0.042)	(0.044)	(0.040)	(0.040)	(0.122)	(0.134)	(0.128)	(0.131)	(0.129)	(0.131)	(0.126)	(0.126)
$\boldsymbol{\beta_{13}}$	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.992**	**0.992**	**0.992**	**0.992**	**0.992**	**0.992**	**0.992**	**0.992**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.049)	(0.045)	(0.047)	(0.047)	(0.048)	(0.047)	(0.049)	(0.049)
β_{14}	0.079	0.109	0.063	0.052	0.052	0.055	0.047	0.047	0.097	0.136	0.112	0.120	0.120	0.129	0.111	0.111
	(0.069)	(0.065)	(0.067)	(0.064)	(0.063)	(0.064)	(0.061)	(0.061)	(0.124)	(0.130)	(0.127)	(0.129)	(0.127)	(0.129)	(0.125)	(0.125)
β_{15}	0.093	0.122	0.075	0.062	0.062	0.066	0.057	0.057	0.101	0.144	0.119	0.125	0.126	0.135	0.115	0.115
	(0.096)	(0.094)	(0.090)	(0.085)	(0.084)	(0.087)	(0.080)	(0.080)	(0.096)	(0.109)	(0.103)	(0.101)	(0.100)	(0.102)	(0.096)	(0.096)
$E(\frac{g}{1+g}\|Y)$	0.996	0.990	0.998	0.999	0.998	0.998	0.999	0.999	0.996	0.990	0.994	0.993	0.992	0.991	0.993	0.993

Table 1.2: Posterior Inclusion Probabilities for Setting 'A' with standard deviations in parentheses. Left Panel corresponds to a signal-to-noise ratio of $\sigma = 1/2$, right panel to a ratio of $\sigma = 1$. Coefficients corresponding to variables of the data-generating model are highlighted in bold in the left column. PIP values exceeding 0.5 are highlighted in **bold**. Results are averaged over 50 Monte Carlo draws.

	signal-to-noise ratio of $\sigma = 2.5$								signal-to-noise ratio of $\sigma = 5$							
	g-RIC	g-UIP	g-E($\frac{g}{1+g}\|y$)	EBL	HG-3	HG-4	HG-RIC	HG-UIP	g-RIC	g-UIP	g-E($\frac{g}{1+g}\|y$)	EBL	HG-3	HG-4	HG-RIC	HG-UIP
$\boldsymbol{\beta_1}$	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.792**	**0.842**	**0.955**	**0.953**	**0.946**	**0.946**	**0.884**	**0.905**
	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.257)	(0.208)	(0.066)	(0.068)	(0.077)	(0.075)	(0.207)	(0.171)
β_2	0.086	0.131	0.317	0.304	0.304	0.330	0.276	0.276	0.038	0.067	**0.512**	0.455	0.445	0.481	0.370	0.380
	(0.106)	(0.126)	(0.149)	(0.153)	(0.151)	(0.151)	(0.150)	(0.150)	(0.037)	(0.058)	(0.149)	(0.171)	(0.166)	(0.159)	(0.185)	(0.180)
β_3	0.094	0.141	0.331	0.316	0.317	0.343	0.289	0.289	0.050	0.081	**0.514**	0.460	0.450	0.486	0.376	0.386
	(0.064)	(0.077)	(0.109)	(0.115)	(0.113)	(0.114)	(0.111)	(0.111)	(0.052)	(0.074)	(0.160)	(0.180)	(0.176)	(0.167)	(0.197)	(0.192)
β_4	0.108	0.158	0.339	0.327	0.327	0.351	0.300	0.300	0.044	0.074	**0.503**	0.449	0.439	0.476	0.366	0.376
	(0.119)	(0.144)	(0.165)	(0.172)	(0.169)	(0.168)	(0.169)	(0.169)	(0.072)	(0.104)	(0.158)	(0.176)	(0.172)	(0.164)	(0.195)	(0.190)
$\boldsymbol{\beta_5}$	**0.796**	**0.826**	**0.871**	**0.873**	**0.870**	**0.875**	**0.865**	**0.865**	0.184	0.236	**0.617**	**0.579**	**0.567**	**0.594**	0.497	**0.509**
	(0.244)	(0.215)	(0.155)	(0.155)	(0.156)	(0.150)	(0.164)	(0.163)	(0.267)	(0.286)	(0.215)	(0.246)	(0.244)	(0.228)	(0.279)	(0.273)
β_6	0.062	0.103	0.282	0.272	0.273	0.298	0.245	0.245	0.048	0.080	**0.515**	0.463	0.453	0.488	0.380	0.390
	(0.044)	(0.068)	(0.115)	(0.118)	(0.117)	(0.119)	(0.113)	(0.113)	(0.070)	(0.100)	(0.185)	(0.208)	(0.204)	(0.194)	(0.226)	(0.221)
$\boldsymbol{\beta_7}$	**0.997**	**0.999**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.506**	**0.585**	**0.845**	**0.829**	**0.817**	**0.829**	**0.750**	**0.766**
	(0.014)	(0.005)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.342)	(0.340)	(0.183)	(0.208)	(0.212)	(0.195)	(0.284)	(0.268)
β_8	0.075	0.121	0.305	0.292	0.292	0.317	0.265	0.265	0.049	0.080	**0.509**	0.460	0.450	0.486	0.376	0.386
	(0.065)	(0.095)	(0.136)	(0.151)	(0.149)	(0.149)	(0.148)	(0.148)	(0.075)	(0.098)	(0.157)	(0.172)	(0.169)	(0.160)	(0.194)	(0.188)
β_9	0.131	0.187	0.369	0.356	0.356	0.380	0.329	0.330	0.045	0.077	**0.508**	0.455	0.445	0.482	0.372	0.382
	(0.161)	(0.199)	(0.217)	(0.219)	(0.216)	(0.213)	(0.218)	(0.218)	(0.061)	(0.094)	(0.165)	(0.176)	(0.173)	(0.165)	(0.198)	(0.192)
β_{10}	0.082	0.127	0.313	0.302	0.303	0.328	0.275	0.275	0.042	0.072	**0.520**	0.465	0.454	0.491	0.377	0.388
	(0.089)	(0.113)	(0.144)	(0.151)	(0.148)	(0.149)	(0.146)	(0.146)	(0.040)	(0.059)	(0.147)	(0.163)	(0.159)	(0.152)	(0.181)	(0.175)
$\boldsymbol{\beta_{11}}$	**0.828**	**0.869**	**0.925**	**0.922**	**0.919**	**0.924**	**0.913**	**0.913**	0.305	0.369	**0.717**	**0.689**	**0.677**	**0.699**	**0.609**	**0.622**
	(0.240)	(0.205)	(0.141)	(0.146)	(0.147)	(0.141)	(0.156)	(0.156)	(0.316)	(0.323)	(0.222)	(0.246)	(0.246)	(0.229)	(0.294)	(0.285)
β_{12}	0.065	0.109	0.301	0.288	0.289	0.315	0.259	0.260	0.054	0.093	**0.533**	0.480	0.469	**0.504**	0.393	0.404
	(0.036)	(0.053)	(0.105)	(0.109)	(0.107)	(0.111)	(0.102)	(0.102)	(0.067)	(0.109)	(0.171)	(0.196)	(0.192)	(0.182)	(0.217)	(0.211)
$\boldsymbol{\beta_{13}}$	0.437	**0.504**	**0.651**	**0.640**	**0.637**	**0.653**	**0.617**	**0.617**	0.083	0.129	**0.558**	**0.511**	0.500	**0.532**	0.426	0.437
	(0.332)	(0.322)	(0.261)	(0.268)	(0.267)	(0.258)	(0.277)	(0.277)	(0.119)	(0.151)	(0.189)	(0.216)	(0.212)	(0.200)	(0.239)	(0.233)
β_{14}	0.098	0.143	0.321	0.309	0.310	0.334	0.283	0.283	0.043	0.072	**0.504**	0.449	0.439	0.476	0.366	0.375
	(0.129)	(0.146)	(0.157)	(0.163)	(0.161)	(0.159)	(0.162)	(0.162)	(0.058)	(0.083)	(0.160)	(0.176)	(0.172)	(0.164)	(0.191)	(0.187)
β_{15}	0.165	0.212	0.386	0.372	0.373	0.396	0.347	0.347	0.074	0.109	**0.520**	0.470	0.461	0.496	0.387	0.398
	(0.147)	(0.156)	(0.148)	(0.160)	(0.158)	(0.156)	(0.161)	(0.161)	(0.114)	(0.131)	(0.155)	(0.175)	(0.171)	(0.162)	(0.195)	(0.191)
Ek*	5.027	5.629	7.711	7.572	7.568	7.844	7.260	7.264	2.359	2.965	8.830	8.167	8.011	8.467	6.928	7.103
Es*	0.996	0.990	0.949	0.955	0.947	0.939	0.956	0.956	0.996	0.990	0.783	0.817	0.795	0.760	0.856	0.849

Table 1.3: Posterior Inclusion Probabilities for Setting 'A' with standard deviations in parentheses. Left Panel corresponds to a signal-to-noise ratio of $\sigma = 2.5$, right panel to a ratio of $\sigma = 5$ noise. Coefficients corresponding to variables of the data-generating model are highlighted in bold in the left column. PIP values exceeding 0.5 are highlighted in **bold**. Results are averaged over 50 Monte Carlo draws. * 'Ek' refers to posterior expected model size $E(k|Y)$, 'Es' to the posterior expected (average) shrinkage factor $E(\frac{g}{1+g}|Y)$.

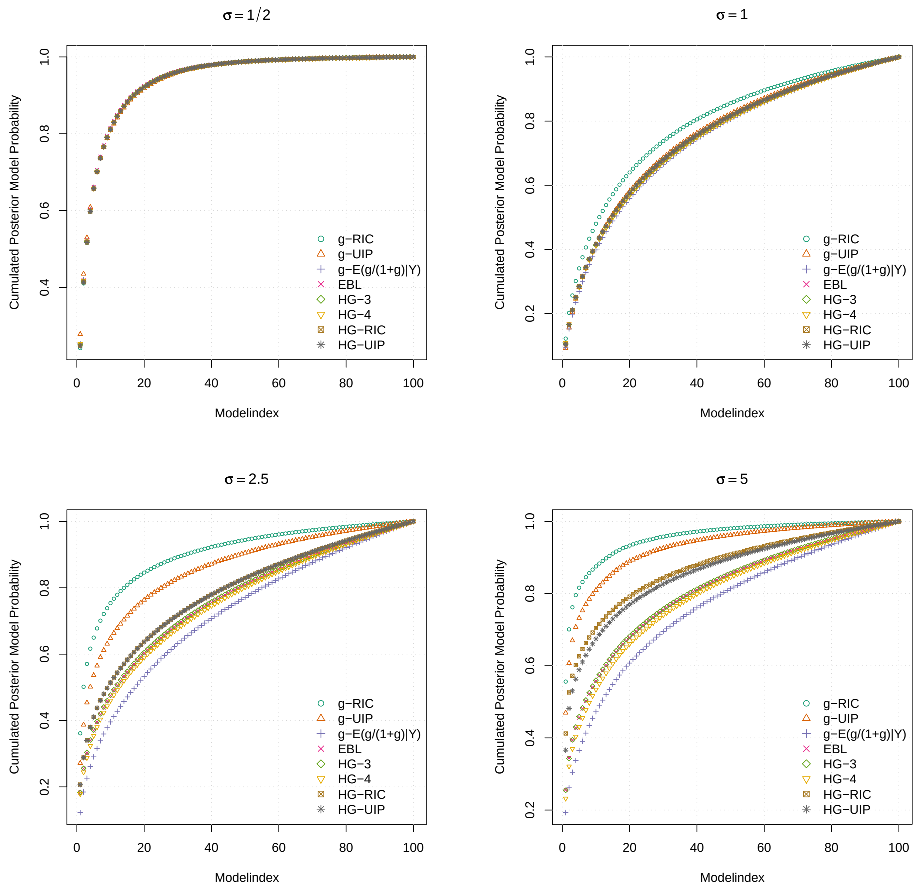


Figure 1.4: Cumulative posterior model probabilities for the best 100 models under Setting 'B'. Top panel corresponds to a noise level of $\sigma = 1/2$ (left) and $\sigma = 1$ (right). Bottom panel corresponds to a ratio of $\sigma = 2.5$ (left) and $\sigma = 5$ (right).

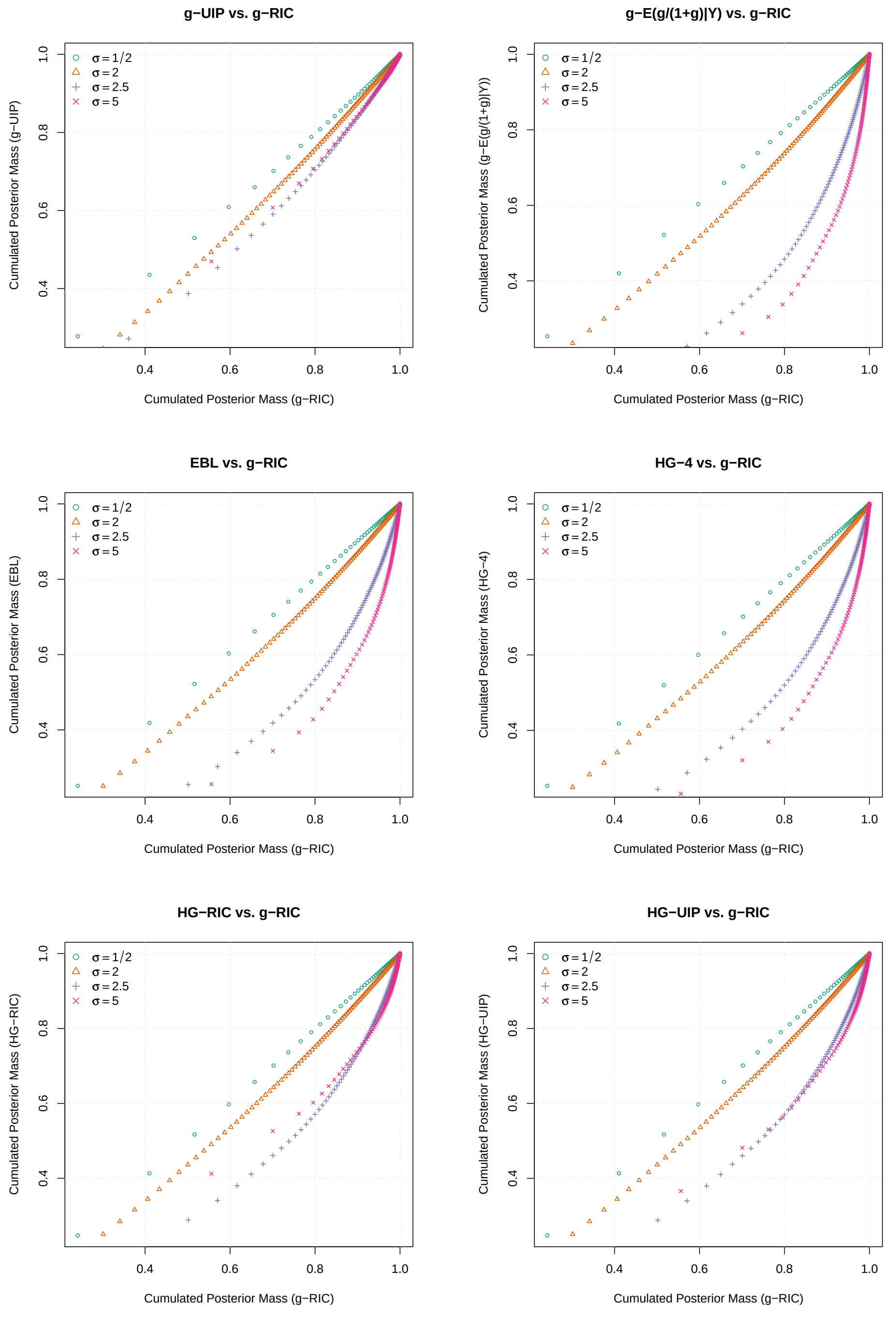


Figure 1.5: QQ-plot of cumulative posterior mass for different g-priors versus that of the g-RIC setting (Setting 'B', based on 50 Monte Carlo draws).
1

	signal-to-noise ratio of $\sigma = 1/2$								signal-to-noise ratio of $\sigma = 1$							
	g-RIC	g-UIP	g-E($\frac{g}{1+g}\|y$)	EBL	HG-3	HG-4	HG-RIC	HG-UIP	g-RIC	g-UIP	g-E($\frac{g}{1+g}\|y$)	EBL	HG-3	HG-4	HG-RIC	HG-UIP
β_1	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
β_2	**0.801**	**0.835**	**0.818**	**0.810**	**0.808**	**0.813**	**0.804**	**0.804**	0.374	0.487	**0.567**	**0.562**	**0.560**	**0.575**	**0.544**	**0.544**
	(0.223)	(0.182)	(0.205)	(0.218)	(0.217)	(0.213)	(0.223)	(0.222)	(0.266)	(0.262)	(0.242)	(0.247)	(0.245)	(0.240)	(0.250)	(0.250)
β_3	**0.762**	**0.802**	**0.781**	**0.773**	**0.772**	**0.777**	**0.767**	**0.767**	0.393	0.480	**0.550**	**0.548**	**0.548**	**0.560**	**0.533**	**0.534**
	(0.239)	(0.195)	(0.220)	(0.228)	(0.227)	(0.222)	(0.232)	(0.232)	(0.295)	(0.270)	(0.243)	(0.248)	(0.247)	(0.241)	(0.253)	(0.253)
β_4	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**	**0.747**	**0.806**	**0.840**	**0.840**	**0.838**	**0.845**	**0.831**	**0.831**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.282)	(0.242)	(0.207)	(0.206)	(0.206)	(0.199)	(0.215)	(0.215)
β_5	**0.885**	**0.897**	**0.891**	**0.889**	**0.888**	**0.889**	**0.886**	**0.886**	0.437	**0.516**	**0.577**	**0.573**	**0.572**	**0.584**	**0.559**	**0.560**
	(0.147)	(0.119)	(0.135)	(0.140)	(0.140)	(0.136)	(0.143)	(0.143)	(0.293)	(0.271)	(0.245)	(0.249)	(0.247)	(0.241)	(0.254)	(0.254)
β_6	**0.943**	**0.952**	**0.948**	**0.947**	**0.946**	**0.947**	**0.944**	**0.944**	0.470	**0.560**	**0.625**	**0.621**	**0.620**	**0.632**	**0.606**	**0.606**
	(0.116)	(0.093)	(0.106)	(0.107)	(0.108)	(0.105)	(0.111)	(0.111)	(0.332)	(0.303)	(0.272)	(0.277)	(0.276)	(0.269)	(0.283)	(0.283)
β_7	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.980**	**0.988**	**0.991**	**0.990**	**0.990**	**0.990**	**0.989**	**0.989**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.039)	(0.023)	(0.018)	(0.020)	(0.021)	(0.019)	(0.022)	(0.022)
β_8	**0.696**	**0.749**	**0.720**	**0.712**	**0.711**	**0.718**	**0.704**	**0.704**	0.325	0.430	**0.509**	**0.506**	**0.506**	**0.520**	0.489	0.490
	(0.234)	(0.190)	(0.216)	(0.221)	(0.220)	(0.215)	(0.225)	(0.225)	(0.237)	(0.242)	(0.227)	(0.233)	(0.231)	(0.227)	(0.236)	(0.236)
β_9	**0.679**	**0.734**	**0.704**	**0.694**	**0.693**	**0.700**	**0.686**	**0.686**	0.255	0.364	0.452	0.449	0.449	0.465	0.431	0.431
	(0.256)	(0.211)	(0.238)	(0.247)	(0.246)	(0.241)	(0.252)	(0.252)	(0.189)	(0.196)	(0.188)	(0.197)	(0.195)	(0.192)	(0.198)	(0.198)
β_{10}	**0.998**	**0.998**	**0.999**	**0.999**	**0.998**	**0.998**	**0.998**	**0.998**	**0.737**	**0.804**	**0.841**	**0.838**	**0.836**	**0.843**	**0.828**	**0.828**
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.271)	(0.225)	(0.190)	(0.192)	(0.193)	(0.186)	(0.200)	(0.200)
β_{11}	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.997**	**0.998**	**0.999**	**0.999**	**0.998**	**0.999**	**0.998**	**0.998**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.022)	(0.013)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)
β_{12}	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**	**0.671**	**0.754**	**0.804**	**0.795**	**0.793**	**0.802**	**0.782**	**0.782**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.300)	(0.248)	(0.207)	(0.217)	(0.218)	(0.210)	(0.227)	(0.227)
β_{13}	**0.989**	**0.990**	**0.989**	**0.989**	**0.989**	**0.989**	**0.989**	**0.989**	**0.728**	**0.773**	**0.804**	**0.801**	**0.800**	**0.806**	**0.793**	**0.794**
	(0.053)	(0.043)	(0.048)	(0.050)	(0.050)	(0.049)	(0.052)	(0.052)	(0.303)	(0.265)	(0.233)	(0.236)	(0.235)	(0.229)	(0.243)	(0.243)
β_{14}	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.993**	**0.996**	**0.997**	**0.996**	**0.996**	**0.996**	**0.996**	**0.996**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.030)	(0.021)	(0.016)	(0.018)	(0.018)	(0.017)	(0.020)	(0.020)
β_{15}	0.486	**0.575**	**0.524**	**0.508**	**0.509**	**0.519**	0.499	0.499	0.285	0.380	0.460	0.457	0.457	0.472	0.441	0.441
	(0.191)	(0.155)	(0.177)	(0.185)	(0.183)	(0.179)	(0.187)	(0.187)	(0.201)	(0.194)	(0.180)	(0.185)	(0.184)	(0.180)	(0.188)	(0.188)
$E(k\|Y)$	13.238	13.530	13.373	13.318	13.313	13.348	13.275	13.275	9.391	10.335	11.015	10.976	10.964	11.089	10.822	10.824
$E(\frac{g}{1+g}\|Y)$	0.996	0.990	0.994	0.995	0.994	0.993	0.994	0.994	0.996	0.990	0.982	0.983	0.980	0.978	0.982	0.982

Table 1.4: Posterior Inclusion Probabilities for Setting 'B' with standard deviations in parentheses. Left Panel corresponds to a signal-to-noise ratio of $\sigma = 1/2$, right panel to a ratio of $\sigma = 1$. PIP values exceeding 0.5 are highlighted in **bold**. Results are averaged over 50 Monte Carlo draws.

	signal-to-noise ratio of $\sigma = 2.5$								signal-to-noise ratio of $\sigma = 5$							
	g-RIC	g-UIP	g-E($\frac{g}{1+g}\|y$)	EBL	HG-3	HG-4	HG-RIC	HG-UIP	g-RIC	g-UIP	g-E($\frac{g}{1+g}\|y$)	EBL	HG-3	HG-4	HG-RIC	HG-UIP
β_1	**0.999**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.819**	**0.851**	**0.944**	**0.940**	**0.930**	**0.927**	**0.804**	**0.857**
	(0.004)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.274)	(0.244)	(0.113)	(0.109)	(0.115)	(0.107)	(0.239)	(0.192)
β_2	0.060	0.105	0.355	0.377	0.373	0.409	0.327	0.328	0.030	0.048	0.412	0.392	0.376	0.423	0.261	0.282
	(0.059)	(0.096)	(0.179)	(0.196)	(0.193)	(0.188)	(0.197)	(0.197)	(0.050)	(0.068)	(0.163)	(0.183)	(0.174)	(0.168)	(0.171)	(0.172)
β_3	0.112	0.165	0.414	0.432	0.426	0.460	0.382	0.382	0.040	0.063	0.422	0.399	0.384	0.428	0.264	0.288
	(0.150)	(0.170)	(0.200)	(0.211)	(0.209)	(0.203)	(0.213)	(0.213)	(0.069)	(0.094)	(0.184)	(0.206)	(0.197)	(0.187)	(0.177)	(0.184)
β_4	0.055	0.105	0.376	0.394	0.389	0.426	0.342	0.342	0.033	0.053	0.436	0.412	0.395	0.442	0.276	0.298
	(0.060)	(0.114)	(0.208)	(0.225)	(0.221)	(0.216)	(0.224)	(0.224)	(0.049)	(0.066)	(0.168)	(0.186)	(0.177)	(0.170)	(0.178)	(0.178)
β_5	0.056	0.093	0.332	0.355	0.351	0.387	0.306	0.307	0.028	0.047	0.420	0.397	0.382	0.428	0.263	0.286
	(0.066)	(0.083)	(0.159)	(0.185)	(0.183)	(0.179)	(0.186)	(0.186)	(0.025)	(0.039)	(0.168)	(0.187)	(0.178)	(0.173)	(0.169)	(0.171)
β_6	0.083	0.128	0.374	0.396	0.391	0.428	0.344	0.345	0.024	0.040	0.395	0.372	0.357	0.404	0.239	0.262
	(0.154)	(0.174)	(0.210)	(0.216)	(0.213)	(0.207)	(0.218)	(0.218)	(0.034)	(0.051)	(0.149)	(0.180)	(0.170)	(0.163)	(0.148)	(0.157)
β_7	0.230	0.304	**0.547**	**0.555**	**0.549**	**0.579**	**0.508**	**0.508**	0.062	0.086	0.446	0.423	0.407	0.451	0.288	0.312
	(0.262)	(0.298)	(0.292)	(0.299)	(0.296)	(0.283)	(0.311)	(0.311)	(0.144)	(0.165)	(0.193)	(0.216)	(0.209)	(0.199)	(0.209)	(0.210)
β_8	0.057	0.095	0.335	0.359	0.355	0.392	0.308	0.309	0.024	0.040	0.396	0.375	0.360	0.407	0.245	0.267
	(0.072)	(0.103)	(0.172)	(0.188)	(0.185)	(0.183)	(0.185)	(0.185)	(0.027)	(0.042)	(0.157)	(0.183)	(0.174)	(0.168)	(0.166)	(0.168)
β_9	0.056	0.097	0.332	0.356	0.353	0.388	0.309	0.309	0.025	0.042	0.397	0.379	0.364	0.411	0.250	0.271
	(0.063)	(0.098)	(0.179)	(0.193)	(0.190)	(0.186)	(0.193)	(0.193)	(0.029)	(0.046)	(0.158)	(0.175)	(0.166)	(0.162)	(0.160)	(0.161)
β_{10}	0.107	0.154	0.402	0.424	0.419	0.453	0.376	0.376	0.046	0.071	0.435	0.417	0.401	0.446	0.286	0.308
	(0.193)	(0.210)	(0.235)	(0.252)	(0.249)	(0.241)	(0.255)	(0.255)	(0.072)	(0.102)	(0.197)	(0.209)	(0.202)	(0.193)	(0.206)	(0.205)
β_{11}	**0.503**	**0.580**	**0.764**	**0.768**	**0.761**	**0.781**	**0.727**	**0.728**	0.138	0.170	**0.524**	**0.501**	0.484	**0.525**	0.360	0.387
	(0.352)	(0.344)	(0.269)	(0.260)	(0.261)	(0.245)	(0.283)	(0.282)	(0.243)	(0.257)	(0.236)	(0.245)	(0.241)	(0.226)	(0.249)	(0.250)
β_{12}	0.115	0.173	0.433	0.451	0.445	0.480	0.399	0.399	0.040	0.065	0.433	0.410	0.394	0.440	0.274	0.297
	(0.164)	(0.197)	(0.241)	(0.243)	(0.239)	(0.234)	(0.243)	(0.243)	(0.067)	(0.098)	(0.173)	(0.196)	(0.188)	(0.179)	(0.187)	(0.187)
β_{13}	0.126	0.178	0.411	0.430	0.426	0.458	0.384	0.384	0.073	0.094	0.436	0.415	0.400	0.444	0.284	0.307
	(0.175)	(0.201)	(0.227)	(0.243)	(0.240)	(0.232)	(0.246)	(0.246)	(0.178)	(0.188)	(0.187)	(0.211)	(0.205)	(0.193)	(0.212)	(0.212)
β_{14}	0.234	0.318	**0.600**	**0.600**	**0.592**	**0.625**	**0.546**	**0.547**	0.051	0.080	0.475	0.448	0.431	0.475	0.308	0.332
	(0.273)	(0.291)	(0.276)	(0.286)	(0.285)	(0.269)	(0.304)	(0.304)	(0.072)	(0.097)	(0.198)	(0.216)	(0.208)	(0.198)	(0.208)	(0.209)
β_{15}	0.045	0.083	0.327	0.350	0.346	0.383	0.300	0.300	0.033	0.051	0.410	0.391	0.376	0.422	0.260	0.282
	(0.039)	(0.066)	(0.151)	(0.169)	(0.167)	(0.163)	(0.170)	(0.170)	(0.073)	(0.092)	(0.159)	(0.181)	(0.173)	(0.168)	(0.173)	(0.173)
Ek*	2.837	3.577	7.001	7.244	7.173	7.649	6.556	6.564	1.465	1.803	6.981	6.670	6.444	7.073	4.662	5.036
Es*	0.996	0.990	0.933	0.926	0.913	0.897	0.931	0.931	0.996	0.990	0.782	0.796	0.767	0.716	0.866	0.850

Table 1.5: Posterior Inclusion Probabilities for Setting 'B' with standard deviations in parenthesis. Left Panel corresponds to a signal-to-noise ratio of $\sigma = 2.5$, right panel to a ratio of $\sigma = 5$ noise. PIP values exceeding 0.5 are highlighted in **bold**. Results are averaged over 50 Monte Carlo draws.
* 'Ek' refers to posterior expected model size $E(k|Y)$, 'Es' to the posterior expected (average) shrinkage factor $E(\frac{g}{1+g}|Y)$.

		g-RIC	g-UIP	g-E($\frac{g}{1+g}$|y)	EBL	HG-3	HG-4	HG-RIC	HG-UIP
$\sigma=\frac{1}{2}$	Min.	0.1349	0.1279	0.1573	0.1888	0.1934	0.1809	0.2096	0.2094
	Mean	0.4618	0.3725	0.5306	0.5951	0.5973	0.5758	0.6220	0.6217
	Max.	0.6297	0.5106	0.7037	0.7669	0.7644	0.7461	0.7845	0.7843
	St.Dev.	0.1342	0.1019	0.1482	0.1551	0.1490	0.1487	0.1484	0.1484
$\sigma=1$	Min.	0.0539	0.0317	0.0426	0.0308	0.0320	0.0289	0.0362	0.0362
	Mean	0.4433	0.3290	0.3922	0.3944	0.3932	0.3690	0.4219	0.4215
	Max.	0.6115	0.4849	0.5578	0.5954	0.5922	0.5658	0.6220	0.6216
	St.Dev.	0.1373	0.1138	0.1283	0.1358	0.1342	0.1293	0.1392	0.1392
$\sigma=2.5$	Min.	0.0021	0.0022	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Mean	0.1201	0.1048	0.0556	0.0665	0.0660	0.0606	0.0721	0.0720
	Max.	0.4609	0.3392	0.1493	0.1978	0.1968	0.1768	0.2216	0.2213
	St.Dev.	0.1133	0.0834	0.0387	0.0487	0.0478	0.0441	0.0524	0.0524
$\sigma=5$	Min.	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Mean	0.0012	0.0023	0.0010	0.0025	0.0024	0.0021	0.0027	0.0028
	Max.	0.0215	0.0324	0.0114	0.0323	0.0308	0.0254	0.0347	0.0347
	St.Dev.	0.0035	0.0057	0.0026	0.0062	0.0060	0.0051	0.0067	0.0067

Table 1.6: Summary statistics of posterior model probabilities for true model based on setting 'A' and 50 Monte Carlo Steps. Top panel corresponds to $\sigma = 1/2$, second panel to $\sigma = 1$, third panel to $\sigma = 2.5$, fourth panel to $\sigma = 5$.

		g-RIC	g-UIP	g-E($\frac{g}{1+g}$|y)	EBL	HG-3	HG-4	HG-UIP	HG-RIC
$\sigma=\frac{1}{2}$	Min.	0.4752	0.6133	0.4704	0.5192	0.5386	0.5175	0.5695	0.5691
	Mean	0.9806	0.9919	0.9807	0.9872	0.9908	0.9902	0.9914	0.9914
	Max.	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	St.Dev.	0.1342	0.1019	0.1482	0.1551	0.1490	0.1487	0.1484	0.1484
$\sigma=1$	Min.	0.1363	0.1115	0.1226	0.1131	0.1188	0.1158	0.1239	0.1238
	Mean	0.9650	0.9552	0.9604	0.9612	0.9626	0.9604	0.9657	0.9656
	Max.	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	St.Dev.	0.1373	0.1138	0.1283	0.1358	0.1342	0.1293	0.1392	0.1392
$\sigma=2.5$	Min.	0.0071	0.0076	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Mean	0.4683	0.5202	0.5516	0.5325	0.5331	0.5274	0.5382	0.5383
	Max.	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	St.Dev.	0.1133	0.0834	0.0387	0.0487	0.0478	0.0441	0.0524	0.0524
$\sigma=5$	Min.	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Mean	0.0067	0.0173	0.0070	0.0203	0.0200	0.0147	0.0284	0.0285
	Max.	0.1643	0.3744	0.0842	0.2835	0.2737	0.1685	0.4719	0.4688
	St.Dev.	0.0035	0.0057	0.0026	0.0062	0.0060	0.0051	0.0067	0.0067

Table 1.7: Ratio of posterior model probability for true model divided by the PMP of the best model (summary statistics for setting 'A', based on 50 Monte Carlo draws). Top panel corresponds to $\sigma = 1/2$, second panel to $\sigma = 1$, third panel to $\sigma = 2.5$, fourth panel to $\sigma = 5$.

	g-RIC	g-UIP	g-E($\frac{g}{1+g}$|y)	EBL	HG-3	HG-4	HG-RIC	HG-UIP
$\sigma=1/2$	-	1.00877	0.99754	0.99798	0.99793	0.99817	0.99771	0.99771
$\sigma=1$	-	1.00347	1.00200	1.00128	1.00219	1.00315	1.00126	1.00127
$\sigma=2.5$	-	0.99501	1.00079	1.00556	1.00320	1.00699	1.00039	1.00042
$\sigma=5$	-	0.99034	1.00697	1.00594	1.00720	1.01958	1.01256	1.00692
$\sigma=1/2$	-	0.99754	0.99926	0.99794	0.99948	0.99910	0.99998	0.99998
$\sigma=1$	-	0.98166	0.97396	0.97316	0.97501	0.97398	0.97648	0.97647
$\sigma=2.5$	-	0.98875	0.97284	0.96760	0.97580	0.97847	0.97631	0.97627
$\sigma=5$	-	0.99578	1.00427	0.99968	1.00747	1.02216	1.01853	1.00976

Table 1.8: Root mean squared errors relative to g-RIC, averaged over 50 Monte Carlo draws of data (y, X) and based on 30 out of forecasts over random sample splits of data under each draw. Values below 1 indicate predictive performance that is superior to the g-RIC setting. Top panel corresponds to setting 'A' and bottom panel to setting 'B'.

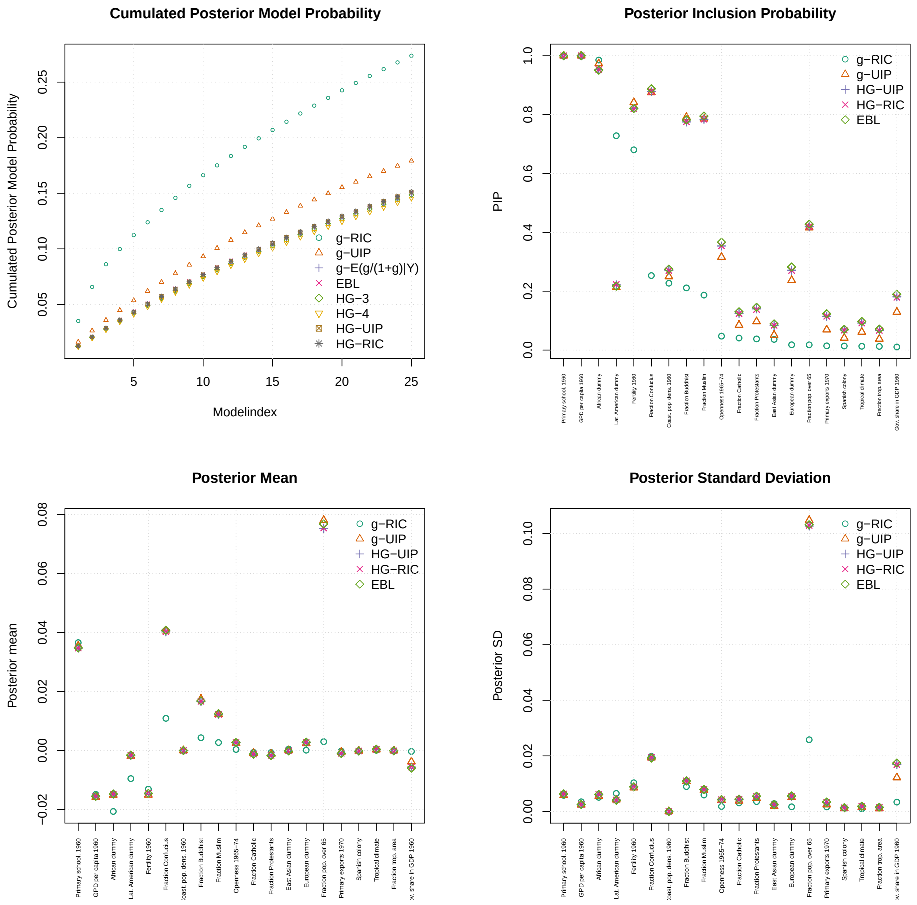


Figure 1.6: Estimation results for the PWT 6.1 data set from section 1.6: Top left panel shows the cumulative posterior mass, left panel the posterior inclusion probabilities for the most important 20 variables, bottom left panel the (expected) standardized coefficients and bottom right panel these standardized coefficients' posterior standard deviation.

	PWT 6.0	PWT 6.1	PWT 6.2	PWT 6.3
g-RIC	0.01384	0.01357	0.01123	0.01476
g-UIP	0.01359	0.01227	0.01070	0.01502
HG-UIP	0.01336	0.01199	0.01053	0.01473
HG-RIC	0.01336	0.01198	0.01054	0.01471
EBL	0.01344	0.01202	0.01061	0.01475

Table 1.9: Root mean squared errors (RMSE) based on 30 random data partitions, as described in section 1.6. Each random sample split assigns 75% as training sample retaining 25% of the data for the forecast evaluation.

	g=K^2	g=N	hyperUIP	hyperBRIC	EBL
PWT 6.0 vs. 6.1	5.4655	1.7971	1.5700	1.5714	1.5580
PWT 6.0 vs. 6.2	4.0735	1.9079	1.6364	1.6336	1.6116
PWT 6.0 vs. 6.3	5.5748	2.6756	1.9654	1.9715	1.9536
PWT 6.1 vs. 6.2	1.6936	1.5184	1.3208	1.3198	1.3037
PWT 6.1 vs. 6.3	2.0679	1.9518	1.6568	1.6573	1.6484
PWT 6.2 vs. 6.3	1.7748	2.2089	1.7516	1.7468	1.7334
PWT Overall Max / Min Ratio	7.3494	3.3100	2.3961	2.3938	2.3565

Table 1.10: Average PIP Max/Min ratios: for each revision pair, these are the mean of the ratio maximum vs. minimum PIP per variable (over all 67 variables).

PWT Revision	g-RIC 6.0	g-RIC 6.1	g-RIC 6.2	g-RIC 6.3	hyper-g (UIP) 6.0	hyper-g (UIP) 6.1	hyper-g (UIP) 6.2	hyper-g (UIP) 6.3
East Asian Dummy	0.85	0.04	0.22	0.18	0.12	0.08	0.11	0.18
Primary Schooling in 1960	0.76	1.00	0.96	1.00	1.00	1.00	1.00	1.00
Fraction of Tropical Area	0.60	0.01	0.04	0.01	0.13	0.07	0.07	0.08
African Dummy	0.23	0.99	0.94	0.96	0.76	0.95	0.91	0.82
GDP in 1960 (log)	0.16	1.00	0.96	1.00	0.95	1.00	1.00	1.00
Latin American Dummy	0.13	0.73	0.36	0.18	0.29	0.22	0.29	0.19
Malaria Prevalence in 1960s	0.13	0.00	0.01	0.01	0.08	0.09	0.10	0.25
Population Density Coastal in 1960s	0.10	0.23	0.03	0.04	0.35	0.27	0.10	0.14
Fraction Buddhist	0.06	0.21	0.17	0.11	0.63	0.78	0.60	0.36
Spanish Colony	0.05	0.01	0.03	0.03	0.29	0.07	0.09	0.11
Higher Education 1960	0.04	0.00	0.00	0.00	0.18	0.06	0.06	0.09
Fraction Confucius	0.04	0.25	0.24	0.20	0.72	0.88	0.80	0.82
Fertility in 1960s	0.02	0.68	0.91	0.91	0.38	0.82	0.88	0.77
Nominal Government GDP Share 1960s	0.02	0.01	0.01	0.00	0.72	0.28	0.38	0.25
Primary Exports 1970	0.02	0.01	0.01	0.09	0.09	0.11	0.11	0.64
Real Exchange Rate Distortions	0.02	0.00	0.00	0.00	0.53	0.07	0.05	0.08
Life Expectancy in 1960	0.02	0.00	0.01	0.00	0.06	0.05	0.05	0.07
Fraction Population In Tropics	0.01	0.00	0.01	0.01	0.07	0.06	0.06	0.08
Openness measure 1965-74	0.01	0.05	0.03	0.09	0.29	0.35	0.24	0.60
Colony Dummy	0.01	0.00	0.00	0.00	0.09	0.07	0.05	0.08
Civil Liberties	0.01	0.00	0.00	0.00	0.14	0.06	0.05	0.06
Fraction Protestants	0.01	0.04	0.04	0.01	0.28	0.14	0.17	0.13
Absolute Latitude	0.01	0.01	0.01	0.01	0.12	0.10	0.10	0.14
Fraction Catholic	0.01	0.04	0.05	0.02	0.29	0.12	0.14	0.13
Years Open 1950-94	0.01	0.01	0.01	0.02	0.13	0.07	0.08	0.17
European Dummy	0.01	0.02	0.01	0.01	0.31	0.27	0.14	0.15
Fraction Muslim	0.01	0.19	0.09	0.05	0.43	0.78	0.58	0.56
Fraction Population Over 65	0.01	0.02	0.01	0.01	0.31	0.42	0.26	0.15
Population Growth Rate 1960-90	0.00	0.01	0.01	0.01	0.09	0.08	0.08	0.14
Fraction Hindus	0.00	0.00	0.00	0.00	0.13	0.09	0.07	0.16
Government Share of GDP in 1960s	0.00	0.01	0.00	0.00	0.12	0.18	0.08	0.14
Air Distance to Big Cities	0.00	0.01	0.01	0.03	0.11	0.09	0.11	0.10
Fraction Population Less than 15	0.00	0.01	0.01	0.01	0.11	0.09	0.08	0.10
Gov. Consumption Share 1960s	0.00	0.01	0.00	0.00	0.10	0.10	0.06	0.11
Fraction GDP in Mining	0.00	0.00	0.00	0.00	0.07	0.05	0.05	0.07
Investment Price	0.00	0.00	0.00	0.00	0.06	0.08	0.05	0.07
Timing of Independence	0.00	0.01	0.01	0.01	0.08	0.08	0.13	0.40
Fraction Speaking Foreign Language	0.00	0.00	0.00	0.00	0.07	0.05	0.06	0.07
Ethnolinguistic Fractionalization	0.00	0.00	0.00	0.00	0.06	0.11	0.07	0.07
Population Density 1960	0.00	0.00	0.00	0.00	0.07	0.08	0.06	0.09
Defence Spending Share	0.00	0.00	0.01	0.00	0.09	0.10	0.11	0.28
Political Rights	0.00	0.00	0.00	0.00	0.10	0.06	0.05	0.08
Population in 1960	0.00	0.00	0.00	0.00	0.07	0.06	0.05	0.10
War Participation 1960-90	0.00	0.00	0.00	0.00	0.06	0.06	0.07	0.06
Tropical Climate Zone	0.00	0.01	0.01	0.01	0.07	0.09	0.12	0.12
Fraction Orthodox	0.00	0.00	0.00	0.00	0.07	0.09	0.08	0.07
Square of Inflation 1960-90	0.00	0.00	0.01	0.01	0.06	0.06	0.08	0.09
Average Inflation 1960-90	0.00	0.00	0.01	0.02	0.06	0.06	0.10	0.09
English Speaking Population	0.00	0.00	0.00	0.00	0.06	0.05	0.05	0.08
Land Area	0.00	0.00	0.00	0.00	0.15	0.18	0.09	0.11
Terms of Trade Ranking	0.00	0.00	0.00	0.00	0.06	0.05	0.05	0.07
Public Education Spending Share in GDP in 1960s	0.00	0.00	0.00	0.00	0.08	0.05	0.05	0.07
Religion Measure	0.00	0.00	0.00	0.00	0.08	0.09	0.06	0.09
Revolutions and Coups	0.00	0.00	0.00	0.00	0.07	0.07	0.07	0.25
Landlocked Country Dummy	0.00	0.01	0.00	0.02	0.10	0.09	0.06	0.24
Fraction of Land Area Near Navigable Water	0.00	0.00	0.00	0.00	0.10	0.09	0.06	0.08
Size of Economy	0.00	0.00	0.00	0.01	0.07	0.10	0.07	0.11
Public Investment Share	0.00	0.00	0.00	0.00	0.07	0.05	0.05	0.07
Socialist Dummy	0.00	0.01	0.00	0.01	0.08	0.19	0.07	0.34
Oil Producing Country Dummy	0.00	0.00	0.00	0.00	0.06	0.05	0.06	0.07
Outward Orientation	0.00	0.00	0.00	0.01	0.07	0.05	0.05	0.07
Hydrocarbon Deposits in 1993	0.00	0.01	0.01	0.01	0.09	0.24	0.17	0.61
British Colony Dummy	0.00	0.00	0.00	0.00	0.07	0.05	0.05	0.06
Capitalism	0.00	0.00	0.00	0.00	0.06	0.10	0.06	0.36
Terms of Trade Growth in 1960s	0.00	0.00	0.00	0.00	0.07	0.05	0.05	0.07
Interior Density	0.00	0.00	0.00	0.00	0.05	0.05	0.04	0.11
Fraction Spent in War 1960-90	0.00	0.00	0.00	0.00	0.06	0.05	0.05	0.07
# Regressors	3.42	5.70	5.30	5.17	12.81	12.60	11.26	14.40
$E(g/(1+g)\|Y)$	0.9998	0.9998	0.9998	0.9998	0.9594	0.9665	0.9625	0.9630

Table 1.11: Posterior inclusion probabilities over PWT revisions. Left panel corresponds to fixed g=K^2, right panel to hyper-g (UIP).

2

The Impact of Data Revisions on the Robustness of Growth Determinants – A Note on 'Determinants of Economic Growth. Will Data Tell?'

Co-authored with:
Martin Feldkircher, *Oesterreichische Nationalbank*

Abstract
Ciccone and Jarociński (2010) show that inference in Bayesian Model Averaging (BMA) can be highly sensitive to small data perturbations. In particular they demonstrate that the importance attributed to potential growth determinants varies tremendously over different revisions of international income data. They conclude that 'agnostic' priors appear too sensitive for this strand of growth empirics. In response, we show that the found instability owes much to a specific BMA set-up: First, comparing the *same* countries over data revisions improves robustness. Second, much of the remaining variation can be reduced by applying an evenly 'agnostic', but flexible prior.

Keywords: Bayesian model averaging, Growth determinants, Zellner's g prior, Model uncertainty.

JEL Classifications: C11, C15, E01, O47.

2.1 Introduction

Ciccone and Jarociński (2010) present an intriguing paper that points to severe weaknesses of Bayesian Model Averaging (BMA). In particular, they criticize the appealing use of non-informative ('agnostic') priors as being plagued by robustness problems. Based on the ubiquitous growth dataset by Sala-i-Martin et al. (2004), the authors analyze the impact of data revisions employing three different vintages of international income data provided by the Penn World Tables (PWT). In this manner, Ciccone and Jarociński (2010) show that small data perturbations can lead to striking differences in posterior inclusion probabilities (PIP) – i.e. the importance of a covariate in explaining the data. The variable 'investment price', for instance, seems very important for growth with a PIP of 0.98 based on PWT 6.1 data, but it exhibits a mere 0.02 under PWT 6.2. Further variables with a worrisome degree of PIP variation reported in Ciccone and Jarociński (2010) are the fraction of Confucian population, population density in 1960, population density in coastal areas in 1960, the fertility rate in 1960, a dummy for east Asia, a dummy for African countries and the fraction of tropical area per country. The conclusions made in Ciccone and Jarociński (2010, p.222) are thus rather pessimistic in that *"margins of error in international income estimates are too large for agnostic growth empirics."*

In a replication exercise we confirm the view that results from 'agnostic' BMA have to be interpreted with caution. In the following sections, however, we point to two caveats to the above findings: First, a considerable share of the PIP variation found by the authors is due to changing sample size and composition over PWT revisions. Secondly, the remaining variation can be attributed to the use of a 'default' prior framework that embodies overly confident prior beliefs – this second caveat constitutes the focus of this note. In response to Ciccone and Jarociński (2010), we show that conditioning on the same country set through all revisions and employing an alternative, flexible prior greatly reduces PIP instability.

2.2 Robustness under Data Revisions

Ciccone and Jarociński (2010) use three different PWT revisions in order to represent income, and cross-check the results with World Bank data. The conditional convergence growth regression employed is of the following form:

$$\Delta y^j = \alpha + \gamma y_0^j + \vec{\beta}_s X_s + \varepsilon, \tag{1}$$

with Δy^j denoting the average annual growth of income per capita over the period from 1960 to 1996 for N countries, α the intercept term and $\mathbf{X}_s = (\mathbf{x}_1 \dots \mathbf{x}_s)$ a matrix whose columns are stacked data for s explanatory variables and ε an error term. Initial income (y_0^j) and income growth (Δy^j) are the only variables that change with revisions. The remaining potential growth determinants grouped in $\mathbf{X}_s$ are the ones originally put forward in Sala-i-Martin et al. (2004) and employed in Ciccone and Jarociński (2010). These 66 variables comprise measures for factor accumulation and convergence (as implied by the Solow growth model), human capital variables, variables measuring political stability and socio-geographical variables. The estimation is carried out for each of the three considered PWT revisions, indexed by $j \in \{\text{PWT 6.0}, \text{PWT 6.1 and PWT 6.2}\}$. In what follows, we denote by σ^2 the variance, N the total number of observations, R_s^2 the OLS R-squared of model s and K the total number of available covariates.

Turning to the econometric framework, Ciccone and Jarociński (2010) use two approaches from the model averaging literature: The 'BACE' methodology proposed by Sala-i-Martin et al. (2004) as well as the popular 'benchmark' BMA type employed by Fernández et al. (2001a). Since both yield broadly similar results in the empirical application by Ciccone and Jarociński (2010, p.223), we follow the purely Bayesian approach akin to Fernández et al. (2001a). This framework relies on Zellner's g prior for the coefficient vector $\vec{\theta}_s|\sigma^2 \sim N\left(\vec{0}_s, \sigma^2 g[X_s' X_s]^{-1}\right)$ with $\vec{\theta}_s \subset (\beta_s, \gamma)$.

2.3 The Impact of Changing Samples – Or How much Variation can be Attributed to Africa?

A special feature of PWT data is that with each vintage the coverage of countries is likely to change. For the data employed by Ciccone and Jarociński (2010), the number of countries ranges from 88 (PWT 6.0) to 79 (PWT 6.2). Ciccone and Jarociński (2010, p.230) argue in favor of keeping sample sizes to their possible maximum, as it might be uncertain which countries may be included in future revisions. While this might have its virtues from the perspective of a forward-looking policy-maker, it has the potential to blur the conclusions on the robustness of the method (BMA).

A closer look at the PWT samples reveals that the countries dropped or added between revisions are mostly African.[1] In terms of growth, these were either very successful or unsuccessful countries with respect to their regional peers. As a consequence, and recently demonstrated in Masanjala and Papageorgiou (2008) and Crespo Cuaresma (2010), growth determinants in Africa systematically differ in various instances and it is difficult to consider this particular set of countries as a randomly chosen subsample of the data.

Varying Sample	$g = K^2$	$g = N$	hyper-g
Overall Max / Min Ratio	4.8580	2.0421	1.5224
PWT 6.0 vs. PWT 6.1	2.1179	1.4123	1.2114
PWT 6.0 vs. PWT 6.2	2.5888	1.6562	1.2778
PWT 6.1 vs. PWT 6.2	3.9466	1.6634	1.3518
Common Sample	$g = K^2*$	$g = N*$	hyper-$g*$
Overall Max / Min Ratio	2.4217	1.7344	1.4013
PWT 6.0 vs. PWT 6.1	1.9675	1.5866	1.3260
PWT 6.0 vs. PWT 6.2	1.8357	1.5434	1.3085
PWT 6.1 vs. PWT 6.2	1.5880	1.2470	1.1451

Table 2.1: Average PIP Max/Min ratios: for each revision pair, the figures above display the mean of the ratio maximum vs. minimum PIP per variable. The asterisk denotes the use of the common country set.

In order to identify the sources of PIP variability, we first replicate[2] the results of Ciccone and Jarociński (2010), employing the changing PWT samples and same prior set-up: a uniform prior[3] on the model space, and the 'benchmark' coefficient prior of Fernández et al. (2001a), that is

[1] Precisely, PWT 6.1 includes the countries of PWT 6.2 plus Botswana, Central African Republic, Mauritania, DR Congo (Zaire) and Papua New Guinea. PWT 6.0 includes the PWT 6.1 countries plus Liberia, Tunisia, West Germany and Haiti.

[2] All computations were carried out with the `R` package `BMS`. The data and detailed instructions for replication are available at `http://bms.zeugner.eu/datarev.php`.

[3] Note that we have omitted the model prior from the discussion since its impact seems to be rather limited as compared to the importance of the g-prior for this application (Ciccone and Jarociński, 2010, p.226).

$g = \max(K^2, N)$. As a measure of PIP instability we calculate the max/min ratio of the posterior inclusion probability over the three data vintages per variable and report its average (over covariates) as the *overall max/min ratio*. The results are summarized in the first column of Table 2.1.

The first column of the top panel ('varying sample') shows by far the greatest overall max/min ratio (and thus PIP variation), implying that the modeling strategy of Ciccone and Jarociński (2010) is indeed the one that is most severely plagued by instability. Furthermore, note that the amount of PIP variation reported in the first column is outstanding for all PWT vintages. That is, the strong variation is not driven by a single PWT vintage, but is a characteristic of the empirical framework employed by Ciccone and Jarociński (2010).

In order to quantify the role played by the changing sample, we replicate the estimations in Ciccone and Jarociński (2010) for the three data revisions but condition on the *same* set of countries. The first column, bottom panel ('common sample') illustrates that conditioning on the same observations per PWT vintage reduces PIP instability by a half. This suggests that the additional countries from PWT 6.0 and PWT 6.1 can be regarded as innovations (outliers) with the potential to change results.

2.4 The use of Default Priors – Or is fixing Prior Beliefs always advisable?

As alluded to before, the BMA framework used by Ciccone and Jarociński (2010) calls for eliciting the key hyperparameter Zellner's g. This parameter g reflects the strength of the researcher's prior guess on slope coefficients. Small values correspond to stronger beliefs that the regression slopes are zero (i.e. the prior is tightened).[1] By construction, the parameter g directly affects the posterior model probability (PMP) – the weight attributed to model M_s – and thus final inference:

$$p(M_s|y^j, X) \propto \left(1 - \tfrac{g}{1+g}\right)^{\frac{k_s}{2}} \left(1 - \tfrac{g}{1+g} R_s^2\right)^{-\frac{N-1}{2}} \tag{2}$$

The choice of g can be crucial for the results: Consider a relatively large g (implying a large shrinkage factor $\frac{g}{1+g}$). As is evident from equation (2), this will not only favor parsimonious models but also amplify any – potentially very small – differences in R_s^2. The resulting distribution of posterior model probabilities will therefore be highly concentrated on the few parsimonious models with the very highest R_s^2.

Figure 2.1 illustrates this effect by plotting cumulative PMPs based on different settings for g. The chart demonstrates that for PWT data, larger g attributes more weight to the first-best model relative to the remaining ones. If the data is dominated by noise, this *supermodel effect* (Feldkircher and Zeugner, 2009) will skew posterior mass to concentrate on a few 'supermodels'. Consequently it will skew the distribution of PIPs and thus amplify variations that may be due to noise. In contrast, employing a smaller g will limit such PIP variations, as exemplified by the second column of Table 2.1 (which uses $g = N$).

A remedy for the supermodel effect: Several 'default' mechanisms have been proposed to elicit g, the most prominent being the 'benchmark prior' by Fernández et al. (2001a), who recom-

[1] Note that in principle, it would be possible to elicit individual priors for the slopes, but we follow Ciccone and Jarociński (2010) and the bulk of the literature in centering all coefficient priors at zero.

mend $g = \max(N, K^2)$. Still, any of these fixed mechanisms risks to set g too small or too large with respect to the noise component in the data.

In response, Feldkircher and Zeugner (2009) propose to forgo fixed g-priors outright and to place a prior distribution on the parameter instead:[1] We follow the fairly general *hyper-g* prior approach put forward by Liang et al. (2008), which belongs to the family of mixtures of g-priors. With a (hyper-)prior on g, the prior on the coefficient vector can be interpreted as a mixture of normal distributions with fatter tails (Ley and Steel, 2011b). Technically, employing the hyper-g prior boils down to placing a Beta prior on the shrinkage factor $g/(1+g) \sim Beta(1, \frac{a}{2}-1)$. Choosing the hyperparameter a accordingly allows for formulating prior beliefs on g that match popular fixed-g settings.[2]

As opposed to fixing beliefs a priori, this hierarchical approach updates the prior beliefs according to the data. In that sense employing the hyper-g prior is less prone to misalignments of data and beliefs. Furthermore, the hyper-g prior is more flexible as it allows for *model-specific* g_s values (and shrinkage factors) that adjust to data quality:[3] If the data is characterized by minor noise, then posterior mass will concentrate on the true model(s) – even more than under fixed settings with large g. Conversely, if noise dominates the data, then posterior statistics will decrease g and PMPs under the hyper-g prior will be distributed more evenly. Even in such a case, BMA under a large fixed g would always come up with clear-cut results (a single model and a few covariates that obtain overwhelming support), not taking into account the small degree of data quality. The hyper-g framework, in contrast, will then point to inconclusiveness mirrored in evenly spread PMPs and PIPs.

The supermodel effect and PWT revisions: In their application, Ciccone and Jarociński (2010) follow the 'benchmark' recommendation by Fernández et al. (2001a) and set $g = \max(N, K^2)$ $= (66+1)^2$. Note that this renders the shrinkage factor very close to unity ($g/(1+g) \approx 0.9998$). We can thus expect PMPs to become highly concentrated, and small differences in R_s^2 to be translated into large differences in PMPs and PIPs. This large shrinkage factor treats the data as very informative with respect to the prior. In fact, however, the quality of the PWT data sets turns out to be rather poor. Using flexible g priors such as the hyper-g prior implies far smaller average shrinkage factors of around 0.95 (Table 2.2). This suggests that in order to avoid the supermodel effect, a fixed-g framework on PWT data should rather elicit $g \approx 19$. That said, any fixed g will always lack the flexibility of the hyper-g prior in adjusting to data quality.

In order to examine the role of the hyper-g prior in the context of growth determinants and PWT revisions, we provide the full results in Table 2.3. The hyper-g prior results greatly reduce the variability of PIPs over revisions for the 8 covariates mentioned before. In particular, the investment price and population density variables as well as the tropical area dummy do not appear to matter for growth under either revision while the African dummy seems robust. As a further observation one might stress that under the hyper-g prior, the PIPs are much larger on average – this results from the data inducing lower posterior shrinkage factors and therefore emphasizing less parsimonious models than under the 'benchmark' priors. The smaller model size penalty results

[1] Note that there are also other flexible prior frameworks, such as the Empirical Bayes (EB) approach. For the sake of brevity, we have omitted the EB results in this note, since they are very similar to those under the hyper-g prior (Feldkircher and Zeugner, 2009, Appendix).

[2] For our purpose, we defined the parameter a such that prior expected shrinkage matches the $g/(1+g)$ used in Ciccone and Jarociński (2010). Experimenting with other parameters a produces results that are very close to the ones reported here.

[3] In this note we refer to data quality by the degree of variation of the dependent variable explained by the data at hand.

Varying Sample	$g = K^2$	$g = N$	hyper-g
PWT 6.0	0.9998	0.9888	0.9103
PWT 6.1	0.9998	0.9882	0.9293
PWT 6.2	0.9998	0.9875	0.9454
Common Sample	$g = K^2*$	$g = N*$	hyper-$g*$
PWT 6.0	0.9998	0.9875	0.9361
PWT 6.1	0.9998	0.9875	0.9476
PWT 6.2	0.9998	0.9875	0.9454

Table 2.2: Average shrinkage factors for three PWT revisions under the benchmark case $g = 67^2$, $g = 79$ and the hyper-g prior (with prior expected shrinkage factor $\frac{g}{1+g} = 67^2/(1+67^2)$). The asterisk denotes the use of the common country set. All results are based on a Metropolis-Hastings model sampler as in Fernández et al. (2001a), with 80 million iterations after 20 million burn-ins.

into much larger posterior model size (≈ 25 vs. $7-9$ under the Benchmark case). As a consequence, comparing *absolute* PIPs across different priors is prone to misleading conclusions since the sum of PIPs is by construction equal to the respective posterior model size – which in turn strongly depends on the value of g.[1]

Consider Figure 2.1 to see how the hyper-g prior induces data-dependent shrinkage. The top panel plots the cumulative PMP based on the PWT 6.1 vintage (varying sample composition). As expected, the benchmark prior ($g = K^2$) results in the by far most concentrated PMP distribution, whereas the hyper-g prior spreads PMPs most evenly. The Figure also shows that the shrinkage factor induced by the fixed $g = N$ setting is still too high to match the PMP distribution of the hyper-g prior. The bottom panel of Figure 2.1 provides the same plot for the sample of common countries, which was already shown to induce less PIP variation. Comparing the two figures reveals that the PMP distribution under neither fixed g prior responds substantially to the change in samples. The only line that adjusts considerably is that of the hyper-g prior. This can be best seen by considering the differences in PMP distributions of the hyper-g prior and the fixed UIP ($g = N$) prior: while this difference is pronounced under the varying sample framework, it is considerably smaller for the common sample data. The smaller noise component inherent in the common sample induces the hyper-g prior to put more weight to the data relative to the prior and thus to skew PMP mass. The fixed prior settings, in contrast, do not allow for adjusting the PMP distribution to a change in the noise component.

To assess the degree of PIP variation for the different prior set-ups, we consider the overall max/min figures provided in Table 2.1. For illustration, compare first the results for the original framework with varying sample size (Table 2.1, top panel). A decrease in g (such as $g = N$) already lowers PIP variation substantially, which demonstrates the supermodel effect. However, the most remarkable reduction is achieved by the hyper-g prior, with a drop in variation of close to 70% compared to the benchmark framework (Table 2.1, top panel, columns 1 and 3). The results for the common samples (Table 2.1, bottom panel) yield a similar picture: The hyper-g prior greatly reduces PIP instability in comparison to the other fixed prior settings. With respect to the prior set-up used in Ciccone and Jarociński (2010) (Table 2.1, bottom panel, column 1), PIP variation is reduced by close to 50% – from already decreased levels due to conditioning on the same countries. Finally, Table 2.1 reveals that the minimal overall max/min ratio is achieved by employing the hyper-g prior coupled with using identical countries over PWT vintages. That said, note that PIP variation is always smaller with the hyper-g framework then under the popular fixed g settings. Thus, even when –

[1] For instance, when $g \to \infty$, all PIPs will tend to zero, thus their absolute differences will vanish.

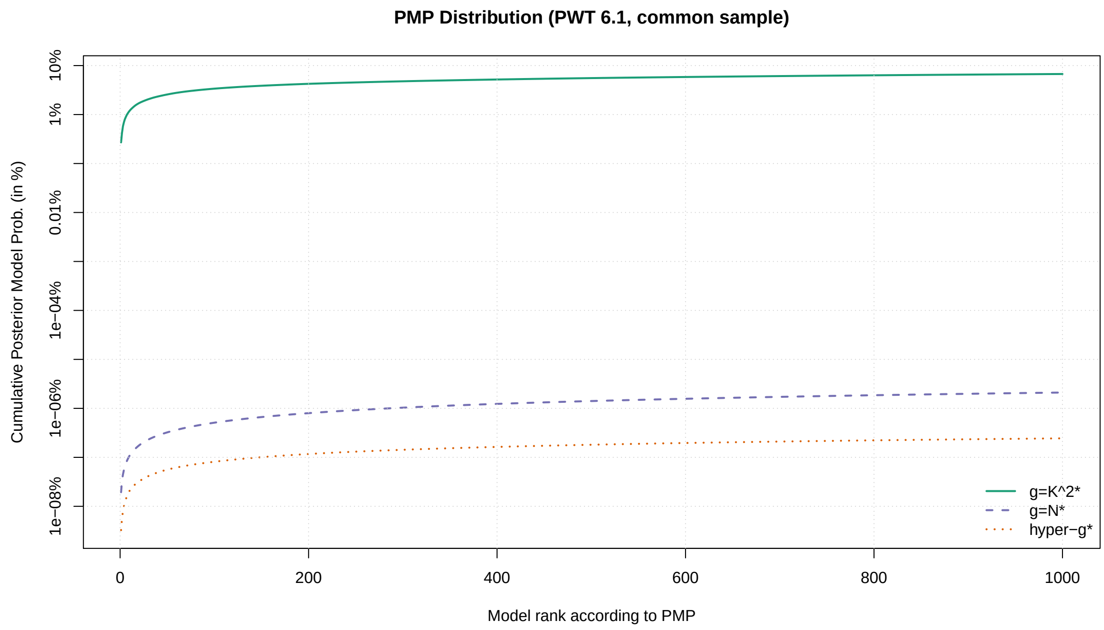
for data availability reasons – sample composition changes with revisions, we strongly recommend the use of a hyper-g prior.

There is one important caveat, though: the quite low shrinkage factor induced by hyper-g not only decreases PIP variation over data revisions, but also over covariates for a given revision. Consequently, this implies that there are less covariates that could be identified as 'considerably more important' than others. As has been noted above, this is a direct result from posterior mass being spread out more evenly over models due to an important noise component in the data. Yet this trait may be desirable, as noisy data should not lead to the striking conclusions from the 'agnostic' approach criticized by Ciccone and Jarociński (2010).

2.5 Conclusion

This note addresses an important issue raised by Ciccone and Jarociński (2010): inference of (agnostic) BMA applied to growth empirics is not robust under small perturbations of international income data. Our response demonstrates that such instability is partly due to the overconfident 'default' g-prior framework employed by the authors. Instead, we propose to rely on the hyper-g prior. While 'agnostic' in the sense of Ciccone and Jarociński (2010), it adjusts to data quality and induces smaller shrinkage factors according to the data's considerable noise component. This in turn renders BMA results considerably more stable over different revisions of PWT growth data. However, reflecting poor data quality, results under hyper-g discriminate far less among covariates. The empirical findings under a flexible prior may thus be characterized as 'robust ambiguity', limiting statements about the importance of growth determinants to a quite small subset of covariates.

2.A Charts and Tables



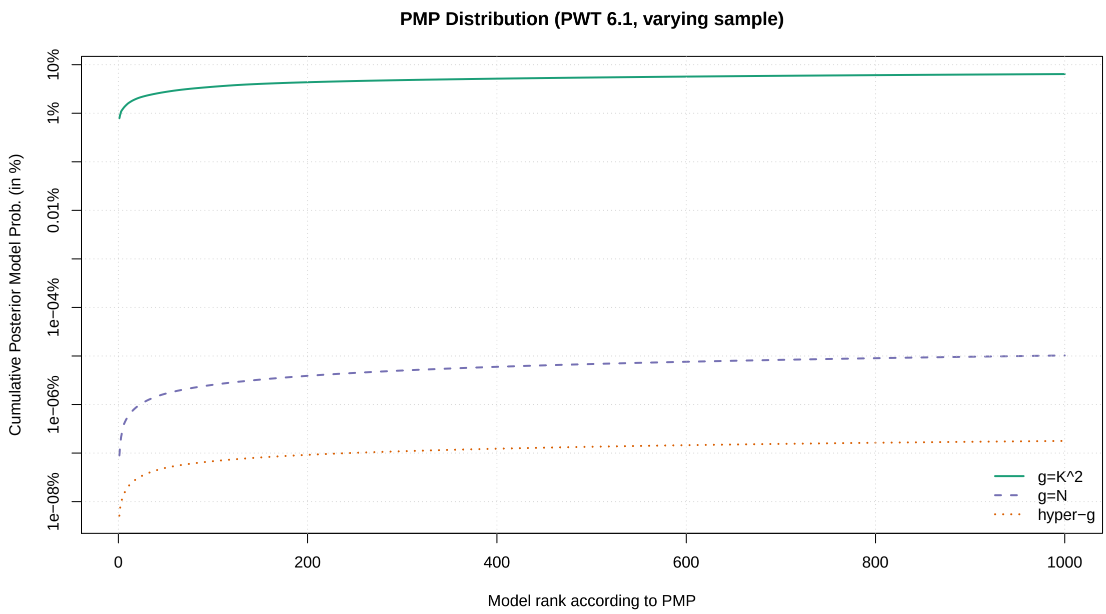


Figure 2.1: Cumulative posterior model probabilities for the best 1000 models (by PMP) under different settings for g. The lower panel displays results for the original PWT 6.1 data with 88 countries. The upper panel shows results for the PWT 6.1 sample with observations restricted to the 'common sample' countries that appear in later PWT revisions (79 observations).

Table 2.3: Posterior Inclusion Probabilities. Left panel corresponds to data set with varying number of countries, center panel to data set with common countries over the three revisions. Right panel displays results of hyper-g BMA over common countries. The asterisk denotes the use of the common country set. Results are based on 80 million posterior draws after a burn-in phase of 20 million draws.

	$g = K^2$			$g = K^2$*			hyper-g*		
PWT revisions	6.0	6.1	6.2	6.0	6.1	6.2	6.0	6.1	6.2
GDP in 1960 (log)	0.73	1.00	1.00	0.81	1.00	1.00	0.87	1.00	1.00
Absolute Latitude	0.04	0.03	0.03	0.06	0.02	0.03	0.38	0.26	0.29
Air Distance to Big Cities	0.05	0.43	0.05	0.07	0.09	0.05	0.34	0.29	0.30
Ethnolinguistic Fractionalization	0.12	0.03	0.04	0.02	0.05	0.04	0.24	0.34	0.28
British Colony Dummy	0.03	0.03	0.02	0.02	0.02	0.02	0.25	0.26	0.23
Fraction Buddhist	0.13	0.13	0.30	0.12	0.17	0.30	0.39	0.59	0.48
Fraction Catholic	0.04	0.02	0.07	0.13	0.06	0.07	0.58	0.39	0.45
Civil Liberties	0.03	0.02	0.02	0.07	0.04	0.02	0.40	0.30	0.24
Colony Dummy	0.04	0.09	0.02	0.12	0.05	0.02	0.36	0.25	0.24
Fraction Confucius	**0.24**	**0.17**	**0.84**	**0.55**	**0.61**	**0.84**	**0.77**	**0.93**	**0.96**
Population Density 1960	**0.11**	**0.71**	**0.02**	**0.02**	**0.04**	**0.02**	**0.29**	**0.25**	**0.24**
Population Density Coastal in 1960s	**0.44**	**0.76**	**0.11**	**0.42**	**0.37**	**0.11**	**0.34**	**0.31**	**0.38**
Interior Density	0.02	0.02	0.03	0.03	0.06	0.03	0.28	0.37	0.26
Population Growth Rate 1960-90	0.02	0.03	0.03	0.03	0.03	0.03	0.30	0.23	0.28
East Asian Dummy	**0.79**	**0.75**	**0.33**	**0.54**	**0.46**	**0.33**	**0.55**	**0.37**	**0.32**
Capitalism	0.02	0.02	0.02	0.02	0.03	0.02	0.24	0.29	0.28
English Speaking Population	0.02	0.02	0.02	0.02	0.02	0.02	0.26	0.28	0.25
European Dummy	0.03	0.04	0.07	0.07	0.08	0.07	0.46	0.50	0.36
Fertility in 1960s	**0.04**	**0.15**	**0.90**	**0.23**	**0.65**	**0.90**	**0.65**	**0.73**	**0.80**
Defense Spending Share	0.02	0.02	0.04	0.03	0.02	0.04	0.31	0.30	0.30
Public Education Spending Share in GDP in 1960s	0.02	0.02	0.02	0.02	0.03	0.02	0.25	0.26	0.25
Public Investment Share	0.07	0.06	0.02	0.03	0.02	0.02	0.33	0.24	0.28
Nominal Government GDP Share 1960s	0.05	0.02	0.29	0.28	0.11	0.29	0.84	0.46	0.74
Government Share of GDP in 1960s	0.08	0.04	0.05	0.09	0.13	0.05	0.31	0.53	0.34
Gov. Consumption Share 1960s	0.12	0.06	0.03	0.05	0.06	0.03	0.29	0.37	0.29
Higher Education 1960	0.07	0.02	0.03	0.14	0.03	0.03	0.60	0.25	0.27
Religion Measure	0.02	0.03	0.02	0.02	0.02	0.02	0.30	0.27	0.24
Fraction Hindus	0.05	0.02	0.03	0.03	0.03	0.03	0.37	0.40	0.33
Investment Price	**0.81**	**0.98**	**0.02**	**0.03**	**0.07**	**0.02**	**0.26**	**0.28**	**0.22**
Latin American Dummy	0.18	0.08	0.34	0.34	0.28	0.34	0.39	0.33	0.34
Land Area	0.02	0.02	0.02	0.04	0.05	0.02	0.49	0.32	0.29
Landlocked Country Dummy	0.02	0.09	0.03	0.03	0.05	0.03	0.31	0.32	0.24
Hydrocarbon Deposits in 1993	0.03	0.13	0.11	0.03	0.22	0.11	0.35	0.73	0.65
Life Expectancy in 1960	0.23	0.26	0.03	0.03	0.03	0.03	0.25	0.23	0.28
Fraction of Land Area Near Navigable Water	0.02	0.05	0.02	0.03	0.04	0.02	0.28	0.26	0.24
Malaria Prevalence in 1960s	0.23	0.03	0.03	0.05	0.03	0.03	0.30	0.26	0.30
Fraction GDP in Mining	0.16	0.28	0.02	0.04	0.02	0.02	0.31	0.22	0.23
Fraction Muslim	0.13	0.22	0.43	0.06	0.49	0.42	0.50	0.88	0.84
Timing of Independence	0.02	0.09	0.14	0.03	0.09	0.14	0.30	0.70	0.85
Oil Producing Country Dummy	0.02	0.02	0.02	0.02	0.02	0.02	0.23	0.22	0.23
Openness measure 1965-74	0.08	0.07	0.16	0.13	0.31	0.16	0.66	0.68	0.62
Fraction Orthodox	0.02	0.02	0.03	0.02	0.02	0.03	0.28	0.35	0.37
Fraction Speaking Foreign Language	0.10	0.04	0.07	0.04	0.03	0.07	0.28	0.26	0.34
Primary Schooling in 1960	0.81	0.99	1.00	0.98	1.00	1.00	0.99	1.00	1.00
Average Inflation 1960-90	0.02	0.02	0.06	0.02	0.03	0.06	0.24	0.25	0.24
Square of Inflation 1960-90	0.02	0.02	0.04	0.02	0.02	0.04	0.24	0.23	0.23
Political Rights	0.07	0.24	0.02	0.06	0.04	0.02	0.43	0.44	0.42
Fraction Population Less than 15	0.04	0.03	0.04	0.03	0.04	0.03	0.42	0.33	0.32
Population in 1960	0.03	0.02	0.02	0.02	0.02	0.02	0.34	0.22	0.23
Fraction Population Over 65	0.03	0.06	0.05	0.05	0.07	0.05	0.48	0.41	0.39
Primary Exports 1970	0.06	0.23	0.28	0.13	0.48	0.28	0.44	0.84	0.76
Fraction Protestants	0.05	0.02	0.07	0.14	0.06	0.07	0.53	0.37	0.44
Real Exchange Rate Distortions	0.10	0.05	0.02	0.11	0.03	0.02	0.60	0.23	0.21
Revolutions and Coups	0.04	0.03	0.03	0.03	0.02	0.03	0.27	0.32	0.34
African Dummy	**0.19**	**0.21**	**0.85**	**0.86**	**0.83**	**0.85**	**0.83**	**0.72**	**0.78**
Outward Orientation	0.04	0.04	0.02	0.02	0.02	0.02	0.26	0.22	0.25
Size of Economy	0.02	0.03	0.04	0.02	0.07	0.04	0.31	0.28	0.28
Socialist Dummy	0.02	0.03	0.03	0.03	0.06	0.03	0.24	0.28	0.22
Spanish Colony	0.12	0.02	0.07	0.28	0.04	0.07	0.41	0.23	0.31
Terms of Trade Growth in 1960s	0.02	0.02	0.02	0.02	0.02	0.02	0.23	0.30	0.28
Terms of Trade Ranking	0.02	0.02	0.02	0.02	0.02	0.02	0.28	0.24	0.22
Fraction of Tropical Area	**0.56**	**0.67**	**0.05**	**0.32**	**0.23**	**0.05**	**0.49**	**0.30**	**0.25**
Fraction Population In Tropics	0.06	0.16	0.04	0.02	0.02	0.04	0.26	0.23	0.28
Fraction Spent in War 1960-90	0.02	0.02	0.02	0.02	0.02	0.02	0.25	0.22	0.26
War Participation 1960-90	0.02	0.02	0.03	0.02	0.02	0.03	0.25	0.22	0.26
Years Open 1950-94	0.12	0.06	0.09	0.12	0.08	0.09	0.31	0.24	0.25
Tropical Climate Zone	0.02	0.03	0.04	0.02	0.04	0.04	0.23	0.36	0.44
Number of Observations	88	84	79	79	79	79	79	79	79

3

Leverage as a Predictor for Real Activity and Volatility

Co-authored with:
Robert Kollmann, *ECARES, Université libre de Bruxelles, Université Paris-Est and CEPR*

Abstract
This chapter explores the link between the leverage of the US financial sector, of households and of non-financial businesses, and real activity. We document that leverage is negatively correlated with the future growth of real activity, and positively linked to the *conditional volatility* of future real activity and of equity returns. The joint information in sectoral leverage series is more relevant for predicting future real activity than the information contained in any individual leverage series. Using in-sample regressions and out-of sample forecasts, we show that the predictive power of leverage is roughly comparable to that of macro and financial predictors commonly used by forecasters. Leverage information would *not* have allowed to predict the 'Great Recession' of 2008-2009 any better than conventional macro/financial predictors.

Keywords: leverage, financial crisis, forecasts, real activity, volatility

JEL Classification: E32, E37, C53, G20

3.1 Introduction

In the years before the recent (2007-09) financial crisis, the leverage of many major financial institutions increased steadily, and reached unprecedented levels. The crisis revealed the fragility of the financial sector, and of many highly indebted non-financial firms and households, and it has triggered the sharpest global recession since the 1930s. Before the crisis, structural macro models largely abstracted from financial intermediaries, and macro forecasting models ignored balance sheet information. The recent dramatic events require a rethinking of the role of finance for real activity. In particular, understanding the link between balance sheet conditions and the real economy has become a key priority.

To explore that link, this paper analyzes the predictive power of leverage for GDP, industrial production, unemployment and physical investment (as well as for equity returns). Leverage is defined as the ratio of an agent/sector's assets to her net worth (assets minus debt). We use quarterly US data (1980-2010), and consider leverage information from the Flow of Funds, for three broad financial sectors (insurance companies, securities brokers-dealers, and commercial banks), as well as for households and for non-financial corporate businesses. We complement that information using the ratio of assets (at book-values) to the *market value* of equity, for financial corporations included in three Dow Jones stock price indices: 'US-Insurance', 'US-Banks' and 'US-Financial Services'. We estimate forecast equations for real activity and equity returns that use these 8 sectoral leverage ratios, and principal components of a set of 30 other macro-financial variables, as predictors. Predictive performance is evaluated using both in-sample fit and (rolling) out-of-sample forecast accuracy. As a robustness check, we apply Bayesian Model Averaging to evaluate the impact of leverage conditional on any potential combination of controls (both in- and out-of-sample).

Our results show that *each* of our 8 leverage variables is negatively related to *future* real activity. This result is not driven by the recent financial crisis. The predictive power of leverage is roughly comparable to that of standard macro-financial forecast variables. Among the 8 leverage series, insurance sector leverage (from Flow of Funds), and the equity-market-value-based leverage measure for banks have the highest out-of-sample predictive ability for GDP. For forecasting real activity, it is advisable to combine the sectoral leverage information, using cross-sectional medians or principal components, instead of using the sectoral leverages series individually as predictors. Thus, the joint information in the sectoral leverage series is more relevant than the information contained in any individual sectoral leverage series. However, despite the high statistical significance of leverage (and of macro-financial factors) in the forecasting regressions, none of the variables considered here would have helped in predicting the 'Great Recession' of 2008-2009.

We also document that increasing leverage at a given date is associated with greater uncertainty about future economic conditions. In particular, leverage is strongly positively related to the absolute value of forecast errors for future real activity (generated by our forecast equations) and to the CBOE equity market volatility index VIX (a measure of expected *future* stock price volatility, derived from option prices). Furthermore, leverage is positively related to the cross-sectional dispersion (across forecasters) of predicted future real activity reported by the Philadelphia Fed Survey of Professional Forecasters (SPF). The link between leverage and conditional future volatility seems consistent with recent theoretical models in which higher leverage amplifies the effect of unanticipated macroeconomic and financial shocks on real activity and asset prices – the idea is that higher leverage makes the economy more fragile.[1]

[1] See, e.g., Krugman (2009), Devereux and Yetman (2010) and Kollmann and Malherbe (2011) for discussions of these mechanisms (and for detailed references).

The work here contributes to key recent strands in the *macro-modeling* and *macro-policy* literatures. Since the crisis, much effort has been devoted to the development of dynamic general equilibrium models with financial intermediaries; e.g., Int Veld et al. (2011) and Kollmann et al. (2011);[1] in those models, leverage is a key state variable for real activity. Our goal here is to identify robust empirical regularities about the link between leverage and real activity that can be used to evaluate those models. In the policy arena, the development of a macroprudential supervision framework (to be implemented by new agencies, such as the European Systemic Risk Board and the US Office for Financial Research) has risen to top priority since the crisis. The monitoring of leverage ratios, to issue early warning indicators of crises, is likely to be a key dimension of the new framework (see Galati and Moessner, 2010). However, our results suggest that the use of *aggregate* leverage information is unlikely to be a panacea for predicting crises.

Our results on the predictive content of leverage for real activity complement a recent study by Adrian and Shin (2010) who argue, based on in-sample fit, that brokers-dealers (and shadow-banking) balance sheets explain future GDP.[2] We conduct a more systematic empirical exploration of the forecasting performance of leverage than these authors, by considering balance sheets for a larger number of sectors, using a broader set of controls, and evaluating both in-sample fit and *out-of-sample* forecast accuracy. Our approach thus seems better suited for evaluating which variables are robustly correlated with real activity. We document (inter alia) that the predictive ability of brokers-dealers is highly sample dependent, and that the *joint* information contained in sectoral leverage series is more relevant for future real activity than the information contained in any individual series.[3]

Section 3.2 describes the leverage data, and Section 3.3 discusses our econometric methodology. Sections 3.4 and 3.5 present the results, and Section 3.6 concludes.

3.2 Leverage data

We construct quarterly time series on the leverage ratios of five major sectors covered by the US Flow of Funds (FoF); specifically, we consider three financial sectors– commercial banks (CB), property and life insurance companies (INS), securities brokers and dealers (SBD)–as well as households (HH) and non-financial corporate businesses (BUS). For each of these sectors, the leverage ratio is defined as: total assets/(total assets − financial liabilities). Asset and liabilities reported in the FoF are partly measured at book values, and may thus differ from market values.[4] We thus complement the FoF leverage measures using the ratios of (book-value) assets to the market value of equity, for financial companies included in three Dow Jones stock price indices (as reported by Datastream): 'US-Banks', 'US-Insurance' and 'US-Financial Services';[5] we refer to these sectors as BNK-MV, INS-MV and FIN-MV, respectively (where 'MV' stands for market value); the corre-

[1] Other contributions include Aikman and Paustian (2006), Van den Heuvel (2008), De Walque et al. (2010), Angeloni and Faia (2009), von Peter (2009), Cúrdia and Woodford (2009), Antipa et al. (2010), Dib (2010), Gerali et al. (2010), Gertler and Kiyotaki (2010), and Meh and Moran (2010).

[2] Adrian et al. (2010) also argue that brokers-dealers leverage predicts equity and bond returns.

[3] Also, as mentioned above, we show that balance sheet information would have failed to predict the crisis, and document that leverage is strongly related to the conditional variability of real activity.

[4] Deviations from market values are likely to be smallest when the balance sheets in a given sector are marked to market and when assets and liabilities are short term.

[5] Datastream provides the aggregate market valuation of the firms included in each of these indices, as well as the corresponding (book-value) assets. The 'US-Banks' index includes commercial banks; 'US-Financial Services' includes investment banks, credit card issuers, and institutions specializing in consumer loans, and thus overlaps only partially with the FoF 'securities brokers-dealers' (SBD) category. These indices only include the major financial institutions, while Flow of Funds data cover all firms in a given sector.

sponding Datastream series are available from 1980q4. (See the Appendix for detailed information on the data.) We thus use data from 1980q4 in the subsequent analysis; our sample ends in 2010q3.

The forecast equations for real activity discussed below are estimated on rolling windows of 40 quarters; given the lag structure of the forecast regressions, the resulting (out-of-sample) forecast evaluation period is 1993q3-2010q3. Figure 3.1 plots the eight sectoral leverage ratios over that period. The mean FoF-based leverage ratios of households (1.2) and of non-financial corporations (2.0) in 1993q3-2010q3 are much lower than those of the financial sectors (CB: 8.9; INS: 7.7; SBD: 27.3). The sample averages of the financial sector leverage measures based on the market value of equity are lower than the FoF-based finance sector leverages (BNK-MV: 5.9; INS-MV: 4.0; FIN-MV: 2.6).[1]

Note also that securities brokers-dealers (SBD) leverage (from FoF) and the financial sector leverage measures based on the market value of equity undergo much bigger fluctuations than the other leverage series. SBD leverage grew very strongly until the crisis, reaching a peak of 55 in 2008q3, and then (after the Lehman bankruptcy) collapsed to about 20. BNK-MV, INS-MV and FIN-MV leverage likewise grew strongly, and peaked in 2009q2 (i.e. at the point in time when bank equity prices reached their lowest values, during the recent crisis), before falling noticeably.[2] By contrast, FoF-based commercial-bank leverage has had a flat trend since about 2005, and held up well during the crisis. This may partly reflect accounting discretion, which has allowed banks to overstate the value of their assets in the crisis (e.g. Huizinga and Laeven, 2009).

Leverage also exhibits interesting correlations with GDP. The year-on-year (YoY) growth rate of securities brokers-dealers leverage is positively correlated with YoY GDP growth (correlation: 0.43), i.e. brokers-dealers leverage is pro-cylical (1993q3-2010q3). CB and INS leverage is a-cylical (correlations with GDP close to zero, and statistically insignificant), while the remaining leverage variables are strongly counter-cylical (median leverage-GDP correlation: -0.50). However, the YoY growth of *all* eight leverage series is *negatively* correlated with *future* YoY GDP growth, at leads greater than 2 quarters. We show below that a significant negative link between leverage and *future* real activity can also be detected, when controlling for other macro/financial variables.

3.3 Econometric methodology

We focus on one-year-ahead forecasts for real activity and equity returns. Following Stock and Watson (2002), we fit forecasting equations of the following form (by OLS):

$$Y_{t+4} - Y_t = \beta_0 + \beta_1 (Y_t - Y_{t-1}) + \beta_2 \Phi_t + \beta_3 \Lambda_t + \epsilon_{t+4} \tag{1}$$

where Y_{t+4} is a measure of real activity in period $t+4$ (to be predicted given period t information). One period represents one quarter in calendar time. Λ_t is the change of an individual sector's log leverage between $t-4$ and t, or the median or first principal component of the (standardized) YoY changes of the eight sectoral log leverage series.[3] Φ_t is a vector of controls, discussed below. Note

[1]This partly reflects the fact that the market value of equity is generally greater than its book value. Leverage measures based on book-value equity (also available from Datastream) are much closer to FoF-based leverage measures: 13.8, 7.5 and 14.0, respectively, for 'US-Banks', 'US-Insurance' and 'US-Financial Services' (1993q3-2010q3).

[2]These movements of the BNK-MV, INS-MV, FIN-MV and SBD leverage measures are largely driven by the sizable fluctuations in these sectors' equity. BNK-MV, INS-MV, FIN-MV leverage are also highly negatively correlated with the overall stock market (the correlation of year-on-year growth of these three leverage measures and the annual Fama-French stock market return is about -0.7).

[3]We also estimated equations with forecast horizons of 1, 2, 6, 8, and 12 quarters, and found that the key results

that, in equation (1), the quarterly first difference of real activity ($Y_t - Y_{t-1}$) is also included as a regressor.[1]

We focus on the following measures of real activity: GDP, industrial production (IP), the unemployment rate (UE) and physical investment (I).[2] The future YoY changes of GDP, IP and I are expressed as annual log growth rates (in %). The forecast equations for unemployment use as a dependent variable the YoY change of the % unemployment rate. We also run the forecasting regression (1) for the % YoY excess equity return (Rx), defined as the difference between the stock market return and the T-bill return (see Appendix).

Due to the upward trends in several of the leverage series (see above), we use the YoY change in (log) leverage as a predictor, in equation (1).[3] (We also estimated forecasting regressions that use the deviation of leverage from a moving average of lagged leverage as a predictor, or the deviation from a linear trend fitted to lagged leverage. The results are very similar to those discussed below.)

Note that log leverage equals the difference between log assets and log equity. We thus also considered forecast equations in which (YoY changes of) log assets and log equity are entered separately as predictors. These specifications yield lower out-of-sample forecast accuracy than models in which log leverage is used as a predictor. We tested whether the coefficient of log equity equals the negative of the coefficient of log assets; for Flow of Funds data, we fail to reject that hypothesis – this suggests that the effect of equity and of assets on future real activity can be subsumed by leverage, consistent with regression equation (1). Hence, the subsequent analysis focuses on leverage as a predictor.

As controls (Φ_t) we use the first four principal components (factors) extracted from a set of macro-financial variables other than leverage, following Stock and Watson (2002).[4] We consider a set of 30 predictors that are widely used in macroeconomic and financial forecasting: quarterly growth rates of NIPA aggregates and price indices, quarterly asset returns etc. (all variables are properly stationarized). See list in Appendix.[5]

We compute out-of-sample measures of forecast accuracy based on a rolling 40-quarter estimation window. As our data set covers the period 1980q4-2010q3, the forecasts based on the rolling window pertain to 1993q3-2010q3 (taking into account the lags in (1)), as mentioned above. We also report the in-sample fit of model (1), based on a regression (non-rolling) for 1993q3-2010q3 (for each dependent variable).

Tables 1-3 report empirical results for different variants of regression (1). Specifically, the model variant referred to as 'Random Walk' only includes the intercept as a regressor, i.e. β_1, β_2 and β_3 are set at $\beta_1 = \beta_2 = \beta_3 = 0$. The 'Just ΔY' model variant also includes the first-difference of the

---

discussed below continue to hold for those horizons. The results are likewise robust to including leverage growth over more than four quarters as a predictor. See discussions below.

[1] Also considered were models in which up to eight lags of $Y_t - Y_{t-1}$ and of Φ_t are included as regressors. The Bayesian Information Criterion (BIC) does not favor inclusion of these lags.

[2] It seems interesting to run the forecasting equation for investment, as investment might be especially sensitive to balance sheet conditions of financial intermediaries and of non-financial firms. Investment, IP, and UE growth rates are strongly correlated with GDP growth rates, but more volatile.

[3] Unit root tests indicate that the 8 individual log leverage series are integrated of order one.

[4] We also considered specifications in which between 1 and 8 factors are used as controls; see discussion below (and Web Appendix). The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) favor the use of 4 factors. Models with 4 factors also have the best out-of-sample forecast performance; that performance decreases noticeably when more factors are included (due to overfitting). We thus use 4 macro-financial control factors in what follows. The fact that a small number of factors has best predictive power for macro variables has been widely documented; see, e.g., Giannone et al. (2005), and Stock and Watson (2008).

[5] The first four principal components of these 30 variables account for 57% of the total variance.

predicted variable ($Y_t - Y_{t-1}$) as a regressor. (All other model variants also include the intercept and the first-difference of the dependent variable as regressors.) The forecast model labeled 'F' adds the four macro-financial factors. 'F, PC-LEV' adds the first principal component of the YoY change of the eight sectoral (standardized) log leverage series to the 'F' model. The 'F, MED-LEV' model variant uses the cross-sectional median of standardized YoY changes of the eight log leverage series as the leverage predictor instead.

Forecast model variants

Model	Restrictions
Random Walk	$\beta_1 = \beta_2 = \beta_3 = 0$
Just ΔY	$\beta_2 = \beta_3 = 0$
F	$\beta_3 = 0$
F, PC-LEV	Λ_t =first principal component of standardized YoY changes in 8 sectoral log leverages
F, MED-LEV	Λ_t =median of standardized YoY changes in 8 sectoral log leverage variables
F, MED-FoF	Λ_t =median of standardized YoY changes in 5 sectoral log leverages from Flow of Funds
F, MED-MV	Λ_t =median of standardized YoY changes in 3 sectoral log leverages based on equity market values
PC-LEV	$\beta_2 = 0$, Λ_t =principal component of YoY changes in 8 sectoral log leverages
MED-LEV	$\beta_2 = 0$, Λ_t =median of standardized YoY changes in 8 sectoral log leverage variables
CB, INS, SBD, HH, BUS, INS-MV, BNK-MV, FIN-MV	$\beta_2 = 0$, Λ_t is the YoY change of one of the eight sectoral log leverage variables

The leverage predictor in the model 'F, MED-FoF' is constructed as the median of YoY changes of log book-value-based leverage measures (from Flow of Funds data). In contrast, the model 'F, MED-MV' uses the median of YoY changes for the three market-value based leverage measures (based on Dow Jones/Datastream data). The entries labeled 'CB', 'INS', etc. pertain to forecast models that use the YoY difference of the corresponding individual sectoral leverage variable as regressors. The Table above summarizes these different model variants.

3.3.1 Benchmarking leverage to controls: Bayesian Model Averaging

Comparing models with leverage ratios against a factor model of macro-financial data ('F') provides a tough benchmark for evaluating the predictive content of leverage growth. However, a comparison with such a factor model risks attributing undue prominence to leverage variables, because the factor extraction method might underweight individual macroeconomic predictors that could outperform the leverage variables.

To address this concern, we apply Bayesian Model Averaging (BMA): Potentially, the 'true' forecasting model as in (1) could be based on any linear combination of leverage variables and individual macro-financial controls. BMA assigns each model a prior probability, which is updated to a data-based *posterior model probability* (PMP) via Bayes' theorem. These PMPs are used to construct a weighted average of model-specific coefficients and statistics, thus effectively accounting for model uncertainty by integrating over the model space (see Appendix).

BMA represents the importance of a variable with its Posterior Inclusion Probability (PIP). Defined as the sum of PMPs from all models that include a particular regressor, its PIP indicates the empirical probability that this regressor is included in the true data-generating model. Note, however, that the sum of PIPs is equal to the posterior weighted average of the number of included variables ('posterior model size'), which can in turn be sensitive to model priors. We therefore focus on the ranking and relative importance of PIPs rather than their absolute values.[1]

Our BMA implementation closely follows the literature standard (e.g. Stock and Watson, 2006). Forecasting equations are estimated by Bayesian regression, and model priors incorporate a model size penalty (see Appendix).[2] Akin to the OLS forecasts, we apply BMA in-sample for the period 1993q3-2010q3, as well as for out-of-sample forecasts based on rolling estimation windows of 40 observations (similar to Faust et al., 2011).[3]

To assess the importance of leverage vs. the individual macro-financial controls, we compare the forecasts from three different BMA settings: 'BMA NO-LEV' applies BMA to just the 30 individual macro-financial controls (omitting the leverage variables). 'BMA MED-LEV' adds the median of eight standardized YoY leverage growth rates as a composite indicator for leverage.[4] Note that this setting attributes the same prior importance to median leverage as to each single macro-financial control. 'BMA 8-LEV' evaluates all combinations of the 30 controls and the eight individual leverage variables.[5]

3.4 Results: leverage as a predictor for real activity and equity returns

Tables 3.1–3.3 show results for forecasting equation (1). Row 1 of Table 3.1 reports root mean squared forecast errors (RMSEs) for the 'Just ΔY' model variant. Henceforth, we take this model variant as a benchmark – in Table 1, we normalize the RMSEs for the other model variants by the RMSE of the 'Just ΔY' variant (see rows 2-18). The Table also presents the relative RMSE of the median forecasts (for GDP, IP, UE and I) reported by the Philadelphia Fed Survey of Professional Forecasters (SPF). The left panel of Table 3.1 reports in-sample RMSEs, while the right panel reports RMSEs of out-of-sample forecasts, based on the rolling 40-quarter estimation windows. Throughout, the forecast evaluation period is 1993q3-2010q3.

[1] While PIP is often interpreted in terms of covariate 'significance', its relationship to the p-values of OLS coefficient t-tests is not straightforward (in the multi-model case).

[2] We use the 'Bayesian regression' framework for estimating individual models. For the model coefficients we specify a Normal-conjugate prior with Zellner's g prior, i.e. our prior expects model slopes to be zero, with a coefficient covariance similar to that of OLS times a shrinkage parameter. The shrinkage parameter is endogenously updated (cf. Appendix).

[3] Note that out-of-sample BMA would imply to evaluate more than a billion models for each of the 69 rolling samples. As this is not computationally feasible, we resort to a well-established MCMC sampler with 2 400 000 iterations to explore the model space (cf. Appendix). For each dependent variable and out-of-sample BMA setting, we thus evaluate 165.6 million regressions.

[4] A BMA version corresponding to the OLS forecast model 'PC-LEV' was omitted for the sake of brevity.

[5] We use the popular binomial model prior as in Sala-i-Martin et al. (2004), imposing a size penalty tuned such that it conforms to a chosen prior expected model size. For comparison with the OLS forecasting models, we set the prior expected model size for all BMA variants to 5 (the number of variables in the forecasting model 'F'). For 'BMA MED-LEV' this implies a prior inclusion probability of 5/31 for leverage, as well as for each macro-financial variable. However, if this prior would be applied indiscriminately to 'BMA 8-LEV', it would imply a higher joint inclusion probability (68%) for all leverage variables than under 'BMA MED-LEV'. Therefore we down-weight the prior on including leverage variables in 'BMA 8-LEV' (cf. Appendix).

3.4.1 In-sample results (forecasting equation (1))

In-sample, models with many regressors achieve the best fit (i.e. the lowest RMSEs). For GDP, industrial production (IP), the unemployment rate (UE), and investment (I), the in-sample forecast regressions with the four macro-financial factors (model variant labeled 'F') generate an RMSE that is about 25%-33% smaller than that of the benchmark 'Just ΔY' model; by contrast, the macro-financial factors do not help a great deal in predicting the excess equity return. In-sample, some individual sectoral leverages too perform well. In particular, FIN-MV leverage stands out, with relative RMSEs for GDP, IP, UE and I in the range 0.77-0.86. INS and SBD leverage yields relative RMSEs of 0.9 for the excess equity returns, and of 0.94-0.96 for GDP. Also, HH leverage is helpful in predicting the unemployment rate, while BNK-MV leverage helps predict GDP. The combined leverage indicators (principal component and median of standardized YoY changes of sectoral log leverages) tend to outperform the individual leverage variables, for all four real activity variables (see 'PC-LEV' and 'MED-LEV' models).

Table 3.2 (left panel) reports estimated slope coefficients for leverage (as well as R^2 coefficients of the corresponding regressions), based on the (non-rolling) regressions for 1993q3-2010q3. Note that almost all the leverage coefficients in the forecast equations for GDP, industrial production, investment and the excess equity return are negative, while the slope coefficients for unemployment are positive. *All* slope coefficients of the median and the principal component of leverage ('MED-LEV' and 'PC-LEV'), and of Flow of Funds insurance leverage are highly statistically significant (for the other individual leverage variables, the slope coefficients in the GDP-regressions are likewise mostly highly significant). We also estimated regressions in which the 8 sectoral leverage variables are included jointly (not reported in Table). Wald tests show that, for each dependent variable, the 8 leverages are overwhelmingly jointly significant (probability value in the range of 10^{-6}).

Turning to the BMA results, we note that the model 'BMA NO-LEV' produces already very low in-sample RMSE (Table 3.1) without taking leverage into account. Including leverage entails RMSE gains similar to the ones achieved in OLS models, and covariate statistics point as well to considerable explanatory power of leverage: The 'BMA 8-LEV' setting (Table 3.11) corroborates the importance of leverage, with very high posterior probabilities for including all leverage variables ('Joint PIP of all Leverage'). The best-performing among these leverage predictors is either FIN-MV or SBD, depending on the variable to be forecasted. Nonetheless, the PIPs for several macro-financial controls typically exceed those for individual leverage variables.[1] However, the setting 'BMA MED-LEV' demonstrates the advantage of combining leverage information. Table 3.1 shows that the median of leverage growth rates is a more important predictor than all macro-financial controls in the case of GDP, industrial production and investment (and it ranks among the top four for stock market returns). In both BMA settings, the coefficients for leverage are negative (positive for unemployment), and their magnitudes (Table 3.12) correspond to the OLS results from in Table 3.2.

The in-sample evidence thus suggests that there exists a highly significant, negative link between leverage and future real activity.

3.4.2 Out-of-sample results – rolling forecast regressions (equation (1))

Out-of-sample forecasting performance based on the rolling regressions is worse than in-sample fit (see right panel of Table 3.1). This is especially the case for models with many regressors. The

[1] Among these, price indexes, interest rates, residential investment, and employment matter most. Note that individual PIP of these variables vary somewhat according to whether 'BMA 8-LEV' or 'BMA MED-LEV' is used.

out-of-sample predictive content for GDP of the model with the four macro-financial factors (model 'F') is very close to that of the ('Just ΔY') benchmark model (relative RMSE: 0.97), although the four factors have non-negligible predictive content for unemployment (relative RMSE: 0.76).

The out-of-sample forecasts generated by the 'MED-LEV' model variant (that uses the cross-sectoral median of YoY sectoral leverage changes as a predictor) likewise outperform the benchmark model; 'MED-LEV' also outperforms the model with four macro-financial factors ('F'), in predicting GDP (relative RMSE: 0.90). When added to the four macro-financial factors, leverage continues to achieve forecast improvements for GDP, as shown by the combined models 'F, PC-LEV' and 'F, MED-LEV', which suggests that leverage contains information on top of established predictors.[1] The combined leverage models 'F, MED-FoF' and 'F, MED-MV' likewise perform well for all four real activity variables; their RMSEs are in the same range, i.e. the performance of book-value and market-value based predictions is roughly comparable. The model variants with the sectoral leverages for INS and SBD (insurance; securities brokers-dealers, from the Flow of Funds) perform marginally better than the benchmark model, but are outmatched by the four macro-financial factors in forecasting GDP. Finally, the efficacy of professional forecasts (SPF) is basically comparable to that of the benchmark model. None of the examined predictors help in forecasting excess stock returns out-of-sample, with the possible exception of the INS and SBD leverage measures.

As a further test of the out-of-sample forecasting capacity of leverage, we use the Clark and West (2007) 'MSPE-adjusted test' to test the null hypothesis that the RMSE of a given model is identical to that of the benchmark model ('Just ΔY'); see Table 3.3. For each dependent variable, the test is separately applied to the different alternative forecast models. For the model variants that include the principal component or the median of YoY changes in sectoral log leverages as a predictor, the p-values of the test are mostly below 0.05 (except for the excess equity return), which suggests that the predictive power of *joint* leverage indicators is statistically significant – the same holds for the macro-financial factors.[2]

We also use the Hubrich and West (2010) 'max-t-stat' test to test the joint null hypothesis that all of the eight models that include a single sectoral leverage variable ('CB', 'INS', ...,'FIN-MV') have the same predictive content as the 'Just ΔY' benchmark model. This test is separately applied for each of the predicted variables. The p-values for GDP, industrial production, the unemployment rate, investment and the excess equity return are 0.026, 0.019, 0.065, 0.035 and 0.068, respectively. These low p-values too suggest that the predictive power of the sectoral leverage information is statistically significant.

Table 3.1 indicates that, out of sample, the BMA set-up without leverage variables ('BMA NO-LEV') performs less well than the 'Just ΔY' model in terms of RMSE. The gains from adding leverage ('BMA MED-LEV', 'BMA 8-LEV') are at most marginal. This comparatively poor forecasting performance of BMA conforms to literature findings (Stock and Watson 2006). Nonetheless, average posterior inclusion probabilities consistently rank the median of leverage growth rates among the three most important predictors (Table 3.14). Likewise the 8 individual leverage variables

[1] That result continues to hold when each of the individual macro-financial variables is used as a predictor (instead of principal components); use of individual macro-financial variables worsens RMSEs, but leverage remains a significantly negative predictor of future real activity growth.

[2] We also used the Clark and West (2007) test to compare models that use the four macro-financial factors and leverage as predictors, against the 'F' model. Tests of model 'F, MED-LEV' vs. 'F' yield p-values 0.08, 0.01, 0.09, 0.01 and 0.65, respectively, for GDP, IP, UE, I and Rx. Thus, joint leverage information has statistically significant predictive value added (for real activity), on top of the macro-financial factors. By contrast, individual sectoral leverage series generally yield no significant predictive improvement, when added to the macro-financial factors. (See Web Appendix for detailed results.)

display high joint PIP, with financial book-value leverage indicators scoring well (Table 3.13).

The rolling regressions again show a negative link between leverage and future real activity. For each model that includes leverage as a regressor, Table 3.2 (right panel) reports the fraction of rolling 40-quarter estimation windows in which the estimated leverage coefficient is negative *and* statistically significant at the 10% level (as well as the fraction of rolling samples in which the slope coefficient is significant, irrespective of sign; see figures in parentheses). In the forecast equations for GDP, industrial production and investment, most slope coefficients of leverage are negative and statistically significant (consistent with this, most leverage coefficients in the forecast equations for unemployment are positive). The median of YoY leverage growth rates (model 'MED-LEV') as well as their principal component (model 'PC-LEV') both feature a negatively significant coefficient (positive for unemployment) for the vast majority of samples.[1]

Figure 3.2 plots standardized regression coefficients of leverage and their p-values, across the rolling estimation windows, for the GDP forecast equations. For most of the sectoral leverage variables, the estimated slope coefficients are negative, across all windows. Hence, the empirical finding that leverage is negatively related to future real activity is not sample dependent – in particular, this result is *not* driven by the financial crisis. However, none of the sectoral leverage variables are highly significant across *all* estimation windows. Note, for example, that the slope coefficient of securities brokers-dealers (SBD) leverage was significant at the beginning and end of the sample, but insignificant (and close to zero) in the middle of the sample. However, jointly the eight sectoral leverage variables are highly significant predictors – and that in *each* of the estimation windows (this is shown by Wald tests not reported here). This again suggests that the *joint* information contained in the eight sectoral leverage series is more relevant for predicting future real activity than the information contained in any individual leverage series.

However, despite the strong (joint) significance of the leverage variables, these variables would not have allowed to predict the 2008-2009 'Great Recession' better than conventional predictors. This is shown in Figure 3.3, which plots the GDP forecasts (rolling window based) generated by the model with the four macro-financial factors ('F'), and by the model with these four factors *and* the median of leverage growth rates ('F, MED-LEV'). Both models fail to predict the dramatic fall in GDP during the recession – in fact, both models yield essentially the *same* predictions for GDP, for 2008-2009. Figure 3.3 reveals that the overall RMSE reduction produced by using leverage information mainly reflects smaller forecast errors made during the early 2000s (after the collapse of the dotcom bubble). The forecasts from BMA point to the same conclusion: Including leverage in BMA improves forecasts marginally during the early 2000s, but both 'BMA NO-LEV' and 'BMA MED-LEV' (Figures 3.6–3.7) miss the recession and subsequent recovery by a wide margin. The striking difference between (PC)OLS and BMA forecasts is that the least squares predictions times the initial growth slump well, but then overshoots (particularly because of the importance of the lagged growth term in (1)), whereas the BMA forecast during the crisis period is determined by the lag indicator (leverage affects the BMA forecast less than the least squares forecast). Both variants, however, fail at gauging the depth and timing of the recession and the scale of recovery. This is partly due to the fact that leverage indicators followed through on the deleveraging process after 2009, whereas the macro-financial control factors did not indicate a massive turnaround ahead of 2008.

[1] When controlling for the macro-financial factors (models 'F, MED-LEV' and 'F, PC-LEV'), the leverage coefficient remains significantly negative for a majority of rolling samples (for each real activity variable).

3.4.3 Robustness: lag structure, non-linearities

The discussion so far has focused on forecast models that use past year-on-year changes of leverage as a predictor of real activity. Due to the sustained buildup of leverage before the crisis, it seems interesting to also consider (past) leverage growth over a period longer than one year, as a predictor of future real activity. We experimented with model variants in which leverage growth over 6, 8 and 12 quarters is used as a predictor–those variants have lower predictive accuracy than the baseline models (with YoY leverage growth). However, the link between past leverage growth and future real activity remains negative and highly significant. Tables 3.4 and 3.5 show results for a model variant in which 8-quarter leverage growth is used as a predictor.

We also considered forecast equations that use, as predictors, past changes of real activity over longer lags ($Y_t - Y_{t-s}$ for $s > 1$) and macro-financial factors (Φ_t) based on changes of variables over more than one quarter.[1] These alternative specifications confirm our results about leverage; in fact, leverage emerges as a slightly more significant negative predictor than in the baseline models. The Web Appendix also reports results for forecast regressions that predict changes of real activity at a longer horizon (than the 4-quarter horizon considered so far). Again, the negative link between leverage and future real activity remains (but it is weaker, the longer the horizon).[2] A VAR analysis likewise confirms this finding: positive innovations to leverage trigger a prolonged decrease in GDP growth (see Web Appendix).

Furthermore, it seems interesting to investigate whether there are non-linearities and asymmetries in the link between leverage and (future) real activity. We thus consider the following extension of equation (1):

$$Y_{t+4} - Y_t = \beta_0 + \beta_1 (Y_t - Y_{t-1}) + \beta_2 \Phi_t + \beta_3 \Lambda_t + \beta_4 f(\Lambda_t) + \epsilon_{t+4} \tag{2}$$

where $f(\Lambda_t)$ represents a non-linear function of the (year-on-year) leverage measure Λ_t. We use $f(\Lambda_t) = \max(0, \Lambda_t)$, to allow responses to differ across positive and negative (YoY) leverage changes (Kilian and Vigfusson, 2009); we also consider $f(\Lambda_t) = (\Lambda_t)^2$, in order to test for a disproportionate impact of large swings in leverage. As shown in Tables 3.6–3.8, the augmented models exhibit considerably worse out-of-sample forecasting performance. However, the link between future real activity growth and leverage growth Λ_t remain negative and significant. By contrast, the coefficients for the asymmetry terms are rarely significant and their signs show no consistent pattern. We also allowed for non-linear effects of leverage changes over lags greater than one year. Again, we find no support for significant non-linearities. Furthermore, we tested for multiplicative interactions between leverage and the other regressors. These interactions terms are rarely significant, and they markedly reduce out-of sample forecast performance (but leverage growth remains strongly negatively significant as a predictor of future real activity growth). See the Web Appendix for the detailed sensitivity and robustness results.

In summary, the in-sample and out-of-sample results suggest very clearly that leverage is a statistically significant negative predictor for future real activity growth. However, *quantitatively*, the

[1] Specifically, the quarterly growth rates and returns used in the construction of the factors (see right-most column in Panel (c) of Appendix) were replaced by growth rates and returns over $s > 1$ quarters.

[2] As mentioned above, we also ran forecast regressions that include leverage and between 1 and 8 macro-financial principal components (factors) as predictors (NB the specifications above use 4 factors). The highly statistically significant negative link between leverage and future real activity holds irrespective of the number of included macro-financial control factors–the slope coefficients of leverage and their p-values are very stable when the number of factors is changed. This holds both for regressions over the whole sample periods, and for the 40-quarter rolling estimation windows. (See Web Appendix.)

effect of using leverage as a predictor is modest – leverage information would not have generated an 'early warning' of the 2008-2009 recession any better than conventional macro-financial predictors.

The negative relation between leverage and future real activity growth seems broadly consistent (at least qualitatively) with the predictions of recent structural macro models with financial intermediaries.[1] Theoretical research also suggests that leverage might matter for the conditional volatility of future real activity and returns: essentially, an increase in leverage today should amplify the effect of future shocks.[2] This would imply a positive link between leverage and uncertainty about future economic conditions. The next Section documents that such a link exists in the data – and that it is powerful.[3]

3.5 Leverage and the conditional variability of real activity and equity returns

We evaluate the link between the date t YoY change in log leverage and the following three measures of uncertainty about future economic conditions:

i The absolute value of date $t+4$ forecast errors (in %) implied by the date t forecasts generated by the forecast models discussed in the previous Section.

ii (ii) The CBOE equity volatility index (VIX) at the end of period t – VIX is an estimate of the future volatility of stock prices (inferred from options prices).

iii (iii) The measure of dispersion (in %), across forecasters, of date t forecasts for real activity growth between t and $t+4$, reported by the Philadelphia Fed Survey of Professional Forecasters (SPF).[4]

Figure 3.4 presents scatter plots of the three measures of conditional future volatility/dispersion against the principal component of the changes in log sectoral leverage between $t-4$ and t (observed at t). (The sample period (t) is 1992q3-2009q3.) Figure 3.5 plots time series of these variables (using the same timing convention). The absolute forecast errors in Figures 3.4–3.5 pertain to GDP; these errors are rolling-window-based, and were generated using the forecast model referred to as 'F' in the previous Section (i.e. the four macro-financial factors are used as predictors). (Plots for errors generated by the other forecast models, and for the other predicted real activity measures are very similar.) Figures 3.4–3.5 show a clear positive link between leverage information at t and

[1] See, e.g., the models in Aikman and Paustian (2006), Meh and Moran (2010) and Kollmann et al. (2011). These models predict that a negative transitory shock to private sector total factor productivity lowers bank leverage. Such a shock also lowers GDP, on impact, but subsequently GDP is predicted to revert to its pre-shock level; hence, leverage is negatively correlated with future GDP growth. (The mechanisms that induce the fall in leverage differ across models; e.g., in Kollmann et al. (2011) it is due to the fact that household savings and the supply of deposit decrease, which leads banks to finance a larger share of their asset holdings by raising equity). An exogenous negative shock to bank capital is likewise predicted to decrease future output. We leave for future research a detailed empirical evaluation of this new class of models.

[2] See, e.g., Krugman (2009), Devereux and Yetman (2010), and Kollmann and Malherbe (2011) for discussions of mechanisms through which leverage may amplify the effect of shocks.

[3] Previous research has documented that the conditional volatility of real activity is time-varying (e.g., Giannone et al. (2008), and Frale and Veredas (2009)). Our results about the link between *leverage* and future conditional volatility of real activity are novel, to the best of our knowledge.

[4] The SPF dispersion measure is the % difference between the 75th and 25th percentiles of the cross-sections of forecasts.

the measures of future conditional variability (and the dispersion of forecasts made at t). The link is very pronounced during the crisis – but it is also clearly present in the pre-crisis period.

Tables 3.9 and 3.10 provide regression evidence on the link between leverage and conditional future volatility/dispersion. Table 3.9 regresses absolute date $t+4$ forecast errors for each of our five dependent variables on (annual YoY changes of) our 8 sectoral log leverage variables observed at t (see first 8 rows of Table); we also regress absolute forecasts errors on all 8 sectoral leverage series *jointly*, and on the principal component and median value of (YoY changes of) sectoral log leverages. Table 3.9 furthermore shows results that obtain when the four macro-financial factors are added as regressors. In Table 10, the cross-sectional dispersion of date t SPF forecasts (for GDP, industrial production, the unemployment rate and investment; see Columns (1)-(4)), as well as the VIX at t (Column 5) are regressed on the regressors used in Table 3.9.

In almost all regressions, the slope coefficients of leverage are positive and highly statistically significant.[1] This result confirms the existence of a powerful positive link between leverage and conditional future variability/dispersion. That link is particularly strong for the leverage factor and median leverage. Each of these two leverage measures alone explains between 20% and 35% of the variances of the absolute GDP forecast errors, of SPF cross-sectional GDP forecast dispersion, and of VIX (see R^2 coefficients).[2] The four macro-financial factors are likewise related to future conditional volatility – but less strongly than leverage (lower R^2s). Furthermore, the principal component and median of the sectoral leverage measures remain highly significant when the four macro-financial factors are added as predictors.

Table 3.15 provides BMA evidence on the link between absolute forecast errors and the median of leverage growth rates as well as the 30 macro-financial controls. The results suggest that several of the macro-financial controls have an impact on conditional output variability that is comparable to that of leverage.[3] Nonetheless, leverage consistently ranks first or second among predictors (except for industrial production) in terms of PIP. Its coefficients are positive and display considerably higher magnitudes than most of the controls. BMA results thus corroborate our conclusion from the OLS forecast exercise that there is a significantly positive relationship between leverage growth and conditional output variability.

3.6 Conclusion

This paper documents a statistically significant negative link between leverage and future real activity, and a significant positive link between leverage and the conditional volatility of future real activity. These links appear particularly clearly when information from sectoral leverage series is combined using cross-sectional medians or principal components. The results here show that the predictive power of leverage is roughly comparable to that of macro and financial predictors commonly used by forecasters. However, leverage information would *not* have allowed to predict the 'Great Recession' of 2008-2009 any better than conventional macro/financial predictors.

[1] There is only one *notable* exception: in about half of the regressions, securities brokers-dealers (SBD) leverage is negatively linked to the volatility/dispersion measures.

[2] As pointed out by a referee, a GARCH-type model could also be used to assess the effect of leverage on volatility. A one-period-ahead GARCH variance equation that includes leverage as a regressor supports the notion that leverage is significantly positively related to future output volatility. (Implementing a GARCH set-up for a horizon of four (or more) quarters raises major challenges beyond the scope of this paper.)

[3] These stem from five covariate groups: interest rates, price indexes, employment, investment, and capacity utilization.

3.A Data sources and definitions of variables

3.A.1 Predicted variables

Series label	***Variable***	***Source***
GDP	Real gross domestic product	Bureau of Economic Analysis
IP	Industrial production index	St. Louis Fed
UE	Civilian unemployment rate, percent	Bureau of Labor Statistics
I	Real gross private domestic investment	Bureau of Economic Analysis
Rx	Excess stock return (French-Fama 'MKT' return – T-bill return)	K. French website

3.A.2 Leverage variables

Series label	***Variable***	***Source***
SBD	Securities Brokers and Dealers leverage ratio	Flow of Funds
CB	Commercial Banks leverage ratio	Flow of Funds
INS	Life and casualty insurance leverage ratio	Flow of Funds
SBD	Securities Brokers and Dealers leverage ratio	Flow of Funds
HH	Households and nonprofit organizations leverage ratio	Flow of Funds
BUS	Non-farm non-financial corporate business leverage ratio	Flow of Funds
BNK-MV	US-Banks index : total assets / equity at market value	Datastream
INS-MV	US-Insurance index : total assets / equity at market value	Datastream
FIN-MV	US-Fin. Services index : total assets / equity at market value	Datastream

3.A.3 Variables used to construct macro-financial factors

Variable	***Source***	***Transformation***
1) Real gross domestic product	Bureau of Econ. Analysis	Quarterly growth rate
2) Real government consumption and investment	Bureau of Econ. Analysis	Quarterly growth rate
3) GDP implicit price deflator	Bureau of Econ. Analysis	Quarterly growth rate
4) Real gross private domestic investment	Bureau of Econ. Analysis	Quarterly growth rate
5) Gross government saving, as share of GDP	Bureau of Econ. Analysis	Quarterly difference
6) Private housing starts of 1-family structures	Bureau of Econ. Analysis	Quarterly growth rate
7) Real personal consumption expenditures	Bureau of Econ. Analysis	Quarterly growth rate
8) Real personal consumption expenditures, durable goods	Bureau of Econ. Analysis	Quarterly growth rate
9) Real private non-residential fixed investment	Bureau of Econ. Analysis	Quarterly growth rate
10) Real private residential fixed investment	Bureau of Econ. Analysis	Quarterly growth rate
11) Real net exports of goods & services, as share of GDP	Bureau of Econ. Analysis	Quarterly difference
12) Total number of employees (non-farm)	Bureau of Labor Statistics	Quarterly growth rate
13) Commodities producer price index	Bureau of Labor Statistics	Quarterly growth rate
14) Civilian unemployment rate, percent	Bureau of Labor Statistics	Quarterly difference
15) Consumer price index, all urban consumers	Bureau of Labor Statistics	Quarterly growth rate
16) Oil price (spot WTI) USD/barrel	Dow Jones & Company	Quarterly growth rate
17) 3-month U.S. T-bill	Federal Reserve Board	Quarterly return
18) 2-year U.S. Treasury bond	Federal Reserve Board	Quarterly return
19) 5-year U.S. Treasury bond	Federal Reserve Board	Quarterly return
20) U.S. Treasury term spread: 10yr – 3month par yield	Federal Reserve Board	Quarterly difference
21) ISM manufacturing inventories index	St. Louis Fed	Quarterly difference
22) ISM manufacturing new orders index	St. Louis Fed	Quarterly difference
23) Industrial production index	St. Louis Fed	Quarterly difference
24) Nominal M2 money stock	St. Louis Fed	Quarterly growth rate
25) Total industry capacity utilization	St. Louis Fed	Quarterly growth rate
26) French-Fama HML factor	K. French website	—
27) French-Fama Momentum factor	K. French website	—
28) French-Fama SMB factor	K. French website	—
29) French-Fama Short-term reversal factor	K. French website	—
30) French-Fama Long-term reversal factor	K. French website	—

3.A.4 Other variables

Variable	***Description***	***Source***
VIX	Equity Volatility Index	Chicago Board Options Exchange
SPF median forecasts investment	Median forecasts for GDP, industrial production, unemployment rate and	Survey of Professional Forecasters, Philadelphia Fed
SPF cross-sectional dispersion of forecasts	% difference between the 75th and 25th percentiles of forecasts	Survey of Professional Forecasters, Philadelphia Fed

Note: The Flow of Funds leverage ratio for commercial banks (CB) displays a break in 1999. We corrected for this break by projecting the CB leverage ratio on a time dummy and a linear and quadratic trend, and then adjusting the raw series for the dummy coefficient.

Section 3.A.3 lists the 30 variables from which the four macro-financial factors (used as predictors) are extracted (principal components). The right-most column lists the data transformations used in constructing the factors. Where applicable, variables were used in seasonally-adjusted form as provided by the data source. Returns on Treasury bonds are derived from constant maturity yield curves (estimated using the methodology of Gürkaynak et al. (2007)), as published on the web page of the Federal Reserve Board.

3.B Implementation of BMA

Bayesian Model Averaging (BMA) addresses uncertainty about which among K regressors to include in a linear model such as equation (1). Each of the 2^K potential variable combinations is considered as a model M_γ to which a prior model probability $p(M_\gamma)$ is assigned. After observing the data y, this prior belief is updated to a posterior model probability via Bayes' theorem:[1]

$$p(M_\gamma|y) = \frac{p(y|M_\gamma)p(M_\gamma)}{\sum_{i=1}^{2^K} p(y|M_i)p(M_i)}$$

The resulting posterior model probabilities then apply to constructing (the posterior distribution for) any statistic of interest θ by weighting the individual model statistics with $p(M_\gamma|y)$, thus integrating over the posterior model distribution. In particular, this approach is useful for model-weighted posterior regression coefficients $p(\beta|y)$.[2]

$$p(\theta|y) = \sum_{\gamma=1}^{2^K} p(\theta|M_\gamma, y)p(M_\gamma|y)$$

The individual forecast models M_γ are estimated via the popular natural-conjugate 'Bayesian regression' (Zellner 1971) with a 'g-prior', as its closed-form posteriors ease the computational burden. We specify prior beliefs on the individual model coefficients as in Fernández et al. (2001a), representing ignorance by centering the prior coefficient distribution at zero. The coefficient variance-covariance prior is set proportional to the one arising from OLS, which simplifies posterior expressions: Model-specific posterior coefficients equal their OLS estimator times a shrinkage factor. The marginal likelihood $p(y|M_\gamma)$ resembles information criteria such as BIC in connecting the OLS R-squared with a size penalty.[3] In this framework, the crucial component is a 'shrinkage' hyperparameter that indicates how tightly the researcher's prior is centered at zero. In order to minimize prior sensitivity, we follow the recommendation by Ley and Steel (2011a) and opt for endogenous updating of this hyperparameter under the 'unit information hyper-prior' formulation by Feldkircher and Zeugner (2009).

We apply BMA to assess one-year-ahead forecast equations as in (1), jointly considering the leverage variables together with the 30 macro-financial controls as described in Section 3.3.1. As it is numerically infeasible to assess the billions of potential models under these settings, we approximate the likelihood surface with a Metropolis-Hastings model sampler as in Fernández et al. (2001a).[4] The BMA set-up discussed so far still requires the crucial choice of the model priors $p(M_\gamma)$. In

[1] Here, $p(y|M_\gamma)$ denotes the posterior marginal likelihood of the individual model M_γ that obtains from integrating the likelihood $p(y|\beta, \sigma, M_\gamma)$ over the posterior distributions of its parameters (β, σ).

[2] Tables 3.11–3.15 in the Appendix report conditional coefficients (i.e. posterior expected coefficients averaged over the models where they are not restricted to zero) for comparison with the OLS forecast models. Many BMA applications present unconditional coefficients instead (i.e. posterior expected coefficients averaged over all models). The unconditional coefficient is defined as the expected coefficient over all models (i.e. including the ones restricting the coefficient to zero) and are equal to the conditional coefficient times PIP.

[3] Note that integration over a model's parameters yields posterior densities of any kind for each model (in particular coefficient densities). This framework naturally extends to formulating a posterior predictive distribution $p(y_{T+h}|\{yt\}_{t=1}^{T}, M\gamma)$, applied in Figure 3.6–3.7.

[4] For each of the rolling samples and each dependent variable, we use 2 000 000 iterations of this sampler, after discarding 400 000 'burn-in' draws. Diagnostic statistics indicate that this approximates the likelihood surface very well for each sampling chain: For the highest-ranking 2 000 models encountered by the sampler, the correlation between MCMC frequencies and analytical PMP is above 0.99 in each case.

most applications, model priors are chosen to not discriminate between potential regressors, but to impose a parameter size penalty in order to reflect a preference for parsimonious models.[1] In our set-up, we follow the popular binomial model prior of Sala-i-Martin et al. (2004), which specifies equal prior inclusion probabilities for each covariate. For 'BMA NO-LEV' and 'BMA MED-LEV', we tune these prior inclusion probabilities such that the prior expected model size is equal to five – which is equal to the number of predictors in the OLS factor model 'F'.[2]

However, applying the same model prior to 'BMA 8-LEV' would imply a relatively high prior joint inclusion probability for leverage (eight times more than that for an individual control). Under such a prior, it would thus be eight times more likely to find at least one leverage variable as important as a single macro-financial control (or as median leverage growth under 'BMA MED-LEV'). We consequently specify a more conservative prior for leverage variables that constrains $p(M_\gamma)$ to zero for any model that includes more than one leverage regressor.[3] At the same time, we keep the prior inclusion probabilities for the controls similar to 'BMA MED-LEV', such that the prior expected parameter is again equal to five. This translates into a prior inclusion probability of just 2.7% for individual leverage variables, compared to 10% for the macro-financial controls.[4]

[1] Note that adjusting priors for parameter size is particularly important when comparing BMA settings with differing numbers of regressors K. If one did not adjust for parameter size, a setting with larger K would expect more predictive power a priori.

[2] We did experiment with other prior expected parameter sizes (as well as with uniform model priors as in Faust et al. (2011)): The conclusions on the relative importance of leverage remain similar, but the absolute levels of posterior inclusion probabilities vary such that results are less comparable to the OLS results (the relative ranking of PIP remains similar to the presented output).

[3] At its extreme, such a prior can be seen as a variant of a 'dilution prior' as in George (2010, chapter 4), that defines models priors according to a collinearity indicator for each model (with the prior used here being a Dirac collinearity prior based on group adherence). Note, however, that in our case 'dilution' only applies to models including leverage variables, and that this prior is combined with a model size-based prior.

[4] A similar model prior is employed for the BMA estimation on conditional forecast variability (Table 3.15).

3.C Charts and Tables

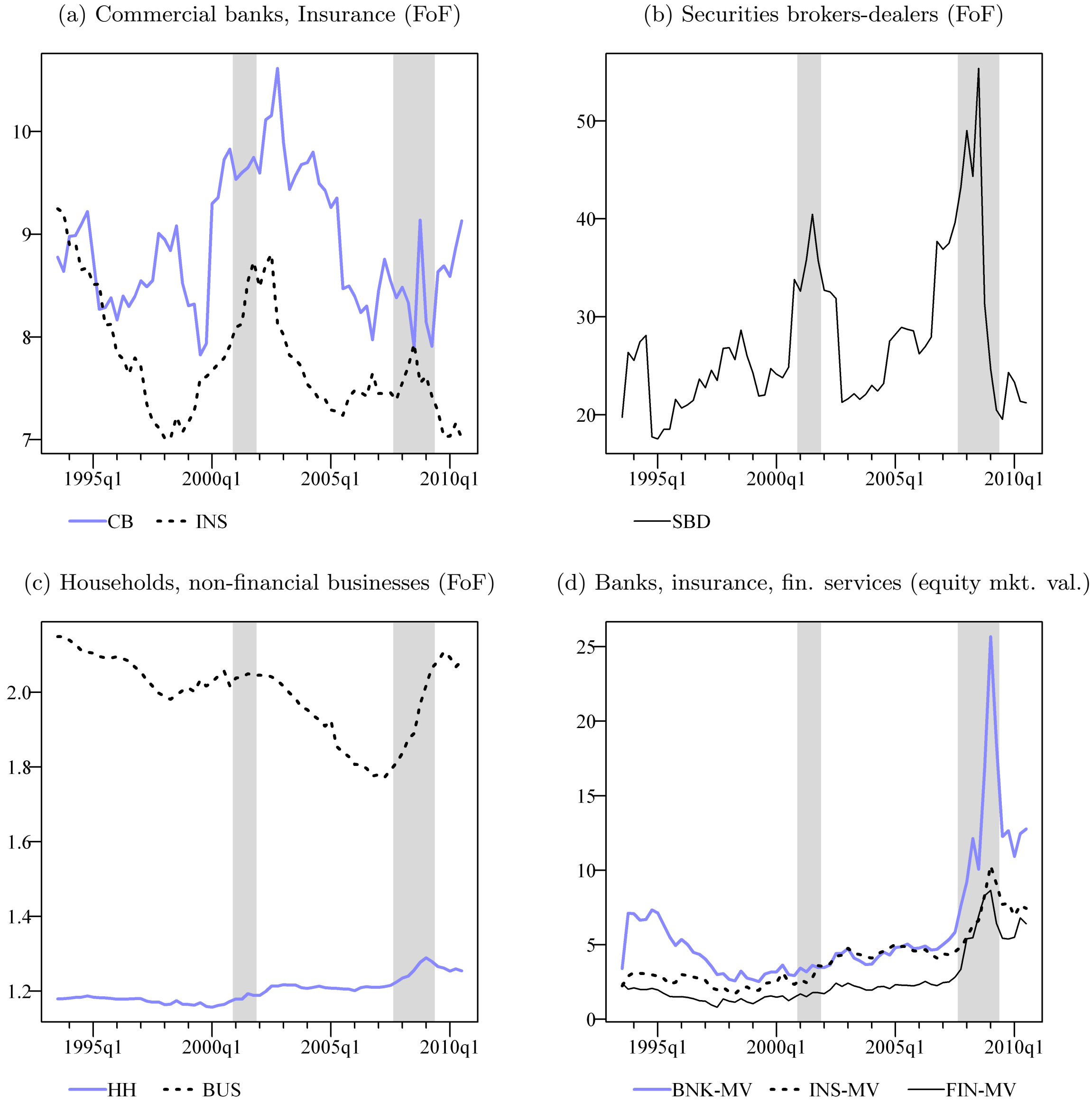


Figure 3.1: Leverage ratios

The Figure plots the time series of leverage ratios for the following sectors – CB: commercial banks (from Flow of Funds, FoF); INS: insurance (FoF); SBD: securities brokers and dealers (FoF); HH: households (FoF); BUS: non-financial corporate businesses (FoF); BNK-MV, INS-MV, FIN-MV: Banks, insurance and financial services, respectively, based on equity market values. Sample period: 1993q3-2010q3. Shaded areas indicate NBER recessions.

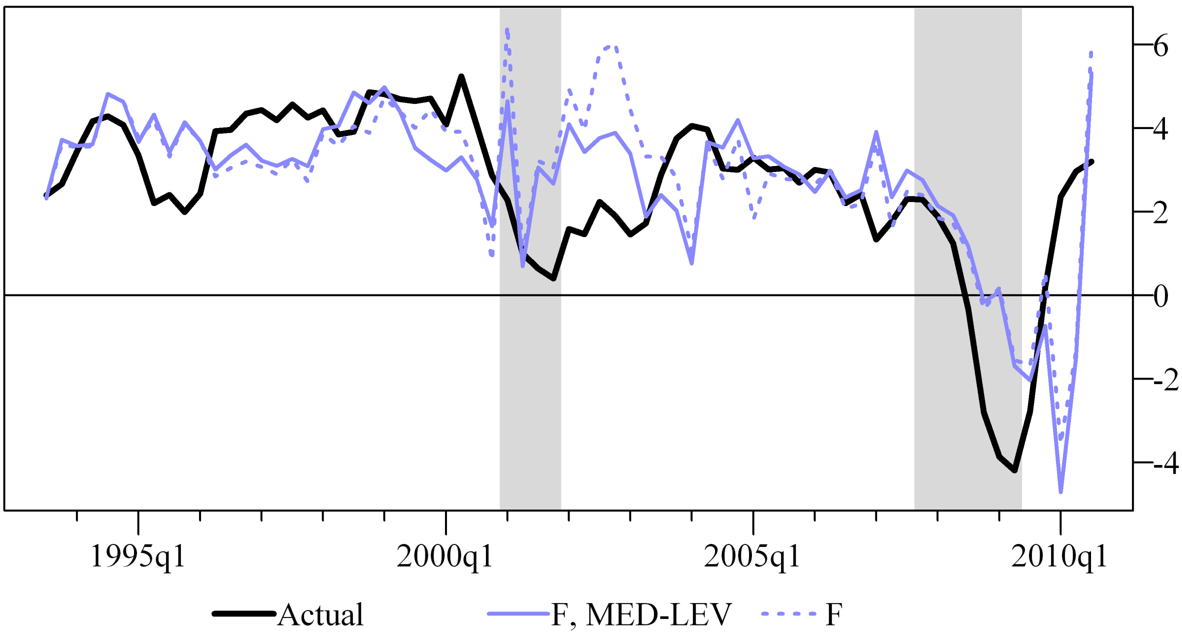


Figure 3.2: Year-on-year GDP growth rate (in %) actual and predicted
The Figure shows actual and predicted year-on-year GDP rates growth (in %), 1993q3-2010q3. The line labeled 'F' shows the prediction (based on 40-quarter rolling estimation windows) generated by a forecast model that includes (as predictors) four factors extracted from a set of 30 macro-financial variables. The line labeled 'F, MED-LEV' shows the prediction obtained by adding the median of the annual growth rates of the 8 sectoral leverage series, as a predictor. Shaded areas indicate NBER recessions.

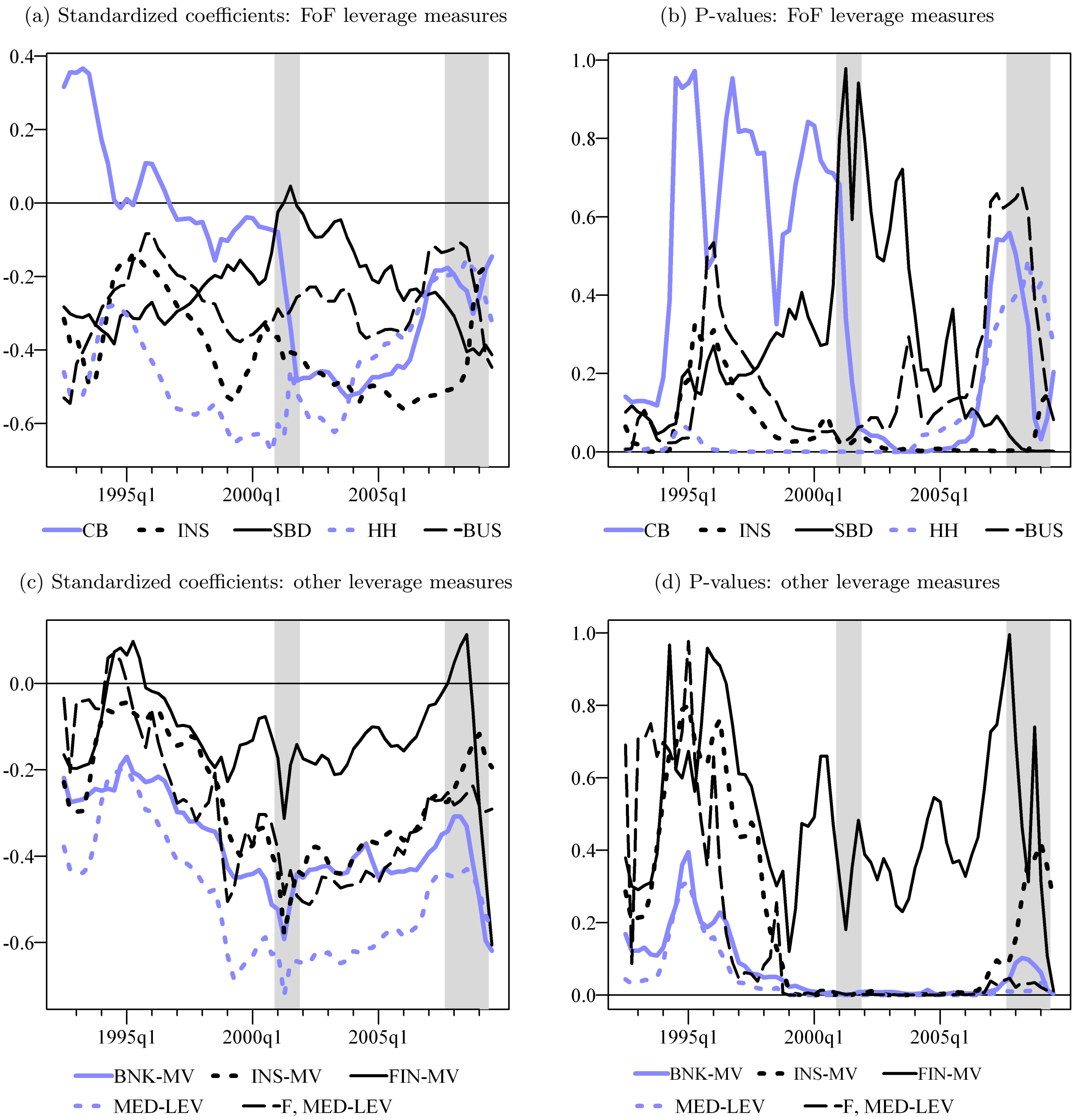


Figure 3.3: Slope coefficients of leverage and p-values, from rolling forecast regressions for GDP
Panels (a) & (c): Standardized regression coefficients of leverage variables in forecast regression for GDP (based on 40-quarter rolling windows). *Panels (b) & (d):* probability-values from Newey-West HAC t-statistics for slope coefficients of leverage (from GDP forecast regressions). Each forecast regression includes the following predictors: a constant, the quarterly first difference of GDP, and one leverage variable. Date (abscissa) indicates final observation of the 40-quarter estimation window for the explanatory variables. Shaded areas indicate NBER recessions.

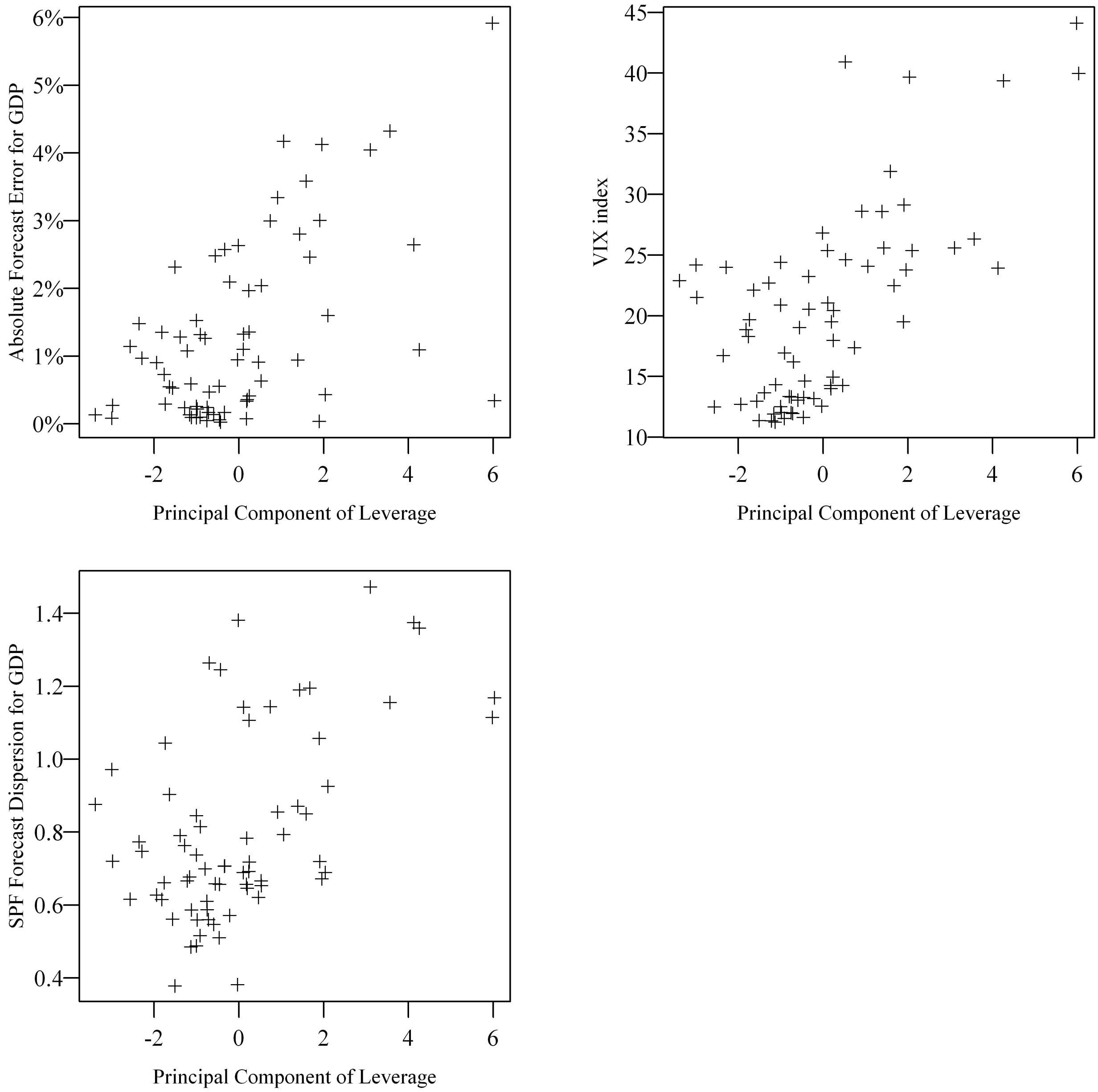


Figure 3.4: Scatter plots of future absolute forecast errors, VIX and forecast dispersion vs. leverage The Figure shows scatter plots of absolute forecast errors for GDP (in %) between t and $t+4$, of Equity Volatility Index (VIX) at t, and of date t cross-sectional dispersion of SPF forecasts of GDP growth between t and $t+4$ vs. the first principal component (of the YoY change) of sectoral log leverages between $t-4$ and t. The forecast errors pertain to forecast model 'F' (four macro-financial factors used as predictors), based on rolling 40-quarter estimation windows. The sample period (t) is 1992q3-2009q3.

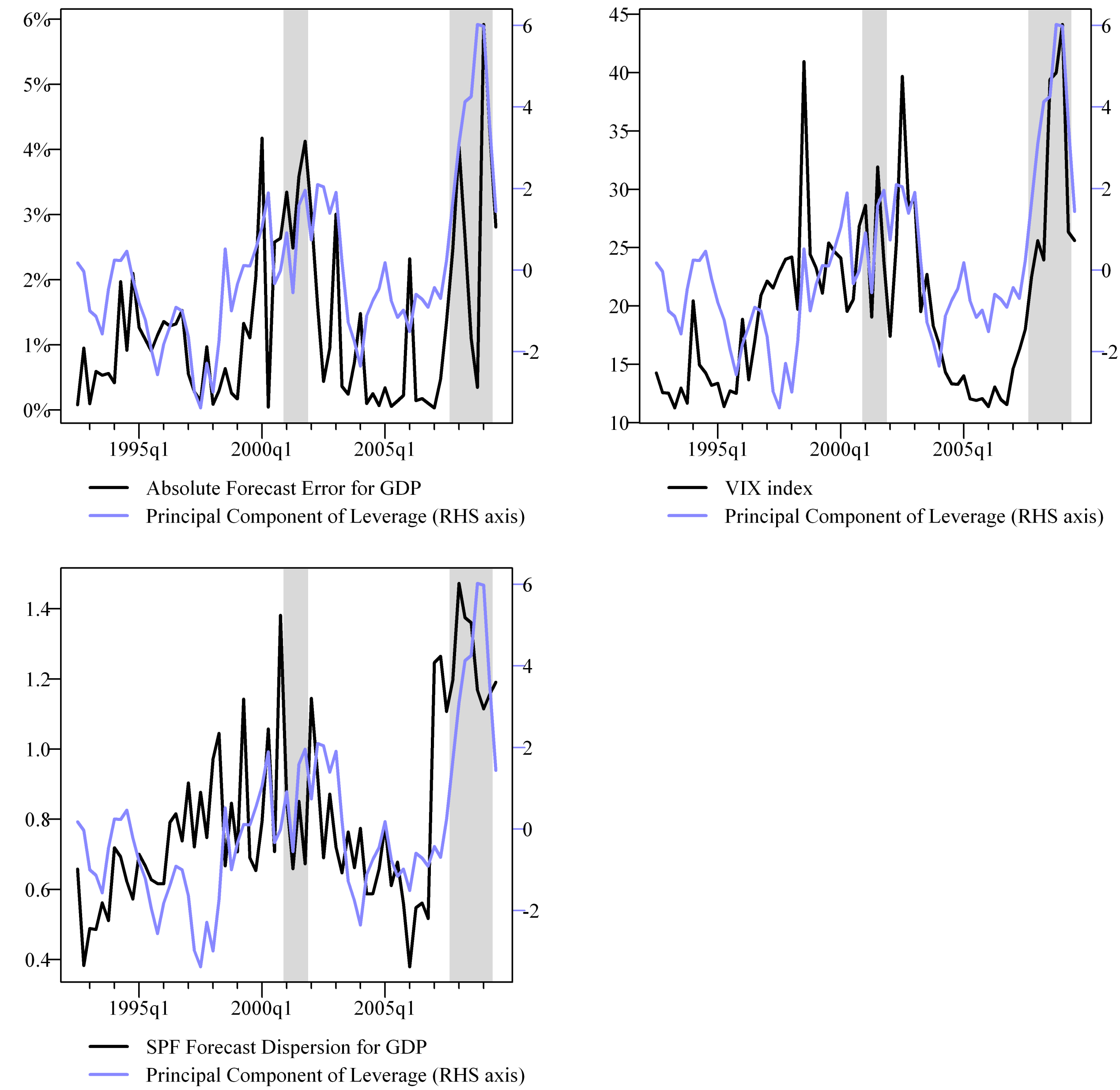


Figure 3.5: Time series plots of absolute future forecast errors, VIX, forecast dispersion and leverage Each panel shows time series plots the first principal component of the YoY change in sectoral log leverages between $t-4$ and t, and another variable. *Panel (a):* absolute forecast error for GDP (in %) between t and $t+4$. *Panel (b):* Equity Volatility Index (VIX) at t. *Panel (c):* SPF date t cross-sectional dispersion of forecasts of GDP growth between t and $t+4$. (Thus: same timing conventions as in Figure 4.) The forecast errors pertain to forecast model 'F' (four macro-financial factors used as predictors), based on rolling 40-quarter estimation windows. The sample period (t) is 1992q3-2009q3. Shaded areas indicate NBER recessions.

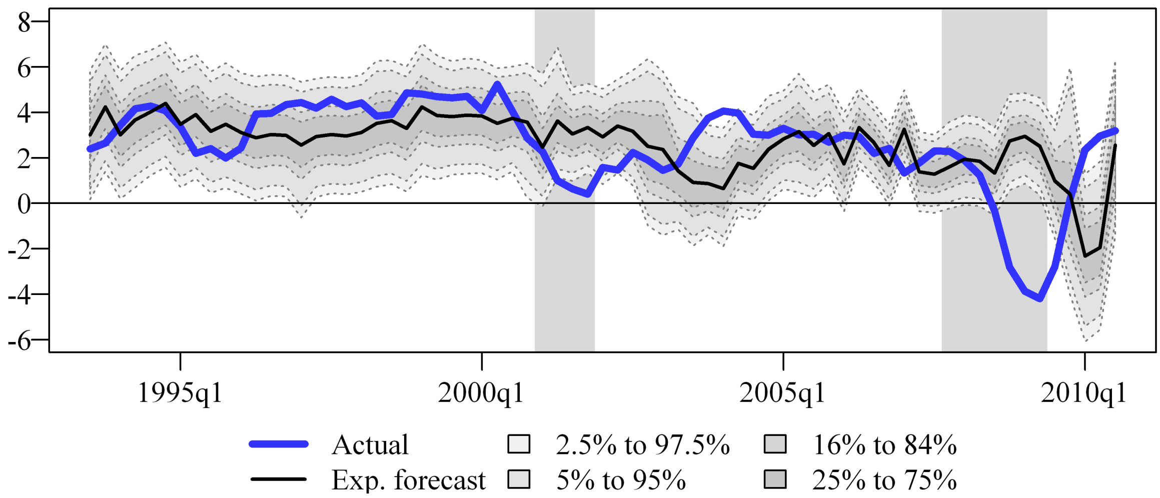


Figure 3.6: Year-on-year GDP growth forecasts from BMA without leverage
The Figure shows actual and predicted year-on-year GDP rates growth (in %), 1993q3-2010q3, based on setting 'BMA no LEV'. The black line shows the expected forecast (based on 40-quarter rolling estimation window) generated by a rolling BMA estimations based over 30 macro-financial variables. The lightly shaded areas indicate intervals of the BMA prediction density (corresponding to standard errors). The vertically oriented shaded areas indicate NBER recessions.

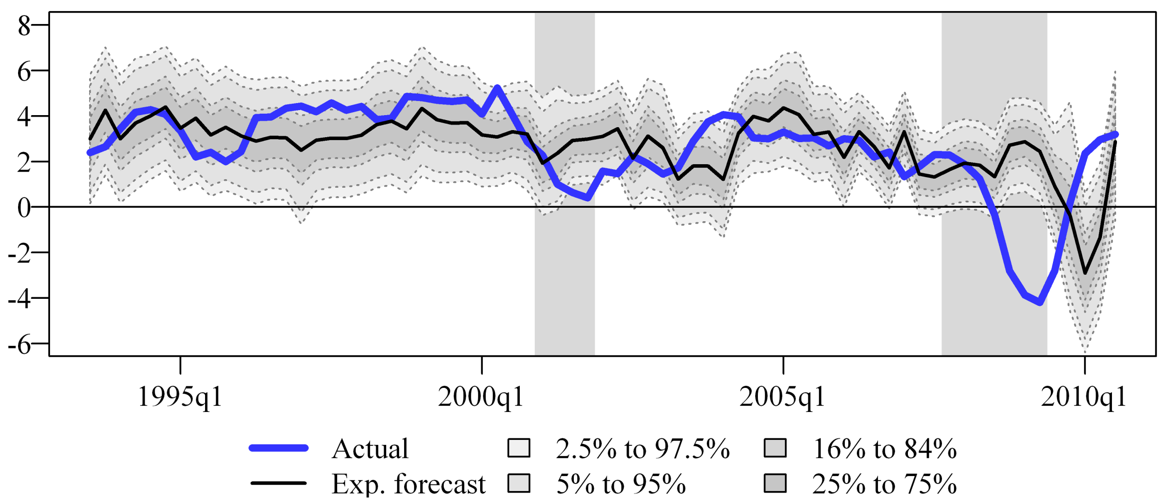


Figure 3.7: Year-on-year GDP growth forecasts from BMA including leverage
The Figure shows actual and predicted year-on-year GDP rates growth (in %), 1993q3-2010q3, based on setting 'BMA MED-LEV'. The black line shows the expected forecast (based on 40-quarter rolling estimation window) generated by a BMA estimations over 30 macro-financial variables as well as MED-LEV (median of the standardized YoY change of the 8 sectoral log leverage series). The lightly shaded areas indicate intervals of the BMA prediction density (corresponding to standard errors). The vertically oriented shaded areas indicate NBER recessions.

Table 3.1: RMSEs of 'Just ΔY' forecast model & relative RMSEs of other models and SPF

	In-sample RMSEs					***Out-of-sample RMSEs***				
Forecast model:	**GDP**	**IP**	**UE**	**I**	**Rx**	**GDP**	**IP**	**UE**	**I**	**Rx**
Just ΔY	*1.77*	*3.9*	*0.87*	*10.16*	*19.63*	*1.91*	*4.27*	*0.95*	*10.93*	*20.96*
Random Walk	1.09	1.11	1.22	1.04	1.00	1.05	1.04	1.10	1.00	0.98
F	0.74	0.78	0.67	0.69	0.92	0.97	0.87	0.76	0.85	1.10
F, PC-LEV	0.68	0.71	0.62	0.64	0.91	0.93	0.87	0.74	0.83	1.17
F, MED-LEV	0.66	0.69	0.60	0.62	0.89	0.90	0.85	0.72	0.80	1.19
F, MED-FoF	0.62	0.70	0.61	0.63	0.80	0.84	0.90	0.83	0.87	1.00
F, MED-MV	0.68	0.71	0.63	0.63	0.90	0.96	0.87	0.76	0.85	1.17
PC-LEV	0.89	0.93	0.88	0.93	0.99	0.96	1.03	0.91	1.00	1.12
MED-LEV	0.85	0.89	0.83	0.88	0.98	0.90	0.94	0.82	0.93	1.14
CB	0.99	0.94	0.96	0.98	0.97	1.04	0.98	1.00	1.03	1.07
INS	0.94	0.95	0.95	0.93	0.90	0.95	0.97	0.97	0.95	0.94
SBD	0.96	0.96	0.94	0.91	0.90	0.99	0.94	0.95	0.89	0.93
HH	0.95	0.98	0.89	0.97	1.00	1.00	1.00	0.92	1.00	1.18
BUS	0.99	1.00	0.99	1.00	1.00	1.19	1.22	1.17	1.30	1.28
BNK-MV	0.88	0.95	0.95	0.93	0.99	0.95	1.08	1.02	1.04	1.10
INS-MV	0.96	0.99	0.99	0.98	1.00	0.97	1.03	1.01	1.00	1.04
FIN-MV	0.86	0.79	0.77	0.81	0.97	1.02	0.93	0.93	0.98	1.06
SPF	NA	NA	NA	NA	NA	1.04	1.03	0.96	0.93	NA
BMA no LEV	0.62	0.63	0.24	0.59	0.84	1.08	1.10	0.98	1.01	1.20
BMA MED-LEV	0.57	0.53	0.24	0.49	0.84	1.03	0.98	0.92	1.02	1.21
BMA 8-LEV	0.60	0.61	0.24	0.55	0.75	1.09	1.20	1.11	1.10	1.26

Note: The first row shows absolute root mean squared forecast errors (RMSEs) of the 'Just ΔY' forecast model. Rows 2-21 show relative RMSEs, with respect to the 'Just ΔY' model. The model variants are listed in the first column (see main text for model descriptions). The row ('SPF') pertains to median SPF forecasts (not available for equity returns). The last three rows ('BMA...') represent RMSE with respect to forecasts (expected values of the predictive densities) from Bayesian Model Averaging. 'In-sample RMSEs' are based on regression (1) estimated for the sample 1993q3-2010q3 (for each dependent variable). 'Out-of-sample RMSEs' are based on (pseudo) out-of-sample forecasts one year ahead, from 40-quarter rolling estimation windows (forecast evaluation period: 1993q3-2010q3). Columns labeled 'GDP',...,'Rx' show RMSE's for the different forecasted variables (IP: industrial production; UE: unemployment rate; I: investment; Rx: excess equity return).

Table 3.2: Regression coefficients of leverage: whole sample and rolling windows

	Whole sample					*% Rolling windows with significant negative coefficients*				
Forecast model	**GDP**	**IP**	**UE**	**I**	**Rx**	**GDP**	**IP**	**UE**	**I**	**Rx**
F, PC-LEV	-0.42 ***	-0.49 ***	0.37 **	-0.38 **	-0.24	0.46	0.68	0.14	0.52	0.19
	{0.62}	{0.60}	{0.74}	{0.61}	{0.18}	{0.46}	{0.71}	{0.54}	{0.52}	{0.30}
F, MED-LEV	-0.45	-0.51	0.40	-0.45	-0.39	0.74	0.70	0.07	0.77	0.45
	{0.63}	{0.62}	{0.76}	{0.65}	{0.21}	{0.74}	{0.70}	{0.52}	{0.77}	{0.58}
F, MED-FoF	-0.45	-0.39	0.31	-0.33	-0.65	0.86	0.72	0.00	0.78	0.71
	{0.67}	{0.60}	{0.75}	{0.63}	{0.36}	{0.86}	{0.72}	{0.49}	{0.78}	{0.81}
F, MED-MV	-0.36	-0.42	0.29	-0.39	-0.24	0.36	0.29	0.33	0.29	0.09
	{0.61}	{0.60}	{0.74}	{0.63}	{0.19}	{0.38}	{0.39}	{0.46}	{0.29}	{0.12}
PC-LEV	-0.53	-0.47	0.60	-0.44	-0.15	0.75	1.00	0.00	0.81	0.30
	{0.34}	{0.31}	{0.48}	{0.19}	{0.02}	{0.75}	{1.00}	{0.52}	{0.81}	{0.42}
MED-LEV	-0.61	-0.56	0.63	-0.57	-0.25	0.87	0.97	0.00	1.00	0.43
	{0.40}	{0.37}	{0.54}	{0.28}	{0.05}	{0.87}	{0.97}	{0.61}	{1.00}	{0.54}
CB	-0.13	-0.30	0.22	-0.20	-0.25	0.32	0.61	0.04	0.59	0.46
	{0.18}	{0.29}	{0.38}	{0.11}	{0.06}	{0.32}	{0.71}	{0.88}	{0.68}	{0.59}
INS	-0.32	-0.30	0.26	-0.38	-0.46	0.78	0.83	0.00	0.62	0.46
	{0.26}	{0.27}	{0.39}	{0.20}	{0.19}	{0.78}	{0.83}	{0.45}	{0.62}	{0.57}
SBD	-0.25	-0.26	0.30	-0.40	-0.44	0.29	0.07	0.00	0.25	0.65
	{0.23}	{0.26}	{0.40}	{0.22}	{0.19}	{0.29}	{0.09}	{0.26}	{0.26}	{0.65}
HH	-0.43	-0.29	0.62	-0.35	0.05	0.81	0.48	0.00	0.70	0.00
	{0.25}	{0.23}	{0.47}	{0.13}	{0.01}	{0.81}	{0.48}	{0.80}	{0.70}	{0.36}
BUS	-0.15	-0.02	0.12	0.04	-0.10	0.46	0.29	0.03	0.17	0.55
	{0.18}	{0.20}	{0.33}	{0.07}	{0.01}	{0.46}	{0.29}	{0.35}	{0.17}	{0.59}
BNK-MV	-0.48	-0.33	0.31	-0.41	-0.14	0.72	0.41	0.00	0.75	0.03
	{0.35}	{0.27}	{0.40}	{0.20}	{0.02}	{0.72}	{0.41}	{0.13}	{0.75}	{0.03}
INS-MV	-0.28	-0.16	0.14	-0.18	-0.04	0.54	0.26	0.00	0.54	0.00
	{0.23}	{0.22}	{0.34}	{0.10}	{0.01}	{0.54}	{0.26}	{0.00}	{0.54}	{0.13}
FIN-MV	-0.57	-0.69	0.62	-0.67	-0.30	0.01	0.86	0.00	0.48	0.01
	{0.39}	{0.50}	{0.60}	{0.39}	{0.07}	{0.01}	{0.86}	{0.25}	{0.48}	{0.16}

Note: The *Left panel* (labeled '*Whole sample*') shows standardized slope coefficients of leverage, from regressions of each dependent variable on lagged leverage and other predictors for the period 1993q3-2010q3 (for each dependent variable). Asterisks indicate significance levels (based on Newey-West HAC t-statistic): * 10%, ** 5%, *** 1%. Numbers in brackets are R^2 coefficients of corresponding regression equations.
The *Right panel* (labeled '*% Rolling windows with significant negative coefficients*') shows shares of leverage coefficients that are significantly smaller than zero at a 10% level (two-sided Newey-West HAC t-test), among the rolling 40-quarter estimation windows; numbers in brackets pertain to the share of estimation windows with significant leverage coefficients at 10% level (i.e. sum of shares for significant negative and positive coefficients).
Columns labeled 'GDP', ...,'Rx' pertain to the different forecasted variables (IP: industrial production; UE: unemployment rate; I: investment; Rx: excess equity return).

Table 3.3: P-values of Clark-West (2007) test of equal predictive accuracy, relative to benchmark 'Just ΔY' model

Forecast model	**GDP**	**IP**	**UE**	**I**	**Rx**
Random Walk	0.61	0.35	0.35	0.33	0.03
F	0.06	0.04	0.02	0.02	0.59
F, PC-LEV	0.05	0.04	0.02	0.02	0.52
F, MED-LEV	0.04	0.03	0.02	0.01	0.59
F, MED-FoF	0.03	0.03	0.02	0.01	0.06
F, MED-MV	0.06	0.06	0.02	0.02	0.67
PC-LEV	0.04	0.31	0.04	0.21	0.62
MED-LEV	0.01	0.04	0.04	0.02	0.61
CB	0.76	0.05	0.18	0.36	0.82
INS	0.00	0.00	0.11	0.01	0.01
SBD	0.07	0.15	0.13	0.12	0.05
HH	0.01	0.03	0.01	0.02	0.97
BUS	0.45	0.69	0.64	0.89	0.64
BNK-MV	0.08	0.51	0.55	0.42	0.79
INS-MV	0.05	0.63	0.65	0.13	0.88
FIN-MV	0.61	0.00	0.11	0.08	0.94

Note: For each model listed in the first column (see main text), and for each of the forecasted variables, the Table reports the p-value of a test of the null hypothesis that that model has the same predictive accuracy (RMSE) as the benchmark 'Just ΔY' model. (The benchmark model nests random walk and is nested in each of the remaining models.) The MSPE-adjusted test statistic of Clark and West (2007) is used. Columns labeled 'GDP', ...,'Rx' show p-values for the different forecasted variables (IP: industrial production; UE: unemployment rate; I: investment; Rx: excess equity return). Out-of-sample forecasts (based on 40-quarter rolling estimation window) are used; the forecast evaluation period is 1993q3-2010q3.

Table 3.4: RMSEs of models that include 8-quarter differences of leverage as predictors

	In-sample RMSEs					***Out-of-sample RMSEs***				
Forecast model:	**GDP**	**IP**	**UE**	**I**	**Rx**	**GDP**	**IP**	**UE**	**I**	**Rx**
Just ΔY	*1.8*	*3.99*	*0.88*	*10.32*	*20.17*	*1.95*	*4.39*	*0.97*	*11.16*	*21.53*
F	0.74	0.78	0.65	0.66	0.92	0.97	0.87	0.74	0.84	1.10
F, PC-LEV	0.69	0.71	0.61	0.65	0.90	0.94	0.89	0.76	0.86	1.16
F, MED-LEV	0.67	0.71	0.60	0.63	0.89	0.89	0.86	0.73	0.82	1.13
PC-LEV	0.96	0.98	0.96	0.99	1.00	1.01	1.06	1.02	1.07	1.16
MED-LEV	0.93	0.98	0.94	0.98	0.99	0.95	1.01	0.97	1.02	1.20

Note: The RMSEs reported in this Table pertain to forecast models similar to (1), but with leverage indicators (Λ_t) based on 8-quarter differences of sectoral leverage series. The first row shows absolute RMSEs of the 'Just ΔY' forecast model (RMSEs are different from those shown Table 3.1, as the forecast evaluation period here was shortened to 1994q3-2010q3 for data availability reasons). The remaining rows show relative RMSEs, with respect to the 'Just ΔY' model. The model variants are listed in the first column (see main text for model descriptions).
'In-sample RMSEs' are based on regressions estimated for the sample 1994q3-2010q3 (for each dependent variable). 'Out-of-sample RMSEs' are based on (pseudo) out-of-sample forecasts one year ahead, from 40-quarter rolling estimation windows.
Columns labeled 'GDP', ...,'Rx' pertain to the different forecasted variables (IP: industrial production; UE: unemployment rate; I: investment; Rx: excess equity return).

Table 3.5: Regression coefficients of leverage indicators (based on 8-quarter differences)

	Whole sample					***% Rolling windows with significant negative coefficients***				
Forecast model	**GDP**	**IP**	**UE**	**I**	**Rx**	**GDP**	**IP**	**UE**	**I**	**Rx**
F, PC-LEV	-0.43 **	-0.52 ***	0.35 **	-0.24	-0.29	0.49	0.71	0.00	0.40	0.37
	{*0.61*}	{*0.60*}	{*0.75*}	{*0.60*}	{*0.19*}	{*0.71*}	{*0.71*}	{*0.46*}	{*0.40*}	{*0.40*}
F, MED-LEV	-0.52	-0.53	0.38	-0.37	-0.40	0.66	0.85	0.00	0.60	0.58
	{*0.64*}	{*0.60*}	{*0.75*}	{*0.63*}	{*0.21*}	{*0.66*}	{*0.85*}	{*0.65*}	{*0.60*}	{*0.71*}
PC-LEV	-0.34	-0.23	0.36	-0.14	-0.10	0.62	1.00	0.00	0.68	0.18
	{*0.24*}	{*0.22*}	{*0.37*}	{*0.08*}	{*0.01*}	{*0.62*}	{*1.00*}	{*0.88*}	{*0.68*}	{*0.18*}
MED-LEV	-0.44	-0.28	0.45	-0.24	-0.13	0.83	1.00	0.00	0.89	0.42
	{*0.29*}	{*0.23*}	{*0.40*}	{*0.10*}	{*0.02*}	{*0.83*}	{*1.00*}	{*1.00*}	{*0.89*}	{*0.60*}

Note: The *Left panel* (labeled '*Whole sample*') shows standardized slope coefficients of leverage indicators, based on 8-quarter differences of sectoral leverage series, for the models considered in Table 4. The coefficients are from regressions of each dependent variable on these leverage indicators and other predictors for the period 1994q3-2010q3. Asterisks indicate significance levels (based on Newey-West HAC t-statistics): * 10%, ** 5%, *** 1%. Numbers in brackets are R^2 coefficients of corresponding regression equations.
The *Right panel* (labeled '*% Rolling windows with significant negative coefficients*') shows shares of leverage coefficients that are significantly smaller than zero at a 10% level (two-sided Newey-West HAC t-test), among the rolling 40-quarter estimation windows; numbers in brackets pertain to the share of estimation windows with significant leverage coefficients at 10% level (i.e. sum of shares for significant negative and positive coefficients). The left column lists model variants, as described in the main text.
Columns labeled 'GDP', ...,'Rx' pertain to the different forecasted variables (IP: industrial production; UE: unemployment rate; I: investment; Rx: excess equity return).

Table 3.6: RMSEs of models that include non-linear transformations of leverage as predictors

	In-sample RMSEs					*Out-of-sample RMSEs*				
Forecast model:	**GDP**	**IP**	**UE**	**I**	**Rx**	**GDP**	**IP**	**UE**	**I**	**Rx**
Just ΔY	*1.77*	*3.90*	*0.87*	*10.16*	*19.63*	*1.91*	*4.27*	*0.95*	*10.93*	*20.96*
F	0.74	0.78	0.67	0.69	0.92	0.97	0.87	0.76	0.85	1.10
F, PC-LEV asym	0.68	0.71	0.59	0.64	0.90	1.02	0.93	0.78	0.93	1.25
F, MED-LEV asym	0.66	0.68	0.57	0.62	0.88	0.95	0.88	0.75	0.86	1.26
PC-LEV asym	0.89	0.93	0.83	0.93	0.98	1.13	1.18	1.08	1.21	1.19
MED-LEV asym	0.85	0.89	0.77	0.88	0.96	0.98	1.02	0.86	1.02	1.20
F, PC-LEV sq	0.68	0.71	0.60	0.64	0.88	1.16	1.04	0.79	1.03	1.40
F, MED-LEV sq	0.66	0.68	0.57	0.62	0.87	1.01	0.98	0.74	0.89	1.49
PC-LEV sq	0.89	0.92	0.85	0.93	0.95	1.35	1.33	1.20	1.35	1.31
MED-LEV sq	0.85	0.89	0.79	0.88	0.93	1.08	1.14	0.91	1.08	1.40

Note: This Table shows RMSEs for forecast models that include non-linear transformations of leverage indicators as predictors; see equation (2). The suffix 'asym' denotes inclusion of the asymmetric term $\max(0, \Lambda_t)$. The suffix 'sq' denotes inclusion of the squared term Λ_t^2. The following predictors are also included: the past change in real activity ($Y_t - Y_{t-1}$), the macro-financial factors (Φ_t) and Λ_t.
The first row shows absolute RMSEs of the 'Just ΔY' forecast model. The remaining rows show relative RMSEs, with respect to the 'ΔY' model. The model variants are listed in the first column (see main text for descriptions).
'In-sample RMSEs' are based on regressions estimated for the sample 1993q3-2010q3 (for each dependent variable). 'Out-of-sample RMSEs' are based on (pseudo) out-of-sample forecasts one year ahead, from 40-quarter rolling estimation windows (forecast evaluation period: 1993q3-2010q3).
Columns labeled 'GDP', ...,'Rx' pertain to the different forecasted variables (IP: industrial production; UE: unemployment rate; I: investment; Rx: excess equity return).

Table 3.7: Regression coefficients of leverage in forecast models that also include non-linear transformations of leverage as predictors

	Whole sample					***% Rolling windows with significant negative coefficients***				
Forecast model	**GDP**	**IP**	**UE**	**I**	**Rx**	**GDP**	**IP**	**UE**	**I**	**Rx**
F, PC-LEV asym	-0.38 **	-0.39 *	0.11	-0.40 **	-0.48	0.28	0.10	0.03	0.23	0.00
	{0.62}	*{0.60}*	*{0.76}*	*{0.61}*	*{0.20}*	*{0.28}*	*{0.12}*	*{0.04}*	*{0.23}*	*{0.07}*
F, MED-LEV asym	-0.36 **	-0.29	0.10	-0.42 **	-0.60 *	0.23	0.10	0.12	0.33	0.09
	{0.63}	*{0.63}*	*{0.78}*	*{0.65}*	*{0.22}*	*{0.23}*	*{0.13}*	*{0.13}*	*{0.33}*	*{0.12}*
PC-LEV asym	-0.58 ***	-0.53 **	0.16	-0.48 **	-0.54 *	0.58	0.13	0.00	0.46	0.03
	{0.34}	*{0.31}*	*{0.53}*	*{0.19}*	*{0.05}*	*{0.58}*	*{0.13}*	*{0.09}*	*{0.46}*	*{0.04}*
MED-LEV asym	-0.59 ***	-0.44 **	0.15	-0.50 ***	-0.61 **	0.65	0.14	0.01	0.42	0.16
	{0.40}	*{0.37}*	*{0.60}*	*{0.28}*	*{0.08}*	*{0.65}*	*{0.14}*	*{0.09}*	*{0.42}*	*{0.17}*
F, PC-LEV sq	-0.43 ***	-0.49 ***	0.35 ***	-0.38 ***	-0.27	0.54	0.71	0.14	0.62	0.22
	{0.62}	*{0.60}*	*{0.76}*	*{0.61}*	*{0.22}*	*{0.54}*	*{0.77}*	*{0.57}*	*{0.62}*	*{0.33}*
F, MED-LEV sq	-0.44 ***	-0.50 ***	0.38 ***	-0.45 ***	-0.39	0.77	0.74	0.12	0.88	0.48
	{0.63}	*{0.63}*	*{0.78}*	*{0.65}*	*{0.25}*	*{0.77}*	*{0.74}*	*{0.61}*	*{0.88}*	*{0.57}*
PC-LEV sq	-0.57 ***	-0.50 **	0.55 **	-0.46 *	-0.35	0.80	0.93	0.00	0.91	0.33
	{0.35}	*{0.32}*	*{0.52}*	*{0.19}*	*{0.10}*	*{0.80}*	*{0.93}*	*{0.55}*	*{0.91}*	*{0.43}*
MED-LEV sq	-0.63 ***	-0.56 ***	0.59 ***	-0.57 ***	-0.39 *	0.86	0.96	0.00	1.00	0.46
	{0.40}	*{0.37}*	*{0.57}*	*{0.28}*	*{0.13}*	*{0.86}*	*{0.96}*	*{0.64}*	*{1.00}*	*{0.58}*

Note: The *Left panel* (labeled '*Whole sample*') shows coefficients (β_3) of leverage in forecasting equation (2) (note: the equation also includes a non-linear function of leverage, as a regressor); estimation period: 1993q3-2010q3.
Asterisks indicate significance levels (based on Newey-West HAC t-statistics): * 10%, ** 5%, *** 1%. Numbers in brackets are R^2 coefficients of corresponding regression equations.
The *Right panel* (labeled '*% Rolling windows with significant negative coefficients*') shows shares of leverage coefficients that are significantly smaller than zero at a 10% level (two-sided Newey-West HAC t-test), among the rolling 40-quarter estimation windows; numbers in brackets pertain to the share of estimation windows with significant leverage coefficients at 10% level (i.e. sum of shares for significant negative and positive coefficients). The left column lists model variants, as described in the main text.
The left column lists model variants (see main text for descriptions). The suffix 'asym' denotes inclusion of the term $\max(\Lambda_t, 0)$; the suffix 'sq' denotes inclusion of the leverage growth indicator squared, $(\Lambda_t)^2$.
Columns labeled 'GDP', ...,'Rx' pertain to the different forecasted variables (IP: industrial production; UE: unemployment rate; I: investment; Rx: excess equity return).

Table 3.8: Regression coefficients of non-linear transformations of leverage

	Whole sample					***% Rolling windows with significant negative coefficients***				
Forecast model	**GDP**	**IP**	**UE**	**I**	**Rx**	**GDP**	**IP**	**UE**	**I**	**Rx**
F, PC-LEV asym	-0.06	-0.15	0.39	0.03	0.36	0.07	0.20	0.01	0.09	0.09
	{0.62}	*{0.60}*	*{0.76}*	*{0.61}*	*{0.20}*	*{0.32}*	*{0.22}*	*{0.17}*	*{0.33}*	*{0.13}*
F, MED-LEV asym	-0.13	-0.33	0.44 *	-0.05	0.33	0.06	0.36	0.00	0.07	0.33
	{0.63}	*{0.63}*	*{0.78}*	*{0.65}*	*{0.22}*	*{0.23}*	*{0.39}*	*{0.22}*	*{0.30}*	*{0.54}*
PC-LEV asym	0.06	0.08	0.62 *	0.04	0.42	0.09	0.19	0.04	0.13	0.14
	{0.34}	*{0.31}*	*{0.53}*	*{0.19}*	*{0.05}*	*{0.33}*	*{0.19}*	*{0.19}*	*{0.33}*	*{0.14}*
MED-LEV asym	-0.02	-0.15	0.65 ***	-0.09	0.42	0.09	0.28	0.00	0.14	0.36
	{0.40}	*{0.37}*	*{0.60}*	*{0.28}*	*{0.08}*	*{0.17}*	*{0.28}*	*{0.39}*	*{0.26}*	*{0.48}*
F, PC-LEV sq	0.02	-0.04	0.19	0.01	0.35	0.10	0.23	0.06	0.09	0.10
	{0.62}	*{0.60}*	*{0.76}*	*{0.61}*	*{0.22}*	*{0.35}*	*{0.28}*	*{0.26}*	*{0.35}*	*{0.22}*
F, MED-LEV sq	-0.02	-0.13	0.22 *	-0.04	0.30	0.07	0.41	0.06	0.07	0.43
	{0.63}	*{0.63}*	*{0.78}*	*{0.65}*	*{0.25}*	*{0.25}*	*{0.48}*	*{0.33}*	*{0.30}*	*{0.64}*
PC-LEV sq	0.11	0.14	0.26 **	0.06	0.35 *	0.14	0.19	0.09	0.13	0.17
	{0.35}	*{0.32}*	*{0.52}*	*{0.19}*	*{0.10}*	*{0.26}*	*{0.19}*	*{0.29}*	*{0.23}*	*{0.19}*
MED-LEV sq	0.09	0.04	0.25 **	0.02	0.34 *	0.14	0.29	0.04	0.16	0.41
	{0.40}	*{0.37}*	*{0.57}*	*{0.28}*	*{0.13}*	*{0.25}*	*{0.29}*	*{0.28}*	*{0.26}*	*{0.54}*

Note: The *Left panel* (labeled '*Whole sample*') shows standardized slope coefficients (β_4) for the non-linear function of leverage, in forecasting equation (2); estimation period: 1993q3-2010q3. Asterisks indicate significance levels (based on Newey-West HAC t-statistics): * 10%, ** 5%, *** 1%. Numbers in brackets are R^2 coefficients of corresponding regression equations.
The *Right panel* (labeled '*% Rolling windows with significant negative coefficients*') shows the share of coefficients for the non-linear functions of leverage that are significantly smaller than zero at a 10% level (two-sided Newey-West HAC t-test), among the rolling 40-quarter estimation windows. Numbers in brackets pertain to the share of estimation windows with significant non-linear terms at 10% level (i.e. sum of shares for significantly negative and positive coefficients).
The left column lists model variants (see main text for descriptions). The suffix 'asym' denotes inclusion of the term $\max(\Lambda_t, 0)$; the suffix 'sq' denotes inclusion of the leverage growth indicator squared, $(\Lambda_t)^2$.
Columns labeled 'GDP', ...,'Rx' pertain to the different forecasted variables (IP: industrial production; UE: unemployment rate; I: investment; Rx: excess equity return).

Table 3.9: Regressions of absolute out-of-sample forecast errors on leverage and macro-financial factors

Forecast model	***GDP***	***IP***	***UE***	***I***	***Rx***
CB	0.22 {.01; .05}	0.38 {.04; .14}	0.24 {.01; .06}	0.25 {.05; .06}	0.42 {.00; .18}
INS	0.26 {.08; .07}	0.21 {.12; .05}	0.25 {.06; .06}	0.36 {.01; .13}	0.29 {.04; .08}
SBD	-0.25 {.19; .06}	0.02 {.87; .00}	0.01 {.91; .00}	-0.07 {.64; .01}	-0.23 {.09; .05}
HH	0.52 {.00; .27}	0.34 {.01; .12}	0.46 {.01; .22}	0.35 {.01; .12}	0.53 {.00; .28}
BUS	0.45 {.00; .20}	0.34 {.02; .12}	0.48 {.00; .23}	0.42 {.00; .18}	0.36 {.00; .13}
BNK-MV	0.39 {.00; .15}	0.29 {.10; .08}	0.40 {.08; .16}	0.26 {.05; .07}	0.42 {.00; .18}
INS-MV	0.44 {.00; .20}	0.28 {.04; .08}	0.24 {.23; .06}	0.28 {.04; .08}	0.48 {.00; .23}
FIN-MV	0.34 {.00; .11}	0.39 {.00; .15}	0.57 {.00; .32}	0.38 {.00; .14}	0.38 {.00; .14}
8 Leverages jointly	{.00; .37}	{.00; .33}	{.00; .55}	{.00; .31}	{.00; .50}
PC-LEV	0.57 {.00; .32}	0.42 {.00; .18}	0.52 {.01; .27}	0.45 {.00; .20}	0.61 {.00; .37}
MED-LEV	0.47 {.00; .22}	0.41 {.00; .17}	0.51 {.01; .26}	0.43 {.00; .18}	0.55 {.00; .31}
F, PC-LEV	{.00; .44}	{.01; .25}	{.01; .33}	{.00; .32}	{.00; .49}
F, MED-LEV	{.00; .37}	{.01; .25}	{.00; .32}	{.00; .29}	{.00; .49}
F	{.00; .19}	{.06; .13}	{.00; .15}	{.01; .15}	{.00; .20}

Note: This Table reports *standardized slope coefficients* of leverage, *p-values* (1st figure in parentheses) and R^2 *coefficients* (2nd figure in parentheses) of regressions of absolute forecast errors for GDP, industrial production (IP), the unemployment rate (UE), gross investment (I) and the equity excess return (Rx), on the variables shown in the first column (a constant is included in all regressions). Columns labeled 'GDP', ...,'Rx' indicate the respective dependent variable. P-values are based on Newey-West HAC t-statistics.
Absolute forecast errors pertains to differences between realizations at $t + 4$ and forecasts made at t; forecasts are generated using the forecast regression referred to as 'F' in the text (based on rolling 40 quarter estimation window), i.e. the four macro-financial factors are used as predictors. Absolute forecast errors are regressed on changes of log leverage between $t - 4$ and t (observed at t). The sample period (t) is 1992q3-2009q3.
The first eight rows use each sectoral leverage variable (YoY changes) as an individual regressor. The row labeled '8 Leverages jointly' uses all 8 leverage series jointly as regressors (in parentheses: p-values of a joint significance test of all 8 leverage variables, based on a Wald test, with HAC covariance matrix). The row labeled 'PC-LEV' pertains to a regression on the first principal component of YoY changes of the 8 sectoral leverage series. The row labeled 'MED-LEV' uses the median of the standardized YoY change of the 8 sectoral log leverage series as a regressor. The next two rows add the four principal macro-financial factors as regressors. The last row (labeled 'F') regresses absolute forecast errors on just the four macro-financial factors. All regressions include a constant.

Table 3.10: Regressions of cross-sectional *dispersion* of SPF forecasts for GDP, IP, UE and I (4 quarters ahead), and of equity price volatility index (VIX), on leverage and macro-financial factors

	Forecast Dispersion				
Forecast model	***GDP***	***IP***	***UE***	***I***	***VIX***
CB	0.01 {.92; .00}	0.05 {.70; .00}	0.01 {.92; .00}	-0.02 {.87; .00}	0.38 {.01; .15}
INS	0.26 {.04; .06}	-0.09 {.55; .01}	0.36 {.04; .13}	0.11 {.40; .01}	0.17 {.28; .03}
SBD	-0.45 {.01; .21}	-0.52 {.00; .27}	0.04 {.68; .00}	-0.39 {.09; .15}	-0.26 {.09; .07}
HH	0.54 {.00; .30}	0.48 {.00; .23}	0.16 {.30; .02}	0.51 {.00; .26}	0.59 {.00; .35}
BUS	0.62 {.00; .39}	0.42 {.09; .18}	0.11 {.48; .01}	0.67 {.00; .45}	0.42 {.02; .18}
BNK-MV	0.33 {.18; .11}	0.24 {.29; .06}	-0.04 {.80; .00}	0.56 {.00; .32}	0.33 {.08; .11}
INS-MV	0.43 {.02; .19}	0.40 {.03; .16}	0.02 {.91; .00}	0.52 {.00; .27}	0.49 {.01; .24}
FIN-MV	0.49 {.00; .24}	0.35 {.00; .12}	0.30 {.00; .09}	0.34 {.03; .12}	0.55 {.00; .30}
8 Leverages jointly	{.00; .55}	{.00; .50}	{.00; .22}	{.00; .59}	{.00; .57}
PC-LEV	0.54 {.00; .29}	0.36 {.05; .13}	0.20 {.19; .04}	0.54 {.00; .30}	0.59 {.00; .35}
MED-LEV	0.45 {.00; .20}	0.30 {.06; .10}	0.16 {.28; .02}	0.49 {.00; .24}	0.55 {.00; .30}
F, PC-LEV	{.00; .34}	{.01; .21}	{.01; .10}	{.00; .36}	{.00; .36}
F, MED-LEV	{.00; .28}	{.01; .18}	{.00; .08}	{.00; .31}	{.00; .32}
F	{.00; .14}	{.00; .13}	{.11: .04}	{.04; .15}	{.08; .06}

Note: This Table reports *standardized slope coefficients* of leverage, *p-values* (1st figure in parentheses) and R^2 *coefficients* (2nd figure in parentheses) of: (i) regressions of the cross-sectional dispersion of 4-quarter-ahead SPF forecasts made at date t for GDP, industrial production (IP), the unemployment rate (UE) and private investment (I), on the change of log leverage between $t-4$ and t; (ii) regressions of the logged CBOE equity price volatility index (VIX) at the end of period t on leverage growth between $t-4$ and t. P-values are based on Newey-West HAC t-statistics. The sample period (t) is 1992q3-2009q3.
The first eight rows use each sectoral leverage variable (YoY changes) as an individual regressor. The row labeled '8 Leverages jointly' pertains to regressions on all 8 individual leverage series jointly (in parentheses: p-values of a joint significance test of all 8 leverage variables, and R^2). The row labeled 'PC-LEV' pertains to a regression of forecast dispersion/VIX on the first principal component of YoY changes of the 8 log leverage series. The row labeled 'MED-LEV' uses the median of the standardized YoY change of the log leverage series as a regressor. The next two rows add the four principal macro-financial factors as regressors. The last row (labeled 'F') regresses forecast dispersion/VIX on just the four macro-financial factors. All regressions include a constant.

Table 3.11: Coefficients and Posterior Inclusion Probabilities for in-sample BMA with individual leverage growth rates

	GDP	**IP**	**UE**	**I**	**Rx**
Gross domestic product	-0.16 {0.09}	-0.06 {0.05}	-0.11 {0.05}	-0.18 {0.07}	-0.05 {0.06}
Government expenditure	-0.14 {0.09}	-0.04 {0.03}	0.05 {0.04}	-0.13 {0.09}	-0.04 {0.06}
GDP implicit price deflator	-0.17 {0.09}	-0.11 {0.06}	-0.09 {0.06}	**-0.24 {0.31}**	**0.23 {0.18}**
Gross private domestic investment	-0.27 {0.10}	-0.04 {0.04}	-0.12 {0.05}	-0.29 {0.15}	0.11 {0.08}
Gross gov't saving (% of GDP)	0.07 {0.05}	0.05 {0.04}	0.10 {0.06}	0.06 {0.05}	0.15 {0.10}
Private housing starts	0.22 {0.09}	**0.22 {0.21}**	**-0.50 {0.98}**	**0.32 {0.31}**	0.06 {0.07}
Personal consumption expenditures	0.22 {0.16}	0.10 {0.05}	0.06 {0.04}	0.22 {0.16}	-0.21 {0.13}
Durable goods expenditures	0.15 {0.11}	0.13 {0.08}	-0.06 {0.04}	**0.20 {0.22}**	-0.02 {0.06}
Non-residential fixed investment	-0.06 {0.05}	-0.06 {0.04}	**0.35 {0.68}**	-0.15 {0.07}	-0.17 {0.10}
Residential fixed investment	**0.47 {0.90}**	**0.26 {0.30}**	-0.20 {0.06}	**0.45 {0.67}**	**0.20 {0.15}**
Net exports (% of GDP)	-0.20 {0.13}	-0.15 {0.11}	-0.05 {0.04}	-0.07 {0.05}	-0.04 {0.06}
Non-farm employment	**0.43 {0.25}**	0.31 {0.07}	**0.40 {0.26}**	0.04 {0.07}	**0.50 {0.59}**
Commodities producer price index	**-0.31 {0.49}**	**-0.31 {0.66}**	**0.36 {0.96}**	**-0.29 {0.35}**	**-0.33 {0.48}**
Civilian unemployment rate	0.36 {0.12}	0.32 {0.12}	**-0.47 {1.00}**	**0.40 {0.46}**	-0.13 {0.11}
Consumer price index	**-0.28 {0.39}**	-0.21 {0.15}	-0.04 {0.08}	-0.23 {0.21}	**-0.30 {0.33}**
Oil price (spot WTI) USD/barrel	0.14 {0.09}	0.16 {0.12}	-0.03 {0.04}	0.14 {0.09}	-0.08 {0.07}
Return on 3-month U.S. T-bill	-0.02 {0.07}	**-0.47 {0.31}**	**0.21 {0.20}**	**-0.32 {0.23}**	**-0.27 {0.23}**
Return on 2-year U.S. Treasury	-0.49 {0.16}	-0.18 {0.20}	**0.13 {0.10}**	-0.24 {0.18}	-0.25 {0.11}
Return on 5-year U.S. Treasury	0.43 {0.12}	-0.15 {0.12}	0.13 {0.10}	0.18 {0.07}	0.15 {0.08}
U.S. Treasury term spread 10Y-3M	-0.10 {0.07}	-0.08 {0.05}	-0.01 {0.04}	-0.17 {0.17}	-0.07 {0.07}
ISM manufacturing - inventories	-0.16 {0.16}	-0.04 {0.04}	0.12 {0.09}	-0.07 {0.05}	**-0.20 {0.21}**
ISM manufactoring - new orders	0.06 {0.04}	0.09 {0.05}	-0.11 {0.08}	0.09 {0.06}	0.02 {0.05}
Industrial production index	**0.81 {0.18}**	**1.09 {0.30}**	-0.05 {0.05}	0.65 {0.15}	0.31 {0.09}
Nominal M2 money stock	-0.03 {0.04}	-0.14 {0.08}	0.13 {0.09}	-0.22 {0.18}	-0.14 {0.10}
Total industry capacity utilization	-0.78 {0.14}	**-1.04 {0.26}**	0.12 {0.05}	-0.52 {0.12}	-0.32 {0.09}
French-Fama HML factor	**-0.21 {0.21}**	**-0.19 {0.20}**	0.09 {0.06}	-0.19 {0.17}	0.00 {0.06}
French-Fama Momentum factor	-0.12 {0.08}	-0.16 {0.10}	0.11 {0.08}	-0.15 {0.10}	-0.17 {0.15}
French-Fama SMB factor	-0.14 {0.08}	-0.11 {0.06}	-0.02 {0.03}	-0.10 {0.06}	0.00 {0.05}
French-Fama Short-term reversal	0.10 {0.06}	0.06 {0.04}	-0.11 {0.06}	0.12 {0.10}	-0.07 {0.07}
French-Fama Long-term reversal	0.00 {0.04}	0.02 {0.03}	-0.01 {0.03}	0.00 {0.04}	0.12 {0.10}
INS	-0.25 {0.04}	-0.22 {0.00}	0.09 {0.00}	-0.26 {0.02}	-0.36 {0.12}
SBD	**-0.33 {0.21}**	-0.27 {0.00}	**0.32 {0.85}**	-0.33 {0.10}	**-0.49 {0.84}**
CB	-0.22 {0.03}	-0.27 {0.00}	0.23 {0.09}	-0.19 {0.00}	-0.23 {0.01}
HH	-0.32 {0.02}	0.00 {0.00}	0.01 {0.00}	-0.26 {0.00}	0.23 {0.00}
BUS	-0.12 {0.00}	-0.22 {0.00}	-0.17 {0.00}	-0.02 {0.00}	-0.07 {0.00}
BNK-MV	-0.34 {0.17}	-0.33 {0.00}	0.08 {0.00}	-0.32 {0.01}	-0.13 {0.00}
INS-MV	-0.21 {0.01}	-0.10 {0.00}	-0.03 {0.00}	-0.14 {0.00}	-0.05 {0.00}
FIN-MV	**-0.38 {0.49}**	**-0.60 {1.00}**	0.23 {0.01}	**-0.49 {0.86}**	-0.19 {0.00}
Joint PIP of all Leverage	**{0.98}**	**{1.00}**	**{0.97}**	**{1.00}**	**{0.99}**

Note: Results from static 'BMA 8-LEV' in-sample forecasts for 5 dependent variables estimated on time period 1993q3-2010q3 (for explanatory variables: 1992q3-2009q3). All covariates have been properly stationarized (cf. data appendix). Model sampling for each of the 5 estimations was carried out with 400,000 burn-ins and 4 million subsequent iterations. The BMA priors are described in the technical appendix.

Numbers denote *posterior standardized conditional coefficients*, numbers in parentheses indicate *Posterior Inclusion Probabilities* (PIP). Note that the higher its PIP, the more posterior importance is attributed to that variable (absolute PIP is less relevant than the PIP ranking). For each dependent variable, the eight largest PIP are highlighted in bold. The last row indicates the posterior probability for inclusion of at least one leverage variable.

Table 3.12: Coefficients and Posterior Inclusion Probabilities for in-sample BMA with median leverage growth

	GDP	**IP**	**UE**	**I**	**Rx**
Gross domestic product	-0.20 {0.13}	-0.05 {0.06}	-0.08 {0.06}	-0.09 {0.07}	-0.04 {0.10}
Government expenditure	-0.13 {0.09}	-0.06 {0.06}	0.10 {0.08}	-0.12 {0.12}	-0.04 {0.10}
GDP implicit price deflator	-0.16 {0.10}	-0.09 {0.06}	0.05 {0.06}	**-0.27 {0.37}**	-0.03 {0.12}
Gross private domestic investment	-0.26 {0.11}	0.00 {0.05}	-0.06 {0.05}	-0.25 {0.16}	0.04 {0.10}
Gross gov't saving (% of GDP)	0.04 {0.05}	-0.05 {0.05}	0.10 {0.07}	0.01 {0.05}	0.08 {0.11}
Private housing starts	0.16 {0.08}	0.25 {0.24}	**-0.51 {0.96}**	0.28 {0.17}	0.09 {0.13}
Personal consumption expenditures	0.12 {0.10}	0.02 {0.06}	0.04 {0.05}	0.15 {0.12}	-0.10 {0.12}
Durable goods expenditures	0.15 {0.12}	0.12 {0.08}	-0.09 {0.06}	0.18 {0.21}	0.01 {0.09}
Non-residential fixed investment	-0.16 {0.08}	-0.12 {0.06}	**0.36 {0.28}**	-0.24 {0.12}	-0.07 {0.11}
Residential fixed investment	**0.44 {0.94}**	**0.36 {0.74}**	-0.26 {0.13}	**0.49 {0.89}**	0.16 {0.17}
Net exports (% of GDP)	-0.18 {0.13}	-0.07 {0.06}	-0.02 {0.05}	0.00 {0.05}	-0.05 {0.10}
Non-farm employment	0.20 {0.08}	0.24 {0.11}	**0.51 {0.79}**	0.02 {0.06}	0.07 {0.11}
Commodities producer price index	**-0.32 {0.58}**	**-0.42 {0.95}**	**0.36 {0.90}**	**-0.33 {0.52}**	**-0.21 {0.26}**
Civilian unemployment rate	**0.32 {0.24}**	**0.47 {0.71}**	**-0.44 {0.99}**	**0.56 {0.87}**	0.02 {0.10}
Consumer price index	**-0.29 {0.37}**	-0.09 {0.08}	0.12 {0.11}	-0.25 {0.20}	**-0.25 {0.38}**
Oil price (spot WTI) USD/barrel	**0.16 {0.18}**	**0.18 {0.27}**	-0.06 {0.06}	**0.19 {0.28}**	-0.02 {0.10}
Return on 3-month U.S. T-bill	-0.02 {0.05}	-0.07 {0.07}	0.21 {0.16}	-0.13 {0.09}	-0.18 {0.17}
Return on 2-year U.S. Treasury	-0.22 {0.12}	**-0.21 {0.28}**	**0.19 {0.29}**	**-0.25 {0.48}**	**-0.37 {0.36}**
Return on 5-year U.S. Treasury	0.08 {0.07}	**-0.25 {0.54}**	**0.18 {0.24}**	0.05 {0.10}	**0.29 {0.18}**
U.S. Treasury term spread 10Y-3M	**-0.14 {0.13}**	-0.06 {0.06}	0.06 {0.06}	-0.15 {0.14}	-0.12 {0.16}
ISM manufacturing - inventories	-0.14 {0.13}	-0.03 {0.05}	0.08 {0.07}	-0.06 {0.06}	-0.08 {0.11}
ISM manufactoring - new orders	0.06 {0.05}	0.15 {0.14}	-0.11 {0.10}	0.09 {0.07}	0.02 {0.09}
Industrial production index	0.42 {0.13}	0.34 {0.26}	-0.18 {0.07}	0.32 {0.22}	**0.62 {0.19}**
Nominal M2 money stock	-0.02 {0.04}	**-0.25 {0.65}**	0.14 {0.11}	**-0.25 {0.53}**	-0.11 {0.13}
Total industry capacity utilization	-0.34 {0.08}	0.04 {0.09}	0.14 {0.07}	0.24 {0.16}	**-0.54 {0.18}**
French-Fama HML factor	**-0.19 {0.24}**	-0.15 {0.20}	**0.15 {0.19}**	-0.14 {0.15}	-0.06 {0.11}
French-Fama Momentum factor	-0.08 {0.06}	-0.12 {0.08}	0.15 {0.12}	-0.09 {0.08}	**-0.15 {0.19}**
French-Fama SMB factor	-0.06 {0.06}	0.00 {0.05}	-0.02 {0.05}	0.00 {0.05}	0.01 {0.09}
French-Fama Short-term reversal	0.13 {0.13}	0.07 {0.06}	-0.11 {0.09}	0.15 {0.19}	0.02 {0.09}
French-Fama Long-term reversal	-0.03 {0.05}	-0.03 {0.05}	-0.01 {0.05}	-0.04 {0.05}	0.06 {0.11}
MED-LEV	**-0.45 {0.99}**	**-0.53 {0.99}**	0.19 {0.14}	**-0.51 {0.98}**	**-0.21 {0.24}**

Note: Results from static 'BMA MED-LEV' in-sample forecasts for 5 dependent variables estimated on time period 1993q3-2010q3 (for explanatory variables: 1992q3-2009q3). All covariates have been properly stationarized (cf. data appendix). Model sampling for each of the 5 estimations was carried out with 400,000 burn-ins and 4 million subsequent iterations. The BMA priors are described in the technical appendix.

Numbers denote *posterior standardized conditional coefficients*, numbers in parentheses indicate *Posterior Inclusion Probabilities* (PIP). For each dependent variable, the eight largest PIP are highlighted in bold.

Table 3.13: Mean coefficients and Posterior Inclusion Probabilities for out-of-sample BMA with eight sectoral leverages

	GDP	**IP**	**UE**	**I**	**Rx**
Gross domestic product	-0.19 {0.10}	-0.06 {0.07}	-0.13 {0.09}	-0.14 {0.07}	-0.01 {0.07}
Government expenditure	-0.10 {0.07}	-0.03 {0.05}	0.02 {0.05}	-0.12 {0.07}	0.11 {0.09}
GDP implicit price deflator	**-0.35 {0.22}**	**-0.38 {0.27}**	**0.28 {0.24}**	-0.22 {0.10}	-0.12 {0.10}
Gross private domestic investment	-0.08 {0.08}	-0.02 {0.05}	-0.10 {0.06}	-0.21 {0.11}	0.03 {0.07}
Gross gov't saving (% of GDP)	0.15 {0.10}	0.13 {0.08}	-0.16 {0.11}	0.12 {0.07}	0.04 {0.09}
Private housing starts	0.08 {0.07}	-0.04 {0.06}	0.00 {0.05}	0.12 {0.09}	0.05 {0.08}
Personal consumption expenditures	0.15 {0.10}	-0.03 {0.07}	-0.20 {0.11}	0.12 {0.11}	-0.20 {0.12}
Durable goods expenditures	0.15 {0.10}	0.06 {0.05}	-0.08 {0.06}	0.27 {0.17}	0.12 {0.09}
Non-residential fixed investment	0.05 {0.07}	0.04 {0.06}	-0.21 {0.10}	0.05 {0.08}	-0.10 {0.09}
Residential fixed investment	**0.41 {0.26}**	0.21 {0.17}	-0.45 {0.17}	**0.47 {0.23}**	0.13 {0.11}
Net exports (% of GDP)	-0.13 {0.10}	-0.13 {0.10}	0.04 {0.12}	0.14 {0.14}	-0.07 {0.10}
Non-farm employment	**0.92 {0.39}**	**0.34 {0.17}**	**-0.58 {0.55}**	**0.49 {0.21}**	**0.22 {0.17}**
Commodities producer price index	-0.08 {0.10}	**-0.29 {0.25}**	0.15 {0.10}	-0.11 {0.09}	0.00 {0.08}
Civilian unemployment rate	**0.36 {0.20}**	-0.06 {0.09}	**-0.23 {0.20}**	**0.43 {0.18}**	**0.05 {0.13}**
Consumer price index	-0.25 {0.15}	-0.20 {0.11}	0.19 {0.10}	-0.11 {0.07}	-0.11 {0.09}
Oil price (spot WTI) USD/barrel	0.04 {0.09}	-0.07 {0.06}	0.07 {0.07}	-0.05 {0.09}	0.00 {0.07}
Return on 3-month U.S. T-bill	**-0.64 {0.46}**	**-0.63 {0.49}**	**0.59 {0.70}**	**-0.70 {0.62}**	-0.11 {0.13}
Return on 2-year U.S. Treasury	**0.42 {0.27}**	0.12 {0.10}	0.49 {0.18}	0.13 {0.11}	**-0.22 {0.18}**
Return on 5-year U.S. Treasury	**0.33 {0.20}**	-0.19 {0.08}	-0.48 {0.18}	0.18 {0.12}	**0.40 {0.22}**
U.S. Treasury term spread 10Y-3M	0.05 {0.07}	0.13 {0.08}	-0.11 {0.09}	0.09 {0.07}	0.04 {0.09}
ISM manufacturing - inventories	-0.13 {0.10}	-0.02 {0.05}	0.04 {0.05}	-0.06 {0.07}	0.03 {0.07}
ISM manufactoring - new orders	0.01 {0.07}	0.08 {0.06}	0.09 {0.06}	0.11 {0.07}	-0.09 {0.09}
Industrial production index	0.64 {0.16}	**0.87 {0.38}**	**-0.41 {0.24}**	**0.40 {0.25}**	**0.77 {0.22}**
Nominal M2 money stock	0.13 {0.09}	0.21 {0.16}	-0.03 {0.11}	0.08 {0.07}	-0.22 {0.12}
Total industry capacity utilization	-0.59 {0.16}	**-0.95 {0.23}**	0.03 {0.14}	-0.31 {0.17}	**-0.74 {0.19}**
French-Fama HML factor	-0.18 {0.17}	0.11 {0.10}	**-0.16 {0.22}**	-0.10 {0.13}	-0.08 {0.09}
French-Fama Momentum factor	-0.07 {0.07}	-0.21 {0.16}	0.14 {0.09}	-0.12 {0.07}	-0.05 {0.08}
French-Fama SMB factor	-0.11 {0.08}	0.05 {0.06}	0.03 {0.04}	-0.03 {0.06}	0.00 {0.07}
French-Fama Short-term reversal	0.19 {0.14}	0.12 {0.08}	-0.15 {0.10}	0.14 {0.08}	-0.07 {0.08}
French-Fama Long-term reversal	-0.09 {0.11}	0.01 {0.08}	0.07 {0.08}	-0.05 {0.08}	0.06 {0.07}
INS	-0.39 {0.10}	-0.37 {0.08}	0.33 {0.08}	**-0.48 {0.19}**	**-0.53 {0.25}**
SBD	-0.33 {0.16}	-0.10 {0.01}	**0.34 {0.21}**	-0.20 {0.03}	-0.34 {0.07}
CB	-0.29 {0.15}	**-0.47 {0.44}**	**0.32 {0.23}**	**-0.46 {0.21}**	-0.44 {0.12}
HH	**-0.39 {0.18}**	**-0.41 {0.18}**	0.35 {0.06}	-0.33 {0.10}	**0.57 {0.37}**
BUS	-0.30 {0.05}	-0.42 {0.14}	0.34 {0.05}	-0.34 {0.07}	-0.25 {0.03}
BNK-MV	-0.50 {0.17}	-0.27 {0.01}	0.28 {0.05}	**-0.45 {0.22}**	-0.09 {0.02}
INS-MV	-0.11 {0.02}	0.10 {0.01}	-0.04 {0.02}	-0.15 {0.01}	0.10 {0.02}
FIN-MV	0.00 {0.02}	-0.29 {0.03}	-0.07 {0.02}	-0.03 {0.01}	0.06 {0.01}
Joint PIP of all Leverage	**{0.87}**	**{0.90}**	**{0.75}**	**{0.87}**	**{0.89}**

Note: Results from rolling 'BMA 8-LEV' 1-year-ahead out-of-sample forecasts for 5 dependent variables over time period 1993q3 to 2010q3. All covariates have been properly stationarized (cf. data appendix). Model sampling for each of the 69 × 5 forecasts was carried out with 400,000 burn-ins and 2 million iterations. The BMA priors are described in the technical appendix. Numbers denote the arithmetic mean (over 69 rolling samples) of posterior *standardized conditional coefficients.*

Numbers in parentheses indicate mean *Posterior Inclusion Probabilities* (PIP) over the 69 rolling samples. Note that the higher its PIP, the more posterior importance is attributed to that variable (absolute PIP is less relevant than the PIP ranking). For each dependent variable, the eight largest PIP are highlighted in bold. The last row indicates the posterior probability for inclusion of at least one leverage variable.

Table 3.14: Mean coefficients and Posterior Inclusion Probabilitites for out-of-sample BMA setting with median leverage growth

	GDP	**IP**	**UE**	**I**	**Rx**
Gross domestic product	-0.17 {0.10}	-0.17 {0.11}	-0.19 {0.14}	-0.13 {0.09}	0.02 {0.10}
Government expenditure	-0.10 {0.09}	-0.02 {0.07}	0.01 {0.05}	-0.14 {0.09}	0.11 {0.12}
GDP implicit price deflator	**-0.33 {0.22}**	**-0.47 {0.33}**	**0.27 {0.21}**	-0.20 {0.11}	-0.06 {0.11}
Gross private domestic investment	-0.08 {0.09}	-0.02 {0.07}	-0.06 {0.06}	-0.21 {0.12}	0.01 {0.10}
Gross gov't saving (% of GDP)	0.16 {0.12}	0.14 {0.11}	-0.16 {0.12}	0.11 {0.08}	0.04 {0.11}
Private housing starts	0.05 {0.08}	-0.03 {0.07}	0.04 {0.05}	0.10 {0.09}	0.01 {0.10}
Personal consumption expenditures	0.12 {0.11}	-0.11 {0.08}	-0.12 {0.08}	0.12 {0.11}	-0.17 {0.14}
Durable goods expenditures	0.19 {0.15}	0.14 {0.08}	-0.05 {0.05}	**0.21 {0.15}**	0.09 {0.11}
Non-residential fixed investment	0.07 {0.08}	-0.01 {0.08}	-0.19 {0.11}	0.06 {0.08}	0.04 {0.10}
Residential fixed investment	**0.42 {0.27}**	**0.39 {0.21}**	**-0.45 {0.19}**	**0.50 {0.26}**	0.01 {0.13}
Net exports (% of GDP)	-0.09 {0.10}	0.02 {0.11}	-0.05 {0.09}	0.21 {0.14}	0.04 {0.12}
Non-farm employment	**0.79 {0.36}**	**0.47 {0.22}**	**-0.64 {0.65}**	**0.35 {0.25}**	**0.42 {0.26}**
Commodities producer price index	0.01 {0.10}	**-0.22 {0.17}**	0.07 {0.10}	-0.01 {0.08}	0.03 {0.11}
Civilian unemployment rate	0.14 {0.15}	0.09 {0.15}	**-0.33 {0.34}**	**0.38 {0.25}**	0.04 {0.14}
Consumer price index	-0.26 {0.17}	-0.23 {0.14}	0.23 {0.12}	-0.07 {0.08}	-0.06 {0.10}
Oil price (spot WTI) USD/barrel	0.00 {0.08}	0.02 {0.07}	0.03 {0.05}	0.01 {0.06}	0.02 {0.09}
Return on 3-month U.S. T-bill	**-0.64 {0.41}**	**-0.53 {0.45}**	**0.63 {0.79}**	**-0.66 {0.72}**	**-0.37 {0.21}**
Return on 2-year U.S. Treasury	**0.52 {0.26}**	0.30 {0.15}	0.34 {0.11}	0.20 {0.11}	**-0.04 {0.20}**
Return on 5-year U.S. Treasury	0.21 {0.13}	-0.31 {0.14}	-0.30 {0.10}	0.06 {0.08}	**0.26 {0.30}**
U.S. Treasury term spread 10Y-3M	0.05 {0.08}	0.20 {0.14}	-0.03 {0.06}	0.11 {0.08}	0.03 {0.12}
ISM manufacturing - inventories	-0.06 {0.07}	0.03 {0.07}	0.04 {0.05}	-0.02 {0.06}	0.00 {0.09}
ISM manufactoring - new orders	0.04 {0.09}	0.14 {0.10}	0.11 {0.07}	0.13 {0.09}	-0.06 {0.11}
Industrial production index	0.27 {0.13}	**0.47 {0.18}**	**-0.27 {0.17}**	**0.29 {0.20}**	**0.24 {0.19}**
Nominal M2 money stock	0.18 {0.12}	0.13 {0.14}	-0.11 {0.13}	0.07 {0.08}	-0.13 {0.12}
Total industry capacity utilization	-0.20 {0.13}	-0.22 {0.13}	-0.16 {0.16}	**-0.16 {0.22}**	**-0.19 {0.17}**
French-Fama HML factor	**-0.22 {0.20}**	**-0.23 {0.23}**	**-0.09 {0.21}**	-0.11 {0.14}	-0.06 {0.11}
French-Fama Momentum factor	-0.07 {0.08}	-0.25 {0.16}	0.13 {0.09}	-0.09 {0.07}	-0.08 {0.12}
French-Fama SMB factor	-0.10 {0.10}	0.00 {0.10}	0.03 {0.05}	-0.02 {0.07}	-0.02 {0.10}
French-Fama Short-term reversal	**0.23 {0.18}**	0.13 {0.11}	-0.13 {0.09}	0.13 {0.09}	-0.08 {0.11}
French-Fama Long-term reversal	-0.04 {0.11}	0.08 {0.12}	0.04 {0.12}	0.01 {0.08}	0.00 {0.09}
MED-LEV	**-0.43 {0.45}**	**-0.45 {0.48}**	**0.35 {0.39}**	**-0.46 {0.62}**	**-0.03 {0.28}**

Note: Results from rolling 'BMA MED-LEV' 1-year-ahead out-of-sample forecasts for 5 dependent variables over time period 1993q3 to 2010q3. All covariates have been properly stationarized (cf. data appendix). Model sampling for each of the 69 $\times$ 5 forecasts was carried out with 400,000 burn-ins and 2 million iterations. The BMA priors are described in the technical appendix. Numbers denote the arithmetic mean (over 69 rolling samples) of *posterior standardized conditional coefficients.*

Numbers in parentheses indicate mean *Posterior Inclusion Probabilities* (PIP) over the 69 rolling samples. For each dependent variable, the eight largest PIP are highlighted in bold.

Table 3.15: BMA coefficients and posterior inclusion probabiltities for the conditional output variability (forecast errors from model 'F')

	GDP	**IP**	**UE**	**I**	**Rx**
Gross domestic product	-0.11 {0.14}	0.12 {0.15}	-0.27 {0.26}	0.03 {0.14}	**0.25 {0.26}**
Government expenditure	-0.05 {0.11}	-0.01 {0.11}	-0.02 {0.11}	0.02 {0.14}	0.01 {0.09}
GDP implicit price deflator	-0.25 {0.30}	**-0.25 {0.36}**	-0.17 {0.21}	-0.10 {0.17}	**-0.15 {0.20}**
Gross private domestic investment	-0.22 {0.21}	0.02 {0.12}	-0.03 {0.13}	-0.08 {0.16}	0.06 {0.11}
Gross gov't saving (% of GDP)	-0.19 {0.21}	-0.12 {0.15}	-0.04 {0.12}	-0.15 {0.23}	**-0.28 {0.62}**
Private housing starts	0.10 {0.13}	0.10 {0.15}	0.03 {0.12}	**0.21 {0.34}**	0.01 {0.09}
Personal consumption expenditures	-0.01 {0.17}	-0.08 {0.14}	**-0.27 {0.38}**	0.00 {0.13}	0.09 {0.11}
Durable goods expenditures	**0.26 {0.55}**	-0.15 {0.20}	-0.12 {0.19}	-0.04 {0.14}	0.00 {0.09}
Non-residential fixed investment	**-0.32 {0.36}**	0.16 {0.16}	**0.28 {0.30}**	0.07 {0.15}	**-0.21 {0.17}**
Residential fixed investment	-0.02 {0.12}	**-0.29 {0.45}**	-0.21 {0.26}	-0.07 {0.16}	-0.12 {0.12}
Net exports (% of GDP)	0.09 {0.13}	0.11 {0.16}	0.03 {0.11}	0.05 {0.14}	0.07 {0.11}
Non-farm employment	**-0.40 {0.35}**	-0.23 {0.17}	-0.16 {0.15}	-0.07 {0.16}	**-0.37 {0.31}**
Commodities producer price index	0.08 {0.13}	**0.41 {0.80}**	**0.29 {0.51}**	0.12 {0.18}	-0.01 {0.10}
Civilian unemployment rate	**0.36 {0.34}**	0.04 {0.12}	0.02 {0.12}	0.07 {0.16}	0.04 {0.11}
Consumer price index	0.27 {0.27}	0.17 {0.19}	0.13 {0.20}	0.05 {0.15}	0.05 {0.10}
Oil price (spot WTI) USD/barrel	**0.29 {0.81}**	0.07 {0.12}	-0.13 {0.18}	**0.13 {0.24}**	0.08 {0.12}
Return on 3-month U.S. T-bill	0.21 {0.28}	**0.24 {0.31}**	-0.01 {0.12}	**0.19 {0.29}**	**0.37 {0.89}**
Return on 2-year U.S. Treasury	**0.37 {0.39}**	0.04 {0.12}	**0.21 {0.35}**	**0.29 {0.30}**	-0.05 {0.12}
Return on 5-year U.S. Treasury	-0.35 {0.29}	-0.05 {0.11}	0.17 {0.26}	**-0.25 {0.24}**	-0.09 {0.12}
U.S. Treasury term spread 10Y-3M	0.11 {0.16}	-0.03 {0.10}	-0.07 {0.13}	0.00 {0.14}	0.03 {0.09}
ISM manufacturing - inventories	0.06 {0.11}	0.02 {0.10}	0.10 {0.14}	0.09 {0.18}	0.05 {0.10}
ISM manufactoring - new orders	0.10 {0.14}	-0.07 {0.13}	**-0.17 {0.29}**	-0.07 {0.16}	0.07 {0.11}
Industrial production index	-0.08 {0.12}	**-0.24 {0.25}**	-0.22 {0.20}	**-0.12 {0.24}**	-0.04 {0.10}
Nominal M2 money stock	-0.02 {0.09}	0.04 {0.10}	-0.15 {0.21}	-0.09 {0.17}	0.10 {0.12}
Total industry capacity utilization	-0.17 {0.15}	**-0.47 {0.73}**	-0.25 {0.26}	**-0.34 {0.46}**	-0.08 {0.11}
French-Fama HML factor	**0.20 {0.43}**	**0.24 {0.56}**	0.16 {0.25}	0.08 {0.17}	0.02 {0.09}
French-Fama Momentum factor	-0.09 {0.13}	0.12 {0.16}	-0.03 {0.13}	0.06 {0.16}	-0.04 {0.10}
French-Fama SMB factor	-0.10 {0.15}	-0.08 {0.14}	**-0.16 {0.27}**	-0.12 {0.21}	0.06 {0.10}
French-Fama Short-term reversal	-0.05 {0.11}	0.11 {0.18}	-0.08 {0.13}	-0.06 {0.15}	**0.15 {0.25}**
French-Fama Long-term reversal	0.04 {0.10}	-0.14 {0.23}	**-0.17 {0.29}**	0.07 {0.16}	0.11 {0.16}
MED-LEV	**0.31 {0.55}**	**0.23 {0.32}**	**0.37 {0.77}**	**0.32 {0.59}**	**0.49 {0.98}**

Note: Results from Bayesian Model Averaging over regressions of absolute forecast errors for GDP, industrial production (IP), the unemployment rate (UE), gross investment (I) and the equity excess return (Rx), on the variables shown in the first column (a constant is included in all regressions). Sample period is 1992q3-2009q3 (1993q3-2010q3 for the dependent variable). Numbers denote posterior *standardized conditional coefficients*, numbers in parentheses indicate *Posterior Inclusion Probabilities* (PIP). Note that the higher its PIP, the more posterior importance is attributed to that variable. For each dependent variable, the eight largest PIP are highlighted in bold.

4

Principal Components and Model Averaging

Co-authored with:
Martin Wagner, Institute for Advanced Studies, Vienna & Frisch Centre for Economic Research, Oslo

Abstract
Model averaging has proven popular for inference with many potential predictors in small samples. However, it is frequently criticized for a lack of robustness with respect to prediction and inference. This chapter explores the reasons for such robustness problems and proposes to address them by transforming the subset of potential 'control' predictors into principal components in suitable datasets. A simulation analysis shows that this approach yields robustness advantages vs. both standard model averaging and principal component-augmented regression. Moreover, we devise a prior framework that extends model averaging to uncertainty over the set of principal components and show that it offers considerable improvements with respect to the robustness of estimates and inference about the importance of covariates. Finally, we empirically benchmark our approach with popular model averaging and PC-based techniques in evaluating financial indicators as alternatives to established macroeconomic predictors of real economic activity.

Keywords: Bayesian model averaging, Frequentist model averaging, principal components augmented regression, financial crisis, business cycle forecasting

JEL Classification: C31, C52, O11, O18, O47

4.1 Introduction

When confronted with a large number of explanatory variables relative to sample size, two approaches are very popular in econometrics. These are, on the one hand, model averaging techniques and factor model approaches on the other.[1] Both approaches come with limitations. Model averaging, be it Bayesian or frequentist, has been criticized as being susceptible to over-fitting (Stock and Watson, 2006) and to lack robustness (e.g., Ciccone and Jarociński, 2010). These problems are of course manifestations of the underlying difficulties of estimation and inference in a setting with many, potentially collinear explanatory variables. Moreover, the computational cost of model averaging is in general large and in applications of typical scale the model space is too large to be fully considered in the averaging procedure. Thus, model averaging often resorts to MCMC sampling schemes, whose approximation difficulties under highly correlated covariates are frequently considered to introduce bias and robustness problems (e.g., Clyde et al., 2010). In contrast, for factor model approaches – typically based on principal components analysis – the computational simplicity and robust performance are considered as main advantages. Whilst factor models are often used for predictions (e.g., Bai and Ng, 2002; Forni et al., 2000), the use of principal components rather than the original variables complicates an assessment of the individual variables' importance in explaining the dependent variable.

In this paper we present an approach that combines the virtues of both model averaging and factor models. The approach is motivated and particularly useful for the following prototypical situation: In many economic applications, the *focus* is only on a relatively small set of (possibly competing) regressors, whereas a large set of other regressors is included to *control* for all unmodeled effects. In such a situation disentangling the effect of the focus variables is difficult when the total number of variables is large compared to the sample size (or even larger than the sample size). In principle, consistent model averaging estimators (cf. Fernández et al., 2001a) allow to asymptotically quantify the effects of the focus variables, but in finite samples two problems arise. First, model averaging over both focus and control variables takes into account models in which potentially important control variables are not included – leading to ensuing omitted variables biases in the coefficient estimators corresponding to the focus variables. Second, and in a sense the opposite to the first problem, Bayesian model averaging with loose priors may lead to large effective model sizes with the ensuing large estimator variances, as has been pointed out by Ciccone and Jarociński (2010). Although several contributions (e.g., Sala-i-Martin et al., 2004; Ley and Steel, 2009) address this feature through 'model prior' structures that allow to steer prior expected model parsimony, such priors at most allow the researcher to favor a particular side in the trade-off between parsimony on the one hand and the inclusion of as many potentially relevant variables as possible on the other.

The simplest approach to overcome the above mentioned problems of model averaging is – in the discussed setup with ex ante defined focus and control variables – to: (i) consider model averaging only over the focus variables and (ii) 'regularize' the regression problem by including principal components computed from the control variables throughout. By conditioning on a (small) number of principal components the effects of the control variables are effectively taken into account (see also Section 2.2). Clearly, the effective number of parameters is reduced by including principal components computed from the control variables rather than the control variables themselves, which will exhibit positive effects on the robustness of model averaging estimators. Moreover, this approach leads to a drastic reduction in the model space and thus in its computational cost. In the approach described so far the number of principal components to be included is yet to be

[1] Of course, there are also other 'regularization' approaches and model selection techniques like (e.g., adaptive LASSO estimators). Primarily for the sake of briefness, this paper will concentrate on the intersection between model averaging and factor models.

determined, with this choice effectively corresponding to a positioning on the trade-off between bias and variance of the estimators. The literature typically resorts to various ad-hoc schemes to determine the number of principal components to be included (see the discussion in Section 2.3). As is exemplified in the simulations in Section 3 these simple procedures still risk to set the number of principal components too low or too high with the ensuing implications for bias and variance.

In order to overcome this limitation with respect to the number of principal components we propose to instead apply model averaging over both the focus variables and the number of principal components to consider. We refer to this approach as principal components-model averaging (PC-MA). This combination of model averaging and factor modeling via principal components has three main advantages: First, conditioning each sub-model for the focus variables on the principal components computed from the control variables is a parsimonious way of reducing potential omitted variables biases for the coefficients corresponding to the focus variables.[1] Thus, PC-MA leads to parsimonious models without the danger of large omitted variables biases, which is a danger for model averaging procedures where averaging is performed over all variables, in particular when in a Bayesian framework the priors are set to small model sizes. Second, averaging also over the principal components leads to a data-driven choice concerning the number of principal components to be included. Here it is worth noting that our proposed prior is flexible enough to reflect a researcher's preferences concerning the inclusion of principal components in a continuous fashion. Third, PC-MA enormously reduces the model space. Denoting the number of focus variables with k_F and the number of control variables with k_C, in standard model averaging the total number of models is $2^{k_F+k_C}$, whereas PC-MA necessitates the evaluation of not more than $2^{k_F} \times (k_C+1)$ models. This means that one does not need to apply approximate MCMC sampling schemes and thereby evades robustness problems that stem from poor mixing of MCMC samplers under a high degree of correlation in the data (cf. Clyde et al., 1996, for a more detailed discussion).

To the best of our knowledge previous research on combining model averaging with principal components is limited to very few contributions. Koop and Potter (2004) perform Bayesian model averaging over principal components computed from all variables and employ their approach to a macroeconomic forecasting exercise along the lines of Stock and Watson (2002). Thus, in our terms, they perform a variant of PC-MA with a zero number of focus variables. There is another important difference, as Koop and Potter (2004) do not take into account the inherent hierarchical structure of the principal components (which further reduces the computational cost, as we will see in Section 4.3.2) and consider only a uniform prior. The approach of Magnus et al. (2010) employs a distinction between focus and control or auxiliary variables, but is in a sense 'dual' to our approach in that they compute principal components from the focus variables after conditioning throughout on (a large number of) control variables. Including a fixed number of control variables is of course very restrictive and furthermore by doing so the researcher risks to end up in an unfavorable point on the bias variance trade-off (especially if the number of control variables is large compared to the sample size). Wagner and Hlouskova (2010) consider model averaging in principal components augmented regressions with a fixed number of principal components included throughout and averaging only over the focus variables. Our paper can be seen as an extension of Koop and Potter (2004) and Wagner and Hlouskova (2010), since (i) we develop a framework to jointly average over both principal components and untransformed variables, and (ii) consider both Bayesian and frequentist model averaging. Finally, the papers by De Mol et al. (2008) and Stock and Watson (2011) are related in that they asses factor models versus Bayesian shrinkage estimators, but they do not directly consider model averaging.

[1]Given that the principal components are mutually orthogonal also implies that the regressions are well-behaved if multi-collinearity between the focus variables and the principal components is not an issue and, of course, if the focus variables are not multi-collinear. Both can easily be checked in any given application.

The paper is organized as follows: Section 4.2 presents the econometric theory and discusses model averaging, principal components augmented regressions and a combination thereof. Section 4.3.2 introduces our PC-MA approach. Section 4.4 presents the simulation design and results. Section 4.5 is devoted to an analysis of financial indicators as an alternative to established macroeconomic predictors of GDP growth by means of the PC-MA approach, and a comparison with alternative approaches. Finally, Section 4.6 briefly summarizes and concludes.

4.2 Description of the Econometric Approach

This section starts by reviewing model averaging and principal components augmented regressions in the first two subsections. It then proceeds to discuss our framework intended to combine the favorable aspects of both, where we consider two cases: (i) model averaging only over the focus variables conditional on a fixed set of principal components computed from the control variables, and (ii) model averaging over both the focus variables and the principal components (PC-MA).

4.2.1 Model Averaging

We consider linear regression models $\mathcal{M}_j$ in partitioned format:

$$y = X_{F,j}\beta_{F,j} + X_C\beta_{C,j} + u_j, \tag{1}$$

with $y \in \mathbb{R}^N$ denoting the dependent variable, $X_C \in \mathbb{R}^{N\times k_C}$ a fixed set of control variables and $u_j \in \mathbb{R}^N$ the error vector. Model uncertainty is present about which focus variables $X_{F,j} \in \mathbb{R}^{N\times k_{F,j}}$ from $X_F \in \mathbb{R}^{N\times k_F}$ should be included in the regression. Under a Bayesian approach, each of these selections, corresponding to a specific model $\mathcal{M}_j$, is assigned a prior model probability $p(\mathcal{M}_j)$. Using Bayes' theorem updating the prior probabilities with the data leads to the posterior model probabilities (see, e.g., Leamer, 1978; Hoeting et al., 1999):

$$p(\mathcal{M}_j|y) = \frac{p(y|\mathcal{M}_j)p(\mathcal{M}_j)}{\sum_i p(y|\mathcal{M}_i)p(\mathcal{M}_i)}, \tag{2}$$

Here, $p(y|\mathcal{M}_j)$ denotes the marginal likelihood of $\mathcal{M}_j$. Given the posterior model probabilities the posterior distribution of any model statistic θ is then obtaining by integration over the model space.

$$p(\theta|y) = \sum_j p(\theta|\mathcal{M}_j, y)p(\mathcal{M}_j|y). \tag{3}$$

Given these basic considerations it is e.g. straightforward to obtain the posterior distributions of the model average coefficients β_{F_j} corresponding to the focus variables. The fact that the posterior distribution is computed (if the whole model space is considered) to 50% of models that do not include any particular focus variable implies that model averaging has a shrinkage property that helps to keep the effective model size small.

Frequentist model averaging operates along the same lines without invoking a Bayesian rationale and using different terminology. The model weights (corresponding in a Bayesian framework to the marginal likelihoods $p(y|\mathcal{M}_j)$ are typically chosen according to some information criterion with an optional weighting factor (corresponding to $p(\mathcal{M}_j)$ in a Bayesian setting). Examples of this approach include Bayesian averaging of classical estimates (BACE) as in Sala-i-Martin et al. (2004), the Mallows model averaging (MMA) approach of Hansen (2007) or the averaging procedures outlined in Claeskens and Hjort (2008). With respect to frequentist model averaging we will focus in the simulations and the empirical application on the smoothed BIC (S-BIC) model averaging scheme discussed in Claeskens and Hjort (2008), which is in many respects similar to BACE. S-BIC model averaging computes the model weights as proportional to the BIC value of the respective model (scaled or smoothed by the sum of the BIC values of all models), i.e.

$$w^{S-BIC}(\mathcal{M}_j) \propto \exp\left(-\frac{1}{2}BIC(\mathcal{M}_j)\right) \propto N^{-\frac{k_j}{2}}(1-R_j^2)^{-\frac{N}{2}}, \tag{4}$$

where k_j is the number of variables included in $\mathcal{M}_j$ and R_j^2 is the corresponding R^2.

The Bayesian model averaging literature provides many alternative specifications of the marginal likelihoods $p(y|\mathcal{M}_j)$, see e.g. Hoeting et al. (1999). Following Fernández et al. (2001a), however, most contributions concentrate on the Zellner (1986) g-prior in the normal-gamma conjugate framework. This setting is conceptually closely related to OLS with shrinkage. Recent research (for an overview see Ley and Steel, 2011b) has extended the g-prior framework by introducing hyper-priors for the shrinkage parameter. In this paper we also use such a 'hyper-g' prior framework by using a beta-prior for the shrinkage parameter (as introduced in Liang et al., 2008).

With respect to the other ingredient in Bayesian model averaging, i.e. the prior model probabilities $p(\mathcal{M}_j)$, the standard choice in the literature is to use a uniform prior attaching the same probability to each and every model. Several authors (e.g., Sala-i-Martin et al., 2004; Ley and Steel, 2009) have criticized this approach since it translates into a prior model size distribution that is tightly concentrated around a mode at half of the focus variables that is averaged over (not counting the included control variables for the moment), which may not be desirable in a situation with a large number of variables relative to sample size. We implement model averaging with both a uniform model prior as well as the models priors proposed by Sala-i-Martin et al. (2004) and Ley and Steel (2009) that allows to control the prior model size distribution.[1]

4.2.2 Principal Components-augmented Regressions

We consider for the moment the full regression including all focus variables X_F and the control variables X_C, i.e.

$$y = X_F\beta_F + X_C\beta_C + u. \tag{5}$$

Without loss of generality we assume that all regressors have mean zero, or equivalently that an intercept is included. The information for regression (5) contained in X_C is equivalently summarized in the set of (orthogonal) principal components corresponding to X_C, as described next. The principal components are the set of transformed variables $\breve{X}_C = X_C O$, with $O \in \mathbb{R}^{k_C \times k_C}$ computed from the eigenvalue decomposition of $\Sigma_{X_C} = X_C'X_C$ (due to the assumption of zero means):

$$\begin{aligned}\Sigma_{X_C} = X_C'X_C &= O\Lambda O' = [O_1\ O_2]\begin{bmatrix}\Lambda_1 & 0\\ 0 & \Lambda_2\end{bmatrix}\begin{bmatrix}O_1'\\ O_2'\end{bmatrix}\\ &= O_1\Lambda_1 O_1' + O_2\Lambda_2 O_2',\end{aligned} \tag{6}$$

where $O'O = OO' = I_{k_C}$ and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_{k_C})$, $\lambda_i \geq \lambda_{i+1}$ for $i = 1, \ldots, k_C - 1$. The partitioning into variables with subscripts 1 and 2 in (6) will become clear in the discussion below. From (6) the orthogonality of the variables in $\breve{X}_C$ is immediate, since $\breve{X}_C'\breve{X}_C = \Lambda$.

Let us consider the case of multi-collinearity in X_C first (which e.g. necessarily occurs when $k_C > N$) and let us denote the rank of X_C with r. Take $\Lambda_1 \in \mathbb{R}^{r\times r}$, hence $\Lambda_2 = 0$ and $X_2'X_2 =$

[1]This will become a more relevant concern in section 4.3.2.

$O_1\Lambda_1 O_1'$. The space spanned by the columns of $X_C \in \mathbb{R}^{N\times k_C}$ coincides with the space spanned by the orthogonal regressors $\tilde{X}_C = X_C O_1 \in \mathbb{R}^{N\times r}$, i.e. with the space spanned by the r principal components. Thus, in this case regression (5) is equivalent to the regression

$$y = X_F\beta_F + \tilde{X}_C\tilde{\beta}_C + u \tag{7}$$

in the sense that both regressions lead to exactly the same fitted values and residuals. Furthermore, in case $[X_F\ \tilde{X}_C]$ has full rank, regression (7) leads to unique coefficient estimates of β_F and $\tilde{\beta}_C$. Therefore, the use of principal components is one computationally efficient way of overcoming multicollinearity. The facts discussed above are, of course, well known results from linear regression theory.

Using principal components in case of full rank of X_C and hence of Σ_{X_C}, however, also has a clear interpretation and motivation. In such a situation replacing X_C by the first r principal components $\tilde{X}_C$ leads to a regression where the set of regressors $\tilde{X}_C$ spans that r-dimensional subspace of the space spanned by the columns of X_C which minimizes the approximation error to the full space in a least squares sense. More formally the following holds true, resorting here to the population level.[1] Let $x_C \in \mathbb{R}^{k_C}$ be a mean zero random vector with covariance matrix Σ_{X_C} (using here the same notation for both the sample and the population covariance matrix for simplicity). For a given value of r consider a decomposition of x_C into a factor component and a noise component, i.e. a decomposition $x_C = Lf + \nu$, where $f \in \mathbb{R}^r$ is random, $L \in \mathbb{R}^{k_C\times r}$ is non-random and $\nu \in \mathbb{R}^{k_C}$ is noise. If the decomposition is such that the factors f and the noise ν are uncorrelated, i.e. orthogonal, then $\Sigma_{X_C} = L\Sigma_f L' + \Sigma_\nu$, with Σ_f denoting the covariance matrix of f and Σ_ν denoting the covariance matrix of ν. Principal components analysis performs such an orthogonal decomposition of x_C into Lf and ν so that the noise component is as small as possible, i.e. it minimizes $\mathbb{E}(\nu'\nu) = \mathrm{tr}(\Sigma_\nu)$. As is well known, the solution is given by $f = O_1'x_C$, $L = O_1$, with $O_1 \in \mathbb{R}^{k_C\times r}$ and $\nu = O_2O_2'x_C$, using the same notation for the spectral decomposition as above.

In many applications it is advisable to compute the principal components not from the covariance matrix but from the correlation matrix. In data sets with variables of substantially different magnitudes computing the principal components based on the covariance matrix leads to essentially fitting the 'large' variables, whereas the computation based on the correlation matrix corrects for scaling differences and leads to a scale-free computation of the principal components. For this reason, we will only consider principal components computed from the correlation matrix (respectively, standardize indicators to unit variance) throughout this paper. To be precise, in this case a so-called weighted principal components problem is solved in which the function to be minimized is given by $\mathbb{E}(\nu'Q\nu) = \mathrm{tr}(Q\Sigma_\nu)$ with $Q = \mathrm{diag}(\sigma^{-2}_{x_C,1},\ldots,\sigma^{-2}_{x_C,k_C})$.[2] This leads to $f = O_1'Q^{1/2}x_C$, $L = Q^{-1/2}O_1$ and $\nu = Q^{-1/2}O_2O_2'Q^{1/2}x_C$. This implies that the principal components to be included in the regressions are given by $\tilde{X}_C = X_CQ^{1/2}O_1$.

Including only r principal components $\tilde{X}_C$ instead of all regressors X_C has a clear interpretation: the principal components augmented regression (PCAR) includes 'as much information as possible' with r linearly independent regressors contained in the space spanned by the columns of X_C. We write the PCAR as:

$$y = X_F\beta_F + \tilde{X}_C\tilde{\beta}_C + \tilde{u}, \tag{8}$$

[1] I.e. we now consider the k_C-dimensional random vector x_C for which a sample X_C of size N is available.

[2] Performing the spectral decomposition on a correlation matrix allows for another simple descriptive criterion concerning the number of principal components. By construction the trace of a correlation matrix equals its dimension, i.e. is equal to k_C. Therefore, if all k_C eigenvalues were equally large, they all would equal 1. This suggests to include as many principal components as there are eigenvalues larger than 1, i.e. to consider the eigenvalues larger than 1 as big and those smaller than 1 as small. The results correspond closely to those obtained with VPC_α with $\alpha = 0.2$.

neglecting in the notation the dependence upon the (chosen) number of principal components r, but indicating by using $\tilde{u}$ the fact that the residuals of (8) in general differ from the residuals of (7). Including only the information contained in the first r principal components of X_C in the regression when the rank of X_C is larger than r of course amounts to neglecting some information and hence leads to different, larger residuals. Thus, in comparison to the full regression (5), the PCAR regression will potentially incur some bias in the estimates which has to be weighed against the benefits of lower estimator variance.

4.2.3 Combining Model Averaging and Principal Components

Performing model averaging in a PCAR framework is straightforward and the simplest approach is to *always* include a certain number of principal components and to average only over the focus variables. The key question is the determination of the number of principal components to be included. The more principal components are included, the lower will be the (omitted variable) bias, but the larger will be the variance and any choice concerning the number of principal components amount to an implicit positioning on the bias-variance trade-off.

In most empirical applications of PCAR, the choice concerning r, i.e. the number of included principal components, rests on the eigenvalues λ_i, where large eigenvalues are attributed to latent factors and small ones to the noise component. The literature provides many approaches and we have experimented with several of them.[1] A widely-used descriptive approach that is found to perform well in practice is given by the variance proportion criterion (VPC):

$$r_{\text{VPC}(\alpha)} = \min_{j=1,\ldots,K_C} \left\{ j \,\middle|\, \frac{\sum_{i=1}^{j} \lambda_i}{\sum_{i=1}^{K_C} \lambda_i} \geq 1 - \alpha \right\}, \tag{9}$$

with $\alpha \in [0, 1]$. Thus, $r_{\text{VPC}(\alpha)}$ is the smallest number of principal components such that a fraction $1 - \alpha$ of the variance is explained. In the simulations and also the applications, setting $\alpha = 0.2$, i.e. explaining 80% of the variance, leads to reasonable numbers of included principal components.

Model averaging conditional on a fixed number $r_{\text{VPC}(\alpha)}$ of principal components is a simple device to tackle (omitted variables) bias of the estimators for the coefficients corresponding to the focus variables whilst keeping the estimator variance low due to the inclusion of only a comparably small number of mutually orthogonal principal components. The number of principal components reflects the 'structure' of the control variables, but the approach described so far has the disadvantage that the relevance of the principal components for the dependent variable is not taken into account in the procedure. This limitation is overcome by performing model averaging not only over the focus variables but over both focus variables and principal components computed from the control variables.

[1] In addition to the results reported in the paper the numvariables ber of principal components has also been determined using the testing approaches of Lawley and Maxwell (1963), Malinowski (1989), Faber and Kowalski (1997), Schott (2006) and Kritchman and Nadler (2008). In a variety of simulations, however, the VPC criterion and a simple eigenvalue test based on the correlation matrix have performed best. Using the VPC criterion can also be interpreted as a *regularization* device for linear regression when the data are in fact not generated by a factor model, whereas the mentioned tests have been derived explicitly for factor models.

4.3 Model averaging over the number of included PCs - A variable group prior

Any choice of the number of included principal components (PCs) r has to a certain extent a heuristic character and has to trade off good approximation (necessary to capture the information contained in all explanatory variables to have small bias) with a sufficiently small number of principal components (necessary for well-behaved regression analysis with low variance). In a model averaging (MA) framework, such uncertainty may be obviously addressed by averaging over models with varying r - i.e., estimating every possible combination of PC-augmented model, for each possible number of included principal components. Keeping up with the logic of PCAR, any combination of focus variables could be conditioned on r PCs with r ranging from 0 to $r_{max} \leq K_C$.[1] This exploitation of the inherent hierarchical ordering of PCs induced by the eigenvalues still presents a computational advantage vs outright model averaging: Applying model averaging over focus variables as well as the number of PCs implies a model space of $2^{K_F} \times (K_C + 1)$ potential models $\mathcal{M}_j$ – a space that is considerably smaller than under outright model averaging over both focus and control sets with $2^{K_F + K_C}$ potential models. As long as there is positive prior mass on each of these models, the posterior estimate of r would approach the 'true' number of latent factors.[2]

4.3.1 Motivation

While such a principal components-model averaging (PC-MA) approach is expected to be more robust than outright model averaging in short samples, both methods should of course asymptotically converge to the same estimates with increasing number of observations. Note, however, that in finite samples, estimates depend on priors – and applications of PC-MA and MA on the same data are only equivalent if the model prior structure is taken into account. Consider, for instance, a model averaging case with one focus and 9 control variables. Under the 'uniform' mode prior (constant prior over all models), the prior model probability of the null model would be 2^{-10}, and the prior inclusion probability for inclusion of at least one control variable would be $1 - 2^{-9}$. Applying PC-MA to the control variables in this case reduces the model space from 2^{10} to $2 * (9 + 1)$. Imposing a uniform model prior on the PC-MA structure would consequently imply a prior model probability for the null model of $\frac{1}{2*10}$, and the prior inclusion probability for the control variables would decrease to $1 - \frac{2}{2*10}$. I.e., PC-MA under a uniform model prior would weaken the prior odds of finding the control variables important, compared to outright model averaging (under the same prior). This issue worsens if the number of considered PCs is smaller then the number of original control variables (for instance, if the number of control variables is larger than the number of observations). In order to avoid such unwanted concentration of prior odds in focus variables, an adjustment of the model prior structure is necessary if the objective is to stay close to the (usually neutral) prior mass distribution from outright model averaging.

Keeping up the prior equivalence with outright model averaging also calls for interior discrimination among models with varying numbers of principal components: Note that eigenvalues represent the degree of overall control group variance embodied in their corresponding principal components. In the previous example with 9 control variables, a uniform model prior implies a prior expected number of 4.5 included control variables (conforming to each control variable's prior inclusion

[1] The maximum number of included PCs r_{max} could be smaller than the number of control variables K_C on which the PCs are based. For robustness reasons and computational simplicity, r_{max} might be chosen according to some suitable cut-off criterion.

[2] Note that 'asymptotic' here refers to increasing sample size while keeping the number of controls and focus variables constant.

probability of $\frac{1}{2}$). Under PC-MA with a uniform prior, the prior expected number of included PCs would equally be 4.5. However, one objective in extracting principal components is to improve robustness by dimensional reduction. In applying such a transformation, a typical research set-up would expect the number of included principal components under PC-MA to be (weakly) less than the prior expected number of included control variables under outright model averaging. Indeed, when considering that the eigenvalue of each principal component represents overall control set variance in terms of number of variables,[1] then such a uniform prior on PCs would imply a prior included number of 'represented' control variables equal to $\frac{\sum_{i=1}^{9}(10-i)\times\lambda_i}{2\times 10} \geq 4.5$.[2] In view of prior equivalence with MA, but also the inherent motivation of cut-off criteria like VPC, it should thus be more appropriate to adjust the prior number of 'represented' control variables to their appropriate scaling under MA. For instance, if all 9 control variables in this example were collinear, then the first PC would represent all control set variance, and including further PCs would be pointless. If, in contrast, the control variables were orthogonal, then each PC would be (up to sign changes) equal to the original untransformed control variables. In such a case, all 9 PCs should be considered equally (under the hierarchical approach described above, this would amount to just considering models that include all nine – or no – PCs) . Note that in these two cases, the prior included number of PCs would differ, but the prior included number of 'represented' control variables would be 4.5 in both.

If the aim is to establish a prior weight of models that conforms to outright model averaging then such a prior should keep constant the prior importance attached to focus variable and the null model, and adjust the set of prior weights on control PCs such that the prior on 'represented' control variables remains unchanged. The following paragraphs develop a more general prior that conforms to such considerations.

4.3.2 A PC-MA group prior

While so far the discussion has focused on extracting PCs from a single set of control variables, this could be generalized to regularizing distinct variable groups: For instance, in order to gauge the effect of institutional settings on economic growth (e.g., Kraay and Tawara, 2010), institutional proxies could be grouped into separate buckets according to their nature, and transformed into principal components to assess the impact of the respective concept. (Such a setting incorporates the notion of untransformed focus variables vs. a PC-transformed control group, since each focus variable can be considered as a singleton variable group.) In view of the implications on prior model mass above, such a setting should be able to mimic the prior mass distribution among groups that obtains under outright model averaging. The following paragraphs lay out a model prior framework that achieves just that.

Suppose the set of explanatory covariates X can be divided into L distinct variable groups, each comprising $K_{C,l}$ variables for $l \in \{1, \ldots, L\}$. For expositional simplicity, let the MA model prior framework be the Sala-i-Martin et al. (2004) binomial model prior $p(k_\gamma) = \theta^{k_\gamma}(1-\theta)^{k_\gamma}$, where k_γ is the number of included variables in model M_γ, and θ denotes a predefined constant prior inclusion probability for each variable. (Note that the uniform model prior is sub-case hereof, with $\theta = \frac{1}{2}$.) Under a transformation of the L variable groups into the PC-MA framework, the model prior should satisfy the following requirements: (a) For each transformed variable group l, a transformation should not affect the prior model probabilities of all other groups and the null

[1] Remember that here, we exclusively consider principal components computed from correlation matrices, thus the sum of their corresponding eigenvalues is restricted to the number of variables that are regularized.

[2] Here, λ_i represent the $i - th$ eigenvalue of the correlation matrix of X_C, ordered by the size of eigenvalues.

model. (b) Additionally, the posterior inclusion probability of each individual PC from group l $PC_{l,i}$ should be proportional to its corresponding eigenvalue $\lambda_{i,l}$ (both requirements might be relaxed, see below).[1]

The following prior satisfies both requirements...[2]

$$p(M_\gamma) = \prod_{l \in M_\gamma} \Theta_l f(M_\gamma, l) \prod_{l \notin M_\gamma} (1 - \Theta_l) \tag{10}$$

$$\text{with} \quad f(M_\gamma, l) \equiv \frac{\lambda_{k_{C,l},l}^{\alpha_l} - \lambda_{k_{C,l}+1,l}^{\alpha_l}}{\lambda_{1,l}^{\alpha_l}} \quad \text{where} \quad k_{C,l} = \arg\min_j \lambda_{j,l} \in M_\gamma$$

... under the following parameter choices:

$$\begin{array}{ll} \text{(A)} & \Theta_l = 1 - (1-\theta)^{K_{C,l}} \quad \forall\, l \\ \text{(B)} & \alpha_l = 1 \quad \forall\, l \end{array}$$

This prior set-up keeps the prior model probabilities of all singleton variable groups (i.e., focus variables) and the null model unchanged with respect to outright model averaging. Moreover, the prior inclusion probability for variable group l (which corresponds to the prior inclusion probability of $PC_{1,l}$) is equal to the prior probability of inclusion of at least one group variable under the MA set-up Θ_l.[3]

Under the lead example used in this paper, with K_F focus variables and K_C control variables, the application of this prior boils down treating each focus variable as a singleton control set. If parameter choices (A) & (B) hold, the $\Theta_F = \theta$ for each focus variable, yielding the following representation of the prior (here, k_F denotes the number of focus variables included in M_γ):

$$p(M_\gamma) = \begin{cases} (1-\theta)^{K_C + K_F - k_F} \theta^{k_F} & \text{if } M_\gamma \text{ includes no PC} \\ (1 - (1-\theta)^{K_C})\theta^{k_F} (1-\theta)^{K_F - k_F} f(M_\gamma, 1) & \text{if } M_\gamma \text{ includes PCs} \end{cases}$$

Straightforward algebra reveals the prior model size under PC-MA (the prior expected number of included focus variables and principal components) as follows:

$$E(k) = \sum_{l=1}^{L} \Theta_l \frac{\sum_{i=1}^{r} \lambda_{i,l}^{\alpha_l}}{\lambda_{1,l}} \tag{11}$$

Under the example with K_F focus variables and one control set with K_C variables, the parameter choices (A) and (B) imply a prior model size as follows:[4]

$$E(k) = E(k_F) + \Theta_1 \frac{K_C}{\lambda_1}$$

The two parameters Θ_l and α_l allow to fine-tune the prior importance of variable groups,with Θ_l steering the importance *between* variable groups and α_i *within* principal components. The smaller

[1] Since for each control variable group, only $r_l \leq K_{C,l}$ eigenvalues can be considered, define $\lambda_{r+1,l} \equiv 0$ for ease of notation.

[2] Here, $l \in M_\gamma$ denotes models that include at least one principal component out of control set l.

[3] Note that condition (a) implies that the prior probability for at least one variable of group l to be included should remain the same under both the MA and the PC-MA representation.

[4] Here, $E(k_F)$ denotes the prior expected number of included focus variables $E(k_F) = \theta K_F$

the parameter Θ_l, the less prior importance is attached to group l. Setting $\Theta_l = 1$, in contrast, forces each model to include at least one principal component of group l. The parameter thus allows to adjust for biased prior probabilities due to group effects, e.g. if two 'equally important' variable groups comprise a different number of individual proxies. The parameter α_l is bijectively related to the prior expected number of included PCs, conditional on inclusion of at least one PC out of group l. While $\alpha_l = 1$ sets the prior inclusion probability of each PC proportional to its eigenvalue, $\alpha > 1$ skews prior mass versus the first-most PCs. With $\alpha_l \to \infty$, the prior set-up concentrates on the first PC of group l, while prior mass on the other PCs will vanish. In contrast, $\alpha_l = 0$ favors models with *all* PCs (of group l), while models with fewer PCs obtain zero prior importance. For practical matters, one might elicit the parameter α_i via specifying the prior expected number of included PCs (of group l, conditional on prior model probabilities of models that include group l).[1] All in all, however, only the parameter choices (A) and (B) allow for equivalence with a prior set-up under outright model averaging. For this reason, we will retain these parameter choices throughout the paper, except where otherwise indicated.[2] Finally, note that the PC-MA prior presented here is built on the binomial Sala-i-Martin et al. (2004) prior for expositional simplicity. It is straightforward to extend this prior in order to achieve equivalence under the MA set-up with other model priors such as in Ley and Steel (2009) (see appendix).

4.4 An Illustrating Simulation

By means of a simulation exercise, we compare the properties of the various PC-MA approaches versus OLS, PCAR and 'standard' BMA, as well as the 'weighted average least squares' (WALS) approach by Magnus et al. (2010). There is, of course, an ad-hoc character about any such exercise – therefore we strive for a set-up that is similar to typical cases from empirical model averaging. As discussed in the introduction, we consider in particular applications where a handful of newly proposed 'alternative' indicators (the focus variables) are evaluated against a large set of established predictors (the controls). (Consider, for instance, the new financial early warning indicators for the business cycle that have been proposed in the wake of the recent financial crisis (e.g., Alessi and Detken, 2011).) Moreover we regard as 'typical' a data set with about a 100 observations and nearly as many potential explanatory variables. The datasets we have in mind are characterized by a rich correlation structure, as well as by non-negligible correlation between focus and control variables. The simulation set-up below formalizes these notions. In order to evaluate the estimation approaches with respect to our two main motivations (robustness of estimates and estimators), we resort to a variant of 'perturbation instability in estimation' by Yuan and Yang (2005): As a proxy for the robustness of estimates, we consider the root mean squared deviations of fitted values from the 'true' simulated values for the dependent variable (note that this corresponds to the notion of out-of-sample prediction errors). For examining the robustness of coefficient estimators, we likewise look at the root mean squared deviation of estimated coefficients for the focus variables from the 'true' coefficients imposed in the simulation set-up.

[1] Although there is a bijection between α_l and the prior number of included PCs (given Θ_l), no general closed-form exists to specify α_l this way (see appendix).

[2] We have experimented with a wider array of prior parameter settings than reported in the following sections. However, results under such priors stay close to the default setting (A) and (B), and do not change at all the qualitative findings discussed below.

4.4.1 The simulation set-up

We structure our simulation as follows: Set sample size at $N = 100$, which is of the order of magnitude encountered in empirical observations.[1] Consider 10 focus variables[2] and 67 'established' control variables. In order to emulate empirical settings, we bootstrap the control variables using the correlation matrix Σ_{SDM} among the 67 covariates in the Sala-i-Martin et al. (2004) dataset. The correlation structure in Σ_{SDM} can be considered quite diverse: While its first principal component (PC) explains 25% of the variance in the (standardized) data set, it takes the first four PCs to account for 50% of its variance, and 23 PCs for 90%. Nonetheless, this correlation matrix was computed on very few observations (88) and consequently displays many very small eigenvalues (i.e., unimportant PCs). This near-collinearity is also expressed by the reciprocal condition number of Σ_{SDM}, which is 9.38×10^{20}.

$$X_C = \text{Std}\left(\tilde{Z}_C \text{ Chol}(\Sigma_{SDM})\right)$$

Here, $\tilde{Z}_C$ denotes a $N \times 67$ vector of IID normally distributed random elements, and Chol denotes the Cholesky decomposition operator. Moreover, let Std be an operator that normalizes the columns of X_C to zero mean and unit variance. The focus variables display a certain correlation among themselves, as well as with the control variables. In order to ensure the positive-definiteness of the result variance-covariance matrix over both controls and focus variables, we opted for the following approach:

$$X_F = \text{Std}\left(\tilde{Z}_F \text{Chol}((1-\rho_F)\mathbf{I}_{10} + \rho_F \mathbf{1}\mathbf{1}') + \theta_{CF} X_C P_C \Lambda_C^{-\frac{1}{2}} (\mathbf{I}_{10}\ \mathbf{0})'\right)$$

Here $\tilde{Z}_F$ denotes an $N \times 10$ matrix with normally IID elements, and ρ_F is a parameter to induce correlation among the independently distributed components of X_F. P_C and Λ_C denote the eigenvector and eigenvector matrix of X_C, respectively – the expression $X_C P_C \Lambda_C^{-\frac{1}{2}} (\mathbf{I}_{10}\mathbf{0})'$ therefore denotes the 10 most important principal components of X_C. The parameter θ_{CF} induces a rich correlation structure among X_C and X_F, and is closely related to the RV coefficient, a measure of similarity between two data sets.[3] Instead of applying θ_{CF} to the principal components of X_C, we considered several alternative specifications which yielded very similar outcomes at a cost of a more clustered covariance matrix between X_F and X_C. Note that in the simulation exercise, we specify the parameter θ_{CF} to account for the total amount variance induced by this structure.

In order to define the response variable, we first specify the model for the 'true' outcome, i.e. without residuals: For a generalized representation of the controls, let the coefficient vector for the control variables $\tilde{b}_C \sim U(0, 4)$ be based on a uniform distribution bounded by zero and four. This resulting summand $X_C\tilde{b}_C$ is pre-multiplied by the scalar θ_C in order to define the relative importance of control vs. focus variables. Include two focus variables in the 'true' model; Since the focus variables generated for X_F are interchangeable, restrict the coefficient vector β_F such that

[1] For instance, the data set by Sala-i-Martin et al. (2004) has 88 observations for 67 regressors, the application by Fernández et al. (2001a) uses 72 observations for 41 covariates, while the model averaging application of Giannone et al. (2010) uses 27 covariates with only 42 observations.

[2] We chose this number as it allows for computationally easy enumeration of all potential models.

[3] The RV coefficient is a multivariate generalization of the R-squared, and is defined as $\frac{Tr(E(x_F' x_C)E(x_C' x_F))}{\sqrt{Tr(E(x_C' x_C)^2)}\sqrt{Tr(E(x_F' x_F)^2)}}$, where $E(x_A' x_B)$ denotes the covariance matrix between X_A and X_B.

its third to tenth element are zero, while the first two elements are specified as $\frac{1}{2}$.[1]

$$y_{\text{true}} = \frac{1}{\sigma_{\text{true}}} \left(X_F \beta_F + \theta_C \text{Std} \left(X_C \tilde{b}_C \right) \right) \tag{12}$$

Here, we divide the result in parentheses by its expected standard deviation σ_{true} in order to render the variance of y_{true} close to one. The 'true' coefficients for the first two focus variables are therefore each $\frac{1}{2\sigma_{\text{true}}}$.[2] The 'observed' response variable is finally generated by $y = y_{\text{true}} + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2 \mathbf{I}_N)$. In the vein of Yuan and Yang (2005), the analysis of simulation results will mostly concentrate on the impact of changes to the variance parameter σ. Note that in this set-up, any decrease of σ has the same impact as an increase of sample size N at constant variance.

4.4.2 Estimation approaches and robustness evaluation

We consider nine estimation approaches for evaluation in the simulation exercise: 'Full OLS' uses all 77 simulated variables (plus an intercept) in ordinary least squares. 'PCAR' applies PC-augmented regression on the focus variables and a number of principal components as implied by a pre-set Variance Proportion Criterion (VPC) of 80% (which is close to the VPC used in comparable applications). 'Standard FMA' applies Frequentist model averaging indiscriminately over focus and control variables.[3] For all cases representing FMA, we apply the S-BIC variant (Claeskens and Hjort, 2008) as discussed in the previous section (for 'standard FMA', these are combined with with uniform model priors). Among the competing methods, we also consider two variants of the weighted-average least squares (WALS) estimator proposed by Magnus et al. (2010), which follows an approach that is directed at the same problem as ours, but follows the opposite approach. The set-up 'WALS w.aux.' is close to the example used by the authors, in that their equivalent of principal-component model averaging is applied over the 10 focus variables[4], after they have been conditioned on the 67 'auxiliary' control variables by means of OLS. Since the large OLS component in 'WALS w.aux.' leads us to expect, a priori, that estimation results will be close to OLS, we also consider another variant of WALS. The model 'WALS no aux.' applies the conceptual equivalent of model averaging indiscriminately over all 77 variables.

We consider four approaches that combine model averaging with principal components according to the concepts introduced in the previous section: 'PC-MA (fixed, FMA)' performs Frequentist model averaging over the focus variables, conditioned on a fixed set of PCs (as given by the VPC of 80%).[5] The approach 'PC-MA (unif., FMA)' applies S-BIC model averaging over both the focus variables and the PCs (up to a maximum PC number as implied by the VPC), with a simple uniform model prior structure. 'PC-MA (prior, FMA)' differs in that it imposes the PC-adjusted prior on the principal components that was introduced in Section 4.2. 'PC-MA (prior, BMA)', in contrast, represents a purely Bayesian approach under state-of-the-art priors: it introduces model-adaptive shrinkage via the hyper-g coefficient prior (Feldkircher and Zeugner, 2009; Liang et al., 2008) and

[1] Note that these coefficients may be fixed to any non-zero value, as the relative importance of X_F vs. X_C is entirely controlled for by the parameter θ_C.

[2] Inspection of the simulation set-up yields $\sigma_{\text{true}} = \sqrt{\theta_C^2 + \frac{1}{4}\text{Var}(X_F \beta_F)}$. For the application at hand, $\text{Var}(X_F \beta_F)$ is closely bounded between 0.7^2 and 0.75^2.

[3] As it is infeasible to enumerate 2^{77} potential models, 'standard FMA' resorts to MCMC sampling for an approximation of posterior results. It relies on the MCMC model sampler used by Fernández et al. (2001a), with 50,000 burn-ins and 450,000 subsequent draws for each simulation run. The other estimation strategies in this section are all estimated by full enumeration of the model space, and thus do not require MCMC sampling.

[4] Note that Magnus et al. (2010) use a different naming convention: Their 'focus variables' conceptually correspond to 'control variables' in this paper, while their 'auxiliary variables' are referred to as 'focus variables' here.

[5] Note that this set-up conforms to the one proposed by Wagner and Hlouskova (2010).

combines the adjusted prior with the binomial model prior by Sala-i-Martin et al. (2004) in order to loosen the prior on expected model parameter size.[1]

Model	Description
full OLS	Ordinary least squares over all 77 simulated variables (plus an intercept).
PCAR	principal components-augmented regression: least squares over 10 'focus' variables and principal components extracted from the 67 'control' variables. The number of included factors conforms to an 80% variance proportion criterion (VPC).
Standard FMA	Outright Frequentist Model Averaging (FMA) over all 77 variables. The FMA variant here is S-BIC as in Claeskens and Hjort (2008) (i.e. using uniform model priors).
WALS w.aux.	'Weighted average least squares' (WALS) as proposed in Magnus et al. (2010), with WALS applied over the 10 'focus' variables, conditioned (via least squares) on the 67 controls as 'auxiliary variables'.
WALS no aux.	'Weighted average least squares' (WALS), applied indiscriminately over all 77 variables.
PCMA (fixed, FMA)	Frequentist model averaging (FMA) as above, applied on the 10 'focus' variables, after having conditioned (via least squares) on a fixed number of principal components extracted from the set of 67 control variables (as in section 4.2.3). The number of included control factors is determined by an 80% VPC.
PCMA (unif., FMA)	Frequentist (S-BIC) model averaging over models with the 10 focus covariates and a variable number of included control principal components, with uniform priors over all model combinations (as mentioned in the motivation of section 4.3).
PCMA (prior, FMA)	Similar to 'PCMA (unif., FMA)', but with the PC-MA model prior as outlined in section 4.3.2).
PCMA (prior, BMA)	Model averaging with the PC-MA prior similar to 'PCMA (prior, FMA)', but with individual models estimated via normal-gamma conjugate Bayesian regression with adaptive shrinkage (Feldkircher and Zeugner, 2009).

Table 4.1: Outline of model variants used for simulations

In our main simulation set-up, we look at the robustness of these estimation strategies under varying degrees of residual variance σ. For any given parameter set $(\sigma, \rho_F, \theta_{CF}, \theta_C)$, we perform 40 bootstrapping runs: Each run of these runs corresponds to a joint draw of the random vectors $\tilde{Z}_F$, $\tilde{Z}_C$ and $\tilde{b}_C$, as well as the error term ε. In order to assess the quality of the estimation strategies, we compute two main indicators over the bootstrapping runs:

- Robustness of estimates: The root mean squared deviation of the fitted values from y_{true} for the robustness of estimates. The slope of this indicator with respect to σ corresponds to the notion of perturbation instability in estimates proposed by Yuan and Yang (2005).[2]

[1] For 'BMA adjusted PC', we tuned the parameters of the beta-binomial model prior such to attain a prior expected model size of 67/2 when taking the prior structure on the PCs into account. The model prior thus represents a uniform distribution over prior model size.

[2] Yuan and Yang (2005) propose to evaluate the robustness of estimates by bootstrapping from a real-world data

- Coefficient robustness: As a proxy for realized coefficient risk, we compute the root mean squared deviation of coefficient estimators from their 'true' values for 1) the focus variables that are included in the 'true' model, and 2) the focus variables that were not considered in constructing y_{true}.

4.4.3 Simulation results

Figure 4.1–4.4 present the simulation results for values of σ in the range $[0.2, 2.2]$. Note that since y_{true} has been modeled to unit expected variance, the parameter is monotonously related to the expected R-squared (squared correlation coefficient) between y_{true} and y. The other parameters have been chosen as to render some standard statistical measures similar to their behavior among subsets of the Sala-i-Martin et al. (2004) data set: Setting $\rho_F = 0.3$ and $\theta_{CF} = 0.6$ yields an RV-coefficient between X_F and X_C that conforms to the RV coefficient between random subsets of 10 and the remaining 57 variables out of the data set. The parameter for the relative importance of X_C in determining y_{true} was set to $\theta_C = 0.2$ in order to make the variance of the two summands in equation (12) comparable.

The robustness of estimates in response to noise increases is displayed in Figure 4.1. As may be expected, 'Full OLS' performs well for small levels of noise, but its performance worsens steeply with increasing levels of σ. In contrast, 'PCAR' is much more more effective for intermediate to large levels to noise, a result that is well-established in the literature (e.g., Stock and Watson, 2006). Interestingly, the pseudo-forecasting results for plain model averaging ('Standard FMA') are very close to those of 'PCAR'. Their slight disparity stems mainly from the differing estimators in focus variables, while the results for the control variables are nearly interchangeable – this finding is not surprising in view of the conclusions reached by De Mol et al. (2008) in a similar context.[1] The 'MA fixed PC' approach intends to combine the virtues of model averaging and PC factor models, and it actually dominates both its siblings by a non-negligible degree. Since the 'WALS w.aux.' approach is characterized by a large OLS component, it does not surprise in hardly outperform 'Full OLS' in terms of robustness. In contrast, the 'WALS no aux.' variant performance astonishes in being considerably worse than the 'PCAR', 'Standard FMA' and 'MA fixed PC' set-ups.

The major improvement in robustness is achieved by the estimation strategies that average over the number of included principal components. Easing the restriction on the number of PCs seems the most important driver of their performance. In contrast, the impact of the various priors seems negligible, which indicates that the data set-up (in particular on the PCs) is strong enough to dominate any well-behaved prior distribution. The outperformance of model averaging with varying PCs may well be due to the fact that the (posterior) expected number of included PCs under these approaches varies between 2 and 5, while the chosen Variance Proportion Criterion induces a number of PCs in 'MA fixed PC' that ranges between 12 and 15 (Note, moreover, that this number is very close to the (posterior) expected number of parameters under 'Standard FMA'). Overall, these results indicate that in terms of robustness for prediction purposes, it might be advisable to forgo the separation of focus variables altogether and directly average over principal components constructed from the unified dataset (X_F, X_C). Yet a major purpose of PC-MA is precisely to improve inference about the importance and coefficients of particular focus covariates.

set and computing the mean squared difference between the (simulated) fitted values for an estimation strategy and the actual OLS fitted values from the initial dataset (divided by the variance σ of the bootstrapping error term). The authors regard the slope of the such constructed indicator with respect to σ at $\sigma = 0$ as proxy for robustness of an estimation strategy.

[1] Note, however, that purely Bayesian model averaging along the lines of 'BMA adjusted PC' (but without a principal component part) performs considerably better than 'Standard FMA'.

In this respect, the results on coefficient estimator robustness provide some insight: Figure 4.2 presents the root mean squared deviation between estimated coefficients[1] for the two included focus variables from their 'true' value $1/2\sigma_{\text{true}}$. The ranking appears broadly similar to the previous case, with OLS worsening its comparative performance with increasing variance. The two WALS variants are virtually indistinguishable and do not represent a major robustness improvement over the remaining methods. Again, the approach 'FMA fixed PC' performs slightly better than 'PCAR'. The indicator is virtually indistinguishable for the two frequentist model averaging approaches with a variable number of PCs ('FMA uniform PC' and 'FMA adjusted PC'), while their fully Bayesian sibling performs considerably better under high levels of noise. The most striking result, though, is the performance of 'Standard FMA', which exhibits much less deviation of the 'true' coefficients from their target. Nonetheless, the outperformance of 'Standard FMA' must be contrasted with the corresponding results from the 'wrong' coefficients in Figure 4.3. Here instead, the PC-MA results outperform the 'Standard FMA' coefficient uncertainty by about the same degree as they underperformed in Figure 4.2. Closer inspection suggests that the differing behavior for 'true' and 'wrong' coefficients results from the fact that the number of included controls/principal components is far less under the PC-MA approaches. We conclude from Figure 4.2 and 4.3 that the advantage of outright 'standard' Model Averaging over PC-based models is its superiority in terms of coefficient uncertainty. However, when model averaging is extended to the number of included PCs, PC-MA performance is comparable to that of outright model averaging. Besides, note that with respect to uncertainty over both 'true' and 'wrong' coefficients, purely Bayesian PC-MA beats its Frequentist peers by a non-negligible margin (which can largely attributed to their differences in prior diffuseness).

Finally, consider the robustness of inference about the importance of focus variables. To that end, we compare the average posterior inclusion probabilities for the 'true' focus variables under the various model averaging approaches.[2] Note that in 'Standard FMA' and all four PC-MA approaches, the posterior expected number of included focus variables varies between 1 and 3 (except for very small noise). Assessing their absolute PIPs therefore leads to the same findings as their relative importance with respect to each other (Feldkircher and Zeugner, 2012, for a related discussion). A first look at Figure 4.4 reveals that all PIPs for the included variables turn out to be quite low. Nonetheless, Figure 4.4 demonstrates that the PC-MA approaches with a variable number of PCs identify the 'true' focus variables far better than 'FMA fixed PC' and 'Standard FMA'. Note that, again, purely Bayesian PC-MA with loose priors priors ('BMA adjusted PC') performs slightly better than its Frequentist analogues under intermediate noise levels.

In addition to the benchmark simulation setting in this Section, we also considered systematic variations of the parameters ρ_F, θ_{CF}, and θ_C under $\sigma = 1$. The relative performance between the seven estimation strategies remains stable over these dimensions, apart from trivial level effects. In particular, changing the relatedness between focus and control variables (by means of θ_{CF}, i.e. the RV coefficient between X_F and X_C) maintains the levels of the robustness indicators at the values observed at point $\sigma = 1$ in Figures 4.1-4.4. Above all, the indifference of results to these other indicators corroborates the findings with respect to robustness to noise (as represented by σ). Principal component-based model averaging with a variable number of PCs is considerably more robust to noise than either model averaging or PC factor models alone. While the quality of PC-MA-based inference about coefficient magnitudes is comparable to outright model averaging, PC-MA (with averaging over PCs) outperforms the latter in inference about the importance of

[1] Note that for this purpose, it is appropriate to take the coefficients for the model averaging approaches conditional on inclusion – i.e., averaged only over models that where the respective coefficient was not restricted to zero.

[2] Note that in the multi-model case, PIPs and t-statistics of OLS models are not readily comparable. We therefore omitted results from 'Full OLS' and 'PCAR' in Figure 4.4.

covariates as predictors.

4.5 Financial Indicators as Business Cycle Predictors

With the onset of the global financial crisis in 2007, the importance of financial markets for economic activity as well as the potential risks emerging from the financial markets for the real economy have become a central concern for policy makers and academic researchers alike. Prior to the crisis, the financial sector has been primarily regarded as facilitating for efficient allocation of resources and risk-sharing across agents and over time, much akin to the logistics sector – but the recent experience has forcefully shown that imbalances on the financial markets may have substantial detrimental effects for real economic activity. In the meantime in the literature a variety of studies has been undertaken trying to quantify the importance of as well as the predictive ability of different financial indicators for real economic activity. Motivated by this example, we assess whether such financial indicators have predictive content for real activity in a simple univariate set-up.

Any econometric assessment of such 'financial' indicators on macroeconomic data faces two challenges: First, macroeconomic data comes at low frequency, while historical data for such financial indicators rarely extends beyond the 1990s. Any comprehensive assessment of financial indicator impact on the business cycle is therefore limited to relatively few observations. This problem is compounded by the fact that not only GDP data, but even some of the proposed financial indicators (e.g., Adrian and Shin, 2010) are only available at quarterly frequency. The second challenge is that a meaningful evaluation of the predictive content of such alternative indicators should be confronted with the large array of 'established' business cycle predictors. Many studies, particularly in the VAR literature, confine themselves to a dozen of established variables (which, together with a dozen financial indicators, would already pose a sizable challenge to the application of simple estimators in a quarterly rolling forecasting framework). But the recent literature found better short-term forecasting performance in models that rely on condensing information from more than 100 predictors (Stock and Watson, 2006). Among these methods rank shrinkage estimators (De Mol et al., 2008), factor-based VARs (Stock and Watson, 2005), or simple dynamic factor models of the type advocated by (Stock and Watson, 2002).

Consequently, this section seeks to address the question whether alternative financial indicators prove to be viable predictors against a large set of control variables, given that historical data availability limits sample size to about 100. The PC-MA framework laid out in this paper lends itself to that purpose, with a separation into control variables along the lines of Stock and Watson (2002) and 'financial' focus variables. The set-up also allows for comparison with other estimation methods evoked in the previous section. For the sake of briefness, we concentrate on forecasting GDP over a 4-quarter forecast horizon.

4.5.1 Data and Set-up

In order to assess their predictive content for the real economy, we consider a variety of financial indicators that have been put forward to capture different aspects and channels through which the financial sector could exert repercussions on the real economy (see Table 4.2 for an overview). We concentrate on indicators with continuous historical data from the 1980s – which considerably reduces the consideration set. Adrian and Shin (2010) include as one of their explanatory variables the growth rate of shadow banking assets, with the shadow banking system comprising ABS issuers, finance companies and funding corporations (SHADBNK_A). The authors argue and demonstrate that the assets of the shadow banking system are an important and informative determinant of funding conditions in the economy.[1] In the same vein, the authors also consider the growth rate of assets in securities brokers-dealers (SBD_A), the set of primary US government bond brokers, as a proxy for leverage in important financial markets intermediaries. The risk-taking channel is in the focus of Altunbas et al. (2010) with the argument being that monetary policy since the 1990s has caused an increase in banks' risk taking due to exceptionally low interest rates. We assess this channel with three variables, namely the delinquency rate on loans secured by real estate (DRSREACBS), the charge-off rate on business loans (STFBQCB), and the rate of non-performing commercial loans (NPCMCM). Real assets, in particular house prices, have regained attention as a predictor financial crises (e.g. Barrell et al., 2010). The house price measure we use is the Standard & Poor's Case-Shiller 10 city home price index (SPCS10RSA). Credit indicators have repeatedly been focused on over decades (e.g., Bernanke, 1993; Bean et al., 2002). While these have been elaborately discussed, we will examine whether simple credit growth indicators deliver predictive content. We focus on two indicators for the US economy that have been repeatedly used for that purpose (e.g., Korobilis, 2009), namely growth in business loans (BUSLOANS) and growth in bank borrowing from the Federal Reserve (BORROW). Gilchrist and Zakrajsek (2010) concentrate on the impact of credit spreads and in particular the so-called excess bond premium on economic activity. This indicators exploit the notion that higher credit spreads reflect decreased aggregate credit quality and higher general risk aversion. Since Gilchrist and Zakrajsek (2010) develop their own credit spread index which is not publicly available we use a very simple, but popular proxy for the credit spread, given by the difference between the 3-month LIBOR and the 3-month treasury bill rate (STED, the 'TED spread'). Clearly, this is only an imperfect measure for credit spread conditions in the real economy, but tentative evidence suggests that such liquid-market indicators co-move strongly and indicate conditions in downstream loan markets well as they do not suffer from survivor bias (as, e.g., in the loan spread). Hatzius et al. (2010) assess the predictive content of financial condition indexes. Financial conditions are essentially the current state of financial variables that impact economic behavior and there (future) economic activity. Typical indexes include variables intended to characterize the supply or demand of financial instruments, see Hatzius et al. (2010) for a broader conceptual discussion. In our analysis we include the Chicago Fed adjusted national financial conditions index (ANFCI). Kollmann and Zeugner (2012) contains a detailed study of the predictive performance (various measures of) leverage, defined as the ratio of a sector's assets to its net worth, for real economic activity respectively its volatility. Amongst the many specifications and combinations investigated in that study, the median of the standardized Year-on-Year (YoY) changes of the book-value based leverage of the financial sectors property and life-insurance companies, securities brokers and dealers, commercial banks, households and non-financial corporate businesses (MEDFoF) performs well. We therefore also consider this specific leverage variable in our study. In total, we dispose of 11 financial indicators with a reasonable amount of historical data. Other indicators were also assessed, but they are either very similar to one of these 11 indicators, or their historical data leaves too few observations to base a quarterly exercise on. Still, the relatively short history of these 11 indicators puts the earliest date for which all are available in their proper transformations at 1988Q1.

For the macroeconomic controls, we resort to the quarterly data set used in Stock and Watson (2011), with 143 macroeconomic indicators. Since their data set ends in 2003, we extend it with data up to 2011Q3 from the same sources (see Stock and Watson (2011, Table B.1) for the list of indicators, sources and transformations).[2] 35 among these variables represents aggregate indicators, which are linear combinations of their sub-components. Following Stock and Watson (2011), we

[1]The size (in terms of assets) of the shadow banking systems has been strongly increasing since the mid 1980s to be of about the same magnitude as the assets of the commercial banks at the eve of the crisis.

[2]The leading indicators 'index of help-wanted advertising in newspapers' and 'ratio of help-wanted ads to unem-

omit this aggregate indicators from estimation, respectively principal components decomposition, which leaves 106 macro-economic indicators (see Stock and Watson, 2011, for the list of included variables).

ployed' were discontinued since 2003 (e.g.,) which reduces the data set used here to 141 variables.

Variable	Acronym	Similar variables used by	Transformation
Growth rate of shadow banking assets; shadow banks are ABS issuers, finance companies and funding corporations	SHADBNK_A	Adrian and Shin (2010)	log-difference
Growth rate of security brokers and dealers assets	SBD_A	Adrian and Shin (2010)	log-difference
Delinquency rate on loans secured by real estate; all commercial banks	DRSEACBS	Altunbas et al. (2010)	first difference
Charge-off rate on business loans; all commercial banks	STFBQCB		first difference
Commercial and industrial loans at all commercial banks	BUSLOANS	Barrell et al. (2010)	log-difference
S&P Case-Shiller 10 city home price index	SPCS10RSA		log-difference
TED Spread (3 month LIBOR - 3 month T-bill rate)	STED	Gilchrist and Zakrajsek (2010)	–
Chicago Fed adjusted national financial conditions index	ANFCI	Hatzius et al. (2010)	–
Median of YoY changes of book-value-based leverage from Flow of Funds data for financial sectors property and life-insurance companies, securities brokers and dealers, commercial banks, households and non-financial corporate businesses	MEDFoF	Kollmann and Zeugner (2012)	–
Total borrowings of depository institutions from the Federal Reserve	BORROW	Korobilis (2009)	log-difference
Nonperforming Commercial Loans from Reports of Condition and Income for All Insured U.S. Commercial Banks	NPCMCM	–	–

Table 4.2: Financial indicators used in empirical application

4.5.2 Methods

For the sake of briefness, we concentrate on forecasting GDP in this exercise. Since, on the one hand, some of the evaluated variables are released only a few months after their reference period, while on the other hand the reduced form factor model approach performs particularly well at shorter-term horizons, we focus on forecasting GDP growth over four quarters (from time t to $t+4$), conditional on information available at time t. The focus of the exercise is to infer the predictive content of financial indicators via their statistical 'significance'. To that end, we concentrate on in-sample results from estimating on the entire sample 1988Q1-2010Q3 for the explanatory variables (respectively 1989Q1-2011Q3 for the response variable). In order to assess the robustness of the methods employed, we also conduct a (pseudo-)out-of-sample prediction exercise based on rolling samples of 40 quarters (which leaves an evaluation sample for the response variable of 2000Q1-2011Q3).

In view of the simulation results from the previous section, we concentrate the following models to the data for comparing their empirical results:[1]

- *DFM*: A simple dynamic factor model in the style of Geweke (1977) and oriented along Stock and Watson (2011). It is estimated with OLS, based on the first 5 principal components (at time t) out of the macroeconomic control data set, plus lagged annual GDP growth (LAGY, $y_t - y_{t-4}$, and most recent quarterly GDP growth (LAGQ, $y_t - y_{t-1}$).[2]

- *PCAR*: Of the PC-augmented regression type, this model adds the 11 financial indicators as regressors to model 'DFM'.

- *BMA*: A Bayesian model averaging framework[3] over both the financial indicators and the individual macroeconomic control variables (plus LAGQ and LAGY).[4] Since the resulting model space is too large for enumeration, we resort to MCMC approximation of the model space.[5]

- *full PC-MA*: The PC-MA framework in developed in section 4.3.2, with the macroeconomic controls as its sole control data set, and the focus variables being the financial indicators (as well as LAGQ and LAGY). The model prior framework conforms to restrictions (A) and (B) in section 4.3.2 and thus allows for the null model, with a prior model size of 5 (i.e. $\theta = 0.042$). For the estimation of individual models , 'full PC-MA' relies on Normal-Gamma Bayesian estimation with a hyper-g prior (Feldkircher and Zeugner, 2009).

- *PC-MA macro*: For comparison with the 'DFM' model, this framework imitates 'full PC-MA', but without the focus variables. I.e., the control data set encompasses all macroeconomic

[1] We did not consider the WALS estimator by Magnus et al. (2010) in the forecasting exercise, as the computational framework provided by the authors does not accommodate for a number of variables that is greater than the number of observations.

[2] The inclusion of these two 'lag' indicators results from BIC optimization of model DFM with varying quarterly and annual lags over the sample 1989Q1-1999Q4.

[3] We also considered Frequentist model averaging (FMA) of the S-BIC type, but rolling out-of-sample prediction of this framework were too far off the charts to be included in the figures. This results mainly from MCMC samplers being caught in local likelihood minima, as S-BIC seems to suffer from a 'supermodel effect' (Feldkircher and Zeugner, 2009). We do h, however, include FMA results in Table 4.8.

[4] The model prior used is of the type Ley and Steel (2009), with a prior model size of 5, and the coefficient prior was set to the hyper-g (UIP) version from Feldkircher and Zeugner (2009).

[5] Here, we use a simple Metropolis-Hastings algorithm (Fernández et al., 2001a) with 300,000 burn-ins and 3,000,000 subsequent model draws for both the in-sample estimation and rolling samples.

controls, but the set of 'focus variables is limited to LAGQ and LAGY. The prior set-up remains as in 'full PC-MA'.

- *PC-MA macrofin*: This framework differs from 'full PC-MA' in that it adds the financial indicators to the control data set rather than the focus variables (i.e., principal components are computed from 106+11 variables.)[1]

4.5.3 Results: Performance of models

Tables 4.3 and 4.4 report the coefficients resulting from the simple dynamic factor model 'DFM' and the variant augmented with financial indicators (model 'PCAR'). Regarding in-sample results, the adjusted R-squared of 'PCAR' versus 'DFM' seems to favor the inclusion of alternative financial indicators. In particular the variables advocated by Adrian and Shin (2010) (who favor them on grounds of in-sample regressions) appear to be significant when controlling for macroeconomic factors, along with the credit growth variables BORROW and BUSLOANS. However, the model 'PCAR' suffers from two major shortcomings in this respect: First, 'PCAR' was estimated with OLS on a short sample of 91 observations, which might lead to over-fitting. Second, the group of alternative predictors comprises 11 variables – it is thus not surprising that several of them turn out to be significant.[2] The suspicion of shortcomings substantiates when comparing the root mean squared errors (RMSE) of out-of-sample forecasts from both models. While 'DFM' performs well with an RMSE of 2.23 (percentage points of growth) over the sample 2000Q1-2011Q3, 'PCAR' is far worse with 4.89. Even when excluding the period since the Lehman brothers collapse (2008Q3-2011Q3), root mean squared errors suggest that a model of the kind of PCAR is not fit for forecasting.

We therefore turn to the model 'full PC-MA', in order to evaluate the predictive content of the various financial indicators. Table 4.5 reports the in-sample results over the sample 1989Q1-2011Q3. The posterior estimate of the number of included principal components[3] is 6.06, whereas the posterior estimate of included financial indicators is 3.19. This results contrasts with a weaker prior expectation of 0.46 included financial indicators and 4.14 principal components. There are, however, stark contrasts within the two sets of variables: The first four principal components receive a very high level of posterior support, while higher-order PCs have considerably smaller posterior inclusion probabilities (PIPs). Among the financial indicators two variables (BORROW and SBD_A) display very high PIPs, even though they are controlled for the PCs. Strikingly, the bulk of posterior mass is concentrated on a tiny set of models: The three models with highest posterior model probabilities together account for 29% of posterior mass and all include exactly four principal components, as well as the two variables BORROW and SBD_A. The question is in how far this effect stems from the selected sample. Figures 4.8 to 4.11 display the PIPs from the rolling samples used for (pseudo-)out-of-sample forecasting. The most striking feature is that the PIPs of the financial indicators seem quite volatile. In particular the PIPs for SBD_A (Figure 4.8) and BORROW (Figure 4.9) surge abruptly in response to the crisis, while they did not seem to matter much among the samples that include only pre-crisis data. The results from rolling sample estimations thus suggest that it is only the data from the crisis episode that lets these two variables seem so important.

[1] For the models 'full PC-MA', 'PC-MA macro' and 'PC-MA macrofin', we impose a maximum number of 23 principal components to average over to keep computational cost low in view of rolling samples (each model space with 13 'focus' variables and 23 PCs requires 196,608 model evaluations). For the evaluation sample 1999Q1-2010Q3, 23 principal components capture 92% of the control set's variance.

[2] Of course, the literature proposes several tests for inference about a set of indicators, e.g., White (2000). But the focus in this paper is on estimation methods, therefore we leave such tests to more detailed empirical research.

[3] This is equivalent to the sum of posterior inclusion probabilities for principal components.

Regarding the principal components, Figure 4.11 displays their PIPs over rolling samples. Almost by definition of the PC-MA frameworks, the first PC remains at a PIP of almost 100% throughout – the second to fourth PC, in contrast seem to start losing their posterior importance even before the onset of the crisis. All in all, this feature can be read as the PC-MA framework attributing less and less posterior mass to macroeconomic controls from mid-2005 on. The results from 'full PC-MA' thus suggest that financial predictors have become more important during the latter half of the 2000s, but it is not clear whether the results on BORROW and SBD_A will remain robust to future developments.

A look at Table 4.8 calls for further caution in this respect: The RMSE of 'full PC-MA' were somewhat larger than those of the simple 'DFM' during the first part of the 2000s. But when taking into account the crisis period, when financial indicators gained in posterior importance at the expense of macroeconomic factors, the RMSE of the PC-MA framework is far worse than that of 'DFM'. Figure 4.7 displays the rolling forecasts from model 'full PC-MA' and offers some insights into this result: The forecast seems to time the economic slump in late 2008 very well, but then overshoots throughout 2009-2010 and completely misses the subsequent recovery. Figure 4.8 shows why: For the rolling samples used for forecasting 2007-2010, the variable DRSREACBS (the mortgage delinquency rate) is the most important predictor (apart from the first macroeconomic PC). In it is in particular the evolution of this variable that is both responsible for the good performance of 'full PC-MA' in forecasting the initial crisis period and the vast overshooting thereafter.

The 'full PC-MA' frameworks thus offers insights on the importance of alternative financial predictors, but does not seem to be a forecasting framework that attains the performance of a simple dynamic factor model (at least during the crisis period). Although it allows for controlling through, and shifting posterior weight to, the factors from the 'DFM' models, it fails to do so properly and time. This behavior is mainly due to the large importance attributed to DRSREACBS in response to its performance in predicting 2007-2008. In order to compare whether the PC-MA framework is viable at all, consider the performance of the framework 'PC-MA macro', which does not include any financial variables. Its in-sample results over 1989Q1-2011Q3 (Table 4.6) seem to support the inclusion of many more PCs than under 'DFM'. Still, its RMSE performance is comparable to the 'DFM' model (Table 4.8). Since model 'PC-MA macro' is based on the same data as 'DFM', the only difference is between the two is how they take the data into account. In particular, model 'PC-MA macro' differs in its time varying shrinkage in terms of posterior inclusion probabilities. A look at its rolling forecasts (Figure 4.6) shows that they differ only slightly from the 'DFM' predictions, but apply much more shrinkage during the crisis episode, when its posterior expected number of included PCs is lower than during the rest of the sample (not displayed). In order to check whether financial indicators help at least in prediction when they are included in the set of macroeconomic controls to be regularized, consider the results for model 'PC-MA macrofin.'. Its out-of-sample RMSE does not differ much from either 'DFM' or 'PC-MA macro' (Table 4.8). Its rolling forecasts, displayed in Figure 4.6, do not appear fundamentally different from the other models, although they appear to perform marginally better during the crisis period. Finally, its in-sample results broadly appear similar to that of 'PC-MA macro' (Table 4.7).

4.5.4 Findings on individual financial indicators

When estimated on the entire sample 1989Q1-2011Q3, PC-MA finds a few financial indicators to be important predictors of GDP growth. The variables BORROW, SBD_A and DRSREACBS enter with negative coefficients, implying that GDP growth is negatively affected by increases in bank borrowing from the Fed, broker-dealers asset growth, and increases in the mortgage delinquency rate. Figure 4.13 displays the posterior densities for these three variables' standardized coefficients, conditional on inclusion.[1] All three coefficients seem to be safely in negative territory. However, as alluded to before, this result might be due to specificities of the sample. Figures 4.8 to 4.10 present the posterior inclusion probabilities for these indicators. All three mentioned variables display sizable PIPs only during the latter half of the rolling samples. Thus the posterior importance attributed to these indicators is mainly due to observations during the crisis period. Figure 4.12 (top panel) sheds more light on the evolution of coefficients: This box plot reports the distribution of (posterior expected values of) coefficients over the rolling samples. And indeed, while the coefficients of both SBD_A and BORROW are negative, they also remain close to zero for the vast majority of samples. In contrast, the (unconditional) coefficients for DRSREACBS and ANFCI are found to remain relatively far from zero for a greater percentage of rolling samples, which implies that they 'matter' for a large part of these samples. For other financial indicators, this notion of importance is less pronounced, largely due to their low PIPs. However, even if several indicators seem to 'matter' less, it is of interest whether they remain consistently above or below zero. The column 'conditional sign certainty' in Table 4.5 shows that apart from the indicators mentioned above, also the (in-sample) coefficients of indicators SHADBNK_A, MEDFoF, STED, and NPCMCM remain at one side of the origin for more than 95% of posterior model mass. Figure 4.12 (bottom panel) broadly confirms this impression, though it casts some doubt on ANFCI.

All in all, it seems that financial indicators could have assisted prediction during the 2000s, but considering them in the simple way pursued here would not have brought much improvement over the prediction performance of simple dynamic factor models. Moreover, a broad view of the results does not suggest that one financial indicator might be particularly adept for prediction purposes. However, the results do indicate consistent directions for indicators, e.g, a negative impact on real activity from mortgage delinquency rates, or leverage growth. Estimating PC-MA over the entire sample 1989Q1-2011Q3 seems to point to SBD_A and BORROW as important predictors. But rolling samples reveal that this result is mainly driven by the crisis period and its aftermath. In that sense, estimating PC-MA in-sample seems to have fallen prone to the same issue as the results reported in Adrian and Shin (2010). However, PC-MA clearly fares better in prediction than outright model averaging on the same data set, which seems to indicate that focus variable coefficients under 'full PC-MA' seem to suffer less from (omitted variable) bias under 'BMA' or 'PCAR'.

4.6 Conclusions

By means of a simulation analysis that is styled after typical model averaging (MA) applications, we find that popular MA techniques yield coefficient estimators that outperform principal-augmented regression in terms of robustness. In contrast, model averaging can be susceptible to robustness problems with respect to estimates and inference about the importance of variables. In response, we propose a framework that apply model averaging over 'focus' variables of interest and PC-based models. A fairly general simulation setting shows that our approach outperforms typical MA and factor-based techniques in terms of estimate robustness and inference. In particular, we demonstrate that a computationally cheap framework that applies model averaging over the number of included principal components can offer considerable robustness improvements. In addition, we propose a prior framework on the principal components that steers the prior concentration of variance in the controls and thus can improve convergence in large datasets. Note that this prior framework can be

[1] The posterior expected value of coefficients conditional on inclusion is equal to their unconditional coefficients (as provided in Table 4.5) divided by their posterior inclusion probabilities.

straightforwardly extended to a MA-based evaluation of competing predictor groups (as is typically the case in the context of growth regressions).

The proposed 'PC-MA' framework is particularly appropriate when evaluating competing alternative predictors conditional on a large set of established potential controls. For instance, an evaluation of recently proposed financial early warning indicators for the business cycle should be controlled for by established predictors. An empirical assessment of several such financial indicators finds evidence for a consistent relation between a handful of indicators and future real activity, but shows that they add little predictive content in comparison with the popular Stock and Watson (2002) dataset.

4.A Figures

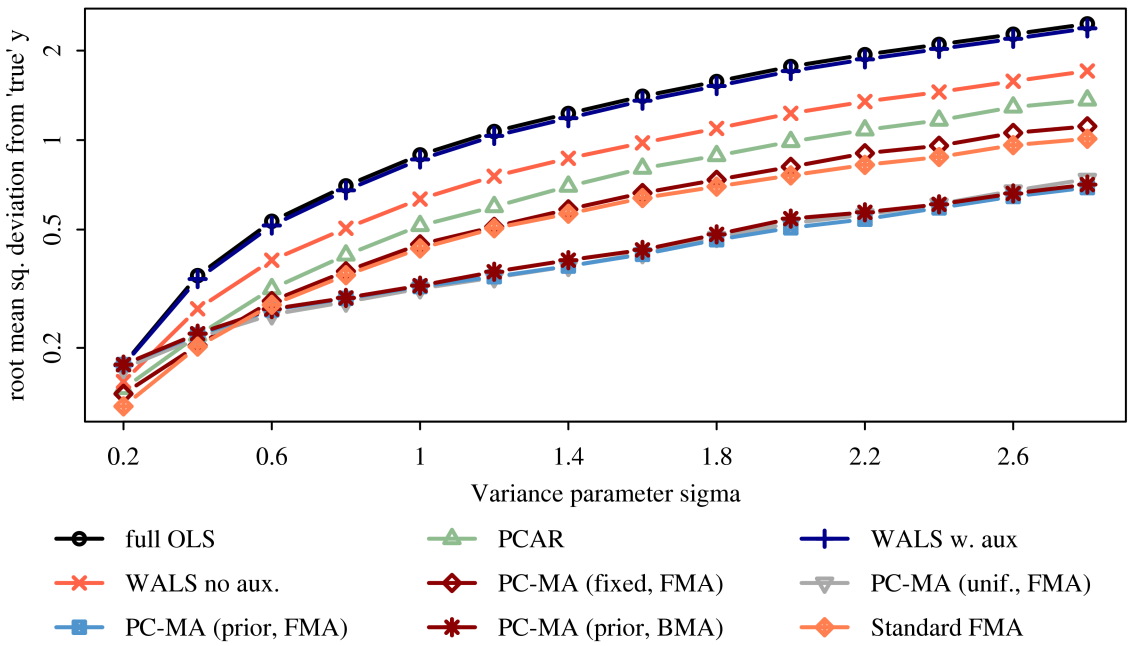


Figure 4.1: *Robustness of estimates:* Mean squared deviation of fitted values from the 'true' response variable according to the simulation set-up. Results are based on the simulations described in Section 4.4, with 50 simulation runs for each value of σ (as indicated on the abscissa).

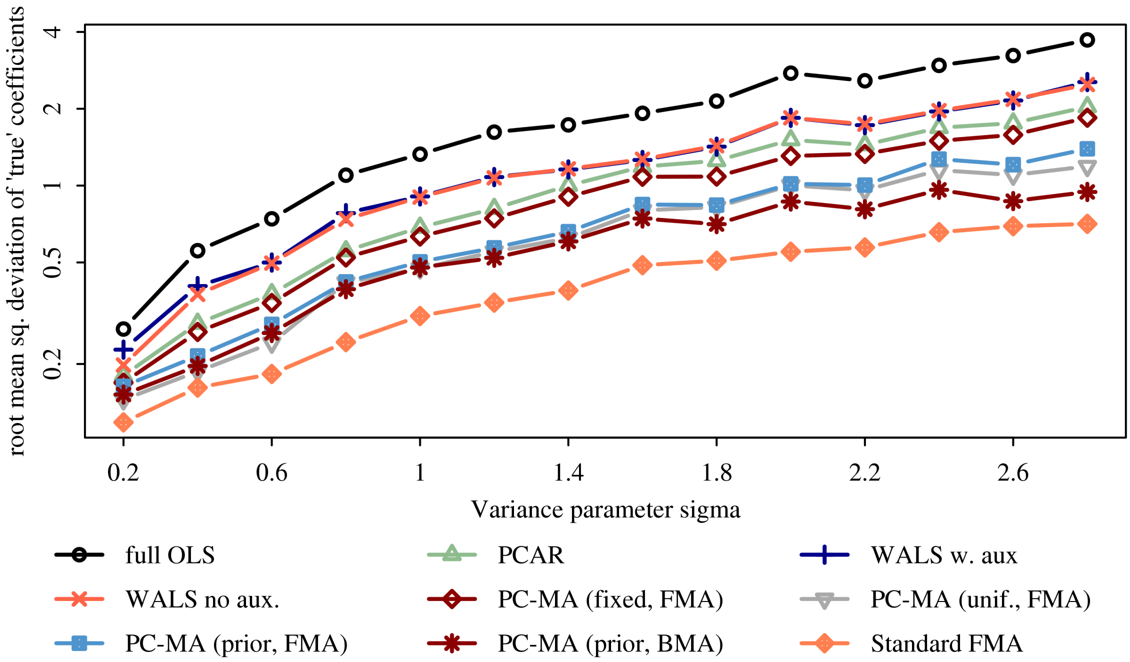


Figure 4.2: *Coefficient uncertainty for 'true' variables:* For focus variables that were included in the 'true' model, the Figure displays the mean squared deviation of coefficients from their values specified in the 'true' model. Results are based on the simulations described in Section 4.4, with 50 simulation runs for each value of σ (as indicated on the abscissa).

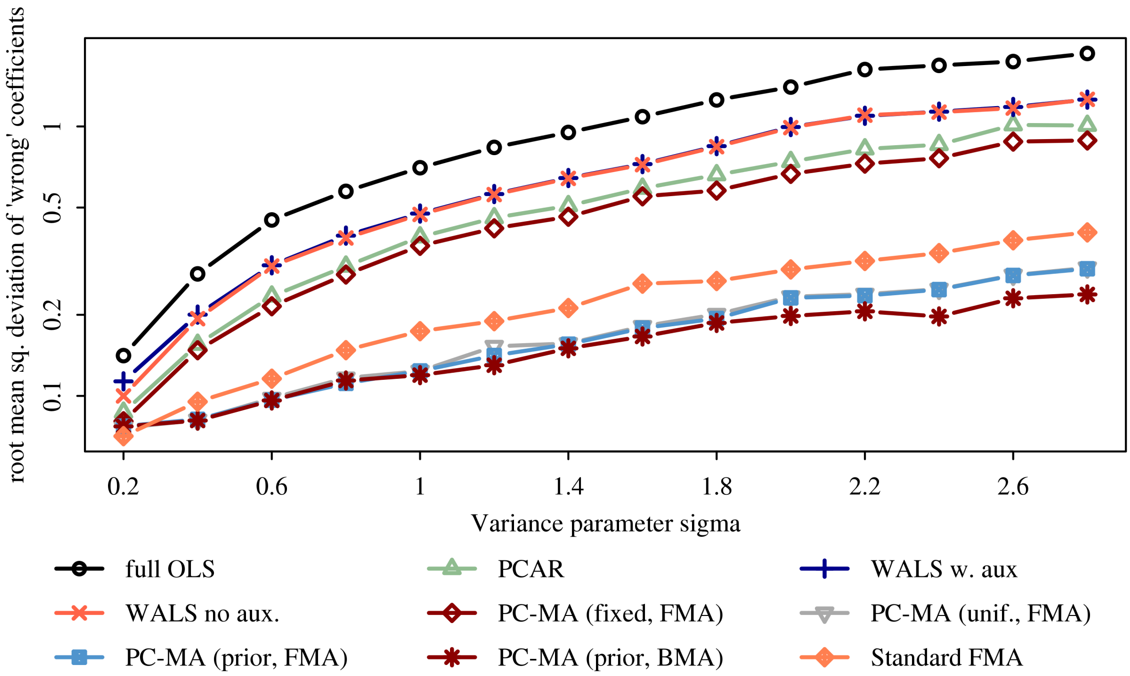


Figure 4.3: *Coefficient uncertainty for 'wrong' variables:* The Figure displays the mean squared deviation of coefficients for focus variables that were not included in the 'true' model. Results are based on the simulations described in Section 4.4, with 50 simulation runs for each value of σ (as indicated on the abscissa).

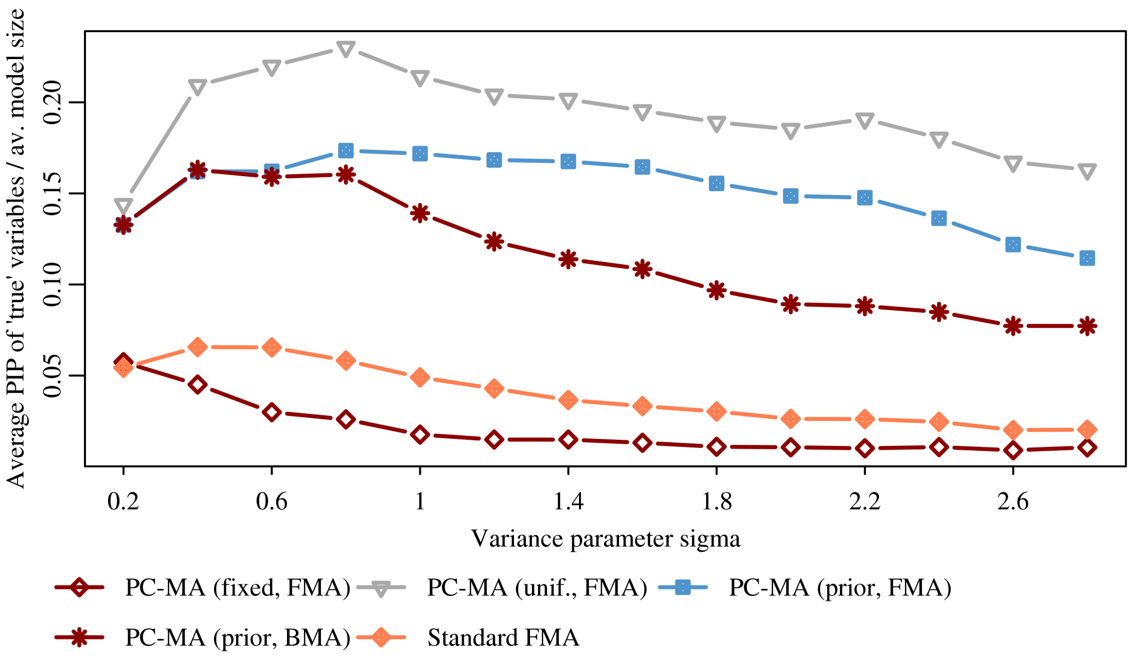


Figure 4.4: *Mean posterior inclusion probability of 'true' coefficients:* results are based on the simulations described in Section 4.4, with 50 simulation runs for each value of σ (as indicated on the abscissa).

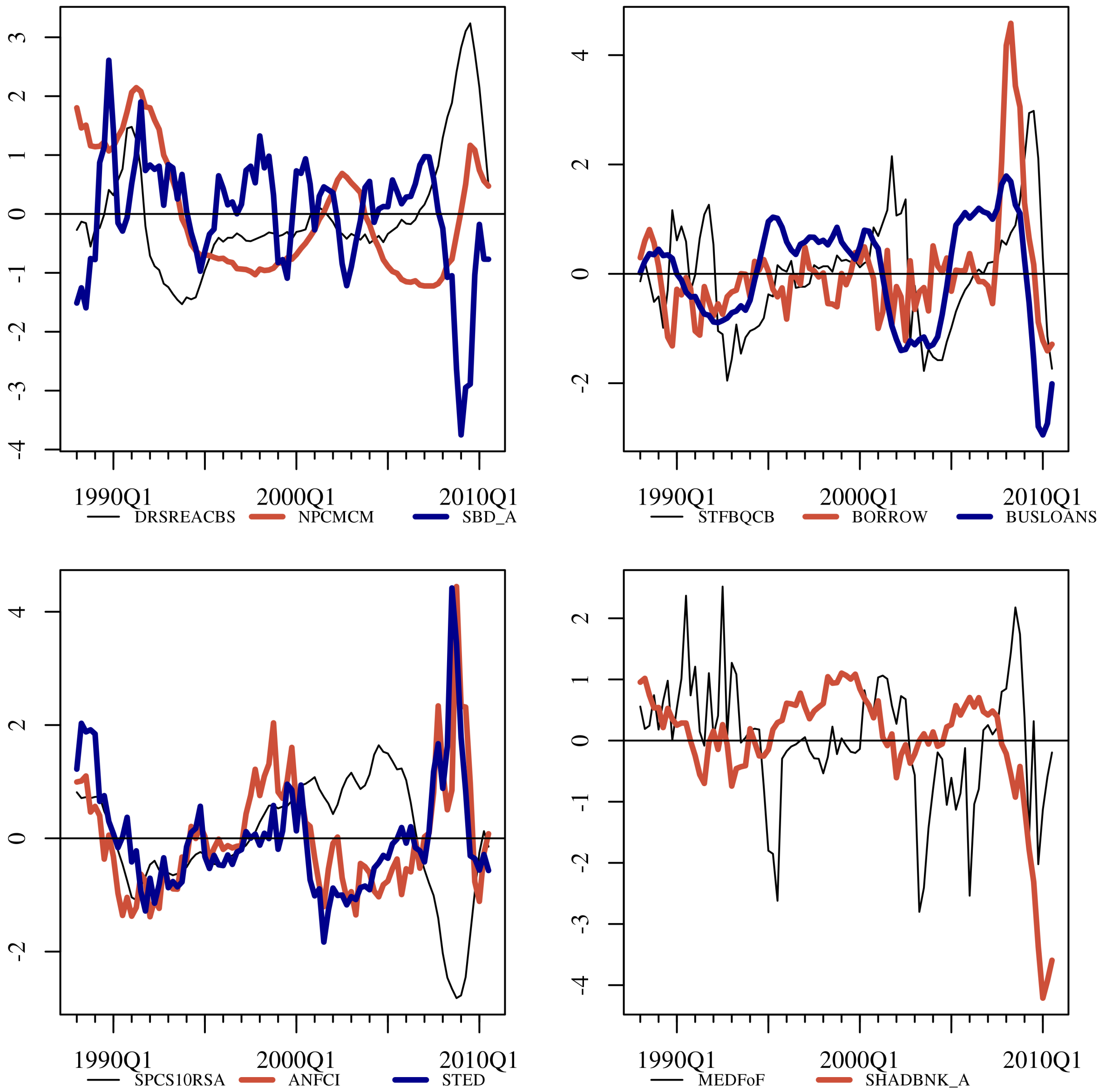


Figure 4.5: *Time series plots of alternative financial indicators*, in standardized (z-scores) form after transformation according to Table 4.2.

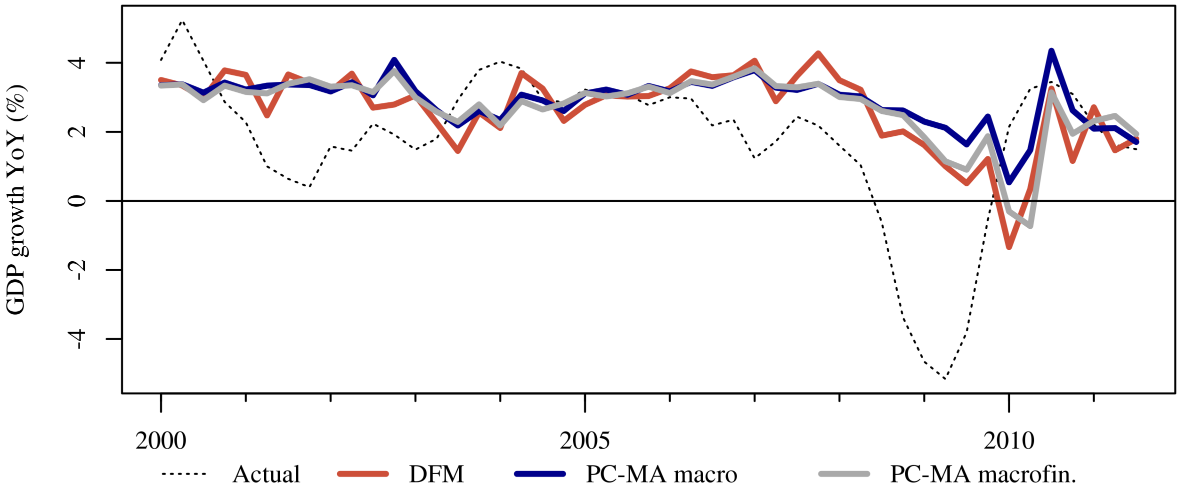


Figure 4.6: *Pure factor model forecasts:* Four quarter-ahead forecasts of annual GDP growth, (pseudo-) out of sample, based on rolling estimation samples of 40 quarters. 'Actual' denotes realized GDP growth. 'DFM' displays its forecasts based on a Stock-Watson-type dynamic factor model with 5 principal components, and GDP growth lags (annual and quarterly). 'PC-MA macro' denotes a PC-MA model (Bayesian estimation as in Table 4.6) that is based on the macroeconomic control variables that underlie the DFM. 'PC-MA macrofin.' is a similar to 'PC-MA macro', but with the control data set (on which principal components are based) extended with alternative financial indicators (Table 4.2)

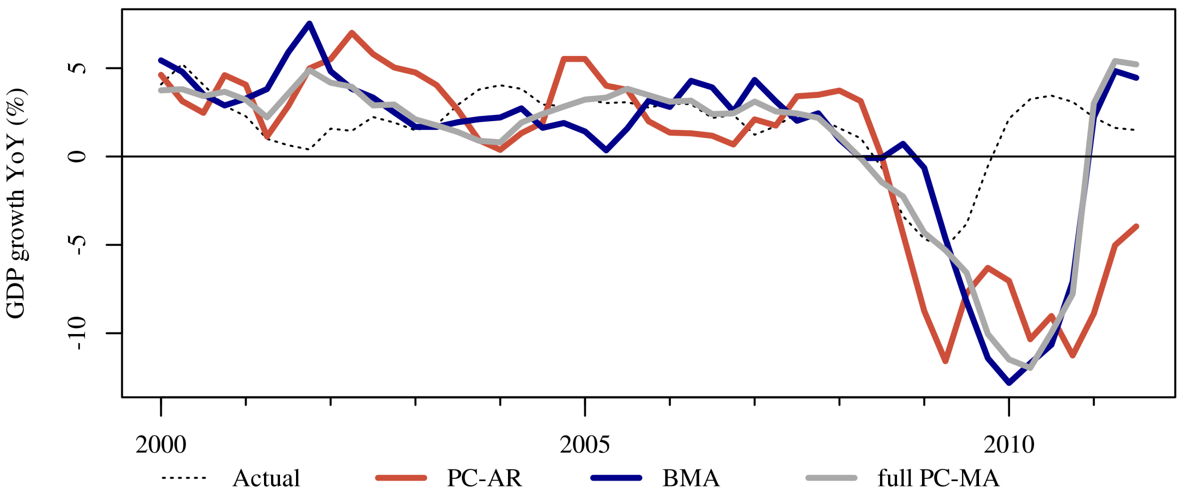


Figure 4.7: *Alternative financial indicators forecasts:* Four quarter-ahead forecasts of annual GDP growth, (pseudo-) out of sample, based on rolling estimation samples of 40 quarters. 'Actual' denotes realized GDP growth. 'PCAR' displays its forecasts based on a Stock-Watson-type dynamic factor model with 5 principal components and GDP growth lags (annual and quarterly) that includes the financial indicators as additional regressors. 'BMA' denotes Bayesian model averaging over both financial indicators and individual variables. 'Full PC-MA' is the model from Table 4.5.

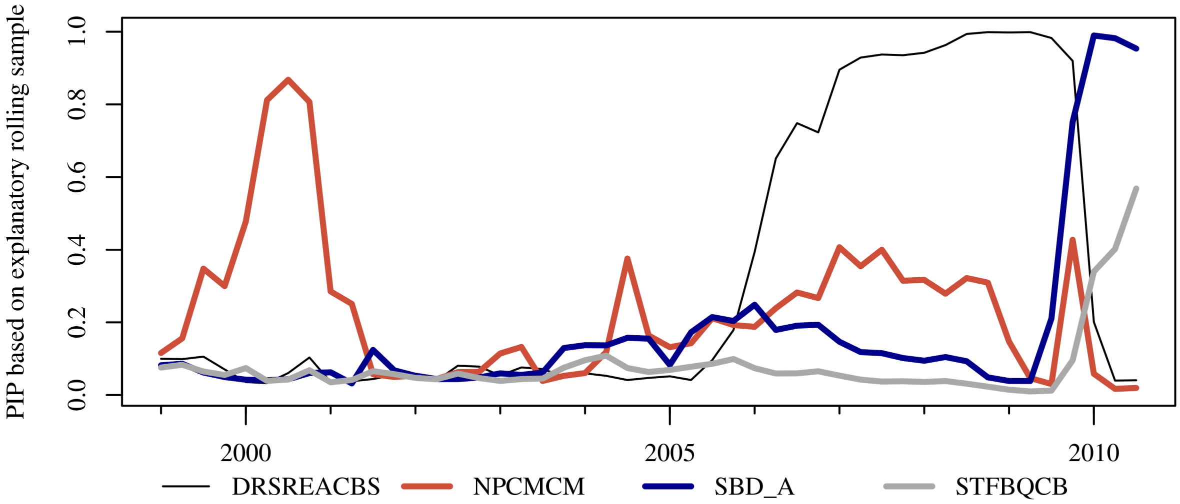


Figure 4.8: *Rolling posterior inclusion probabilities (PIPs) of financial indicators (1):* PIP of four financial indicators over rolling evaluation of the 'full PCMA' framework. Dates on the abscissa denote the end-point of each 40-quarter rolling sample used for estimating the PIPs.

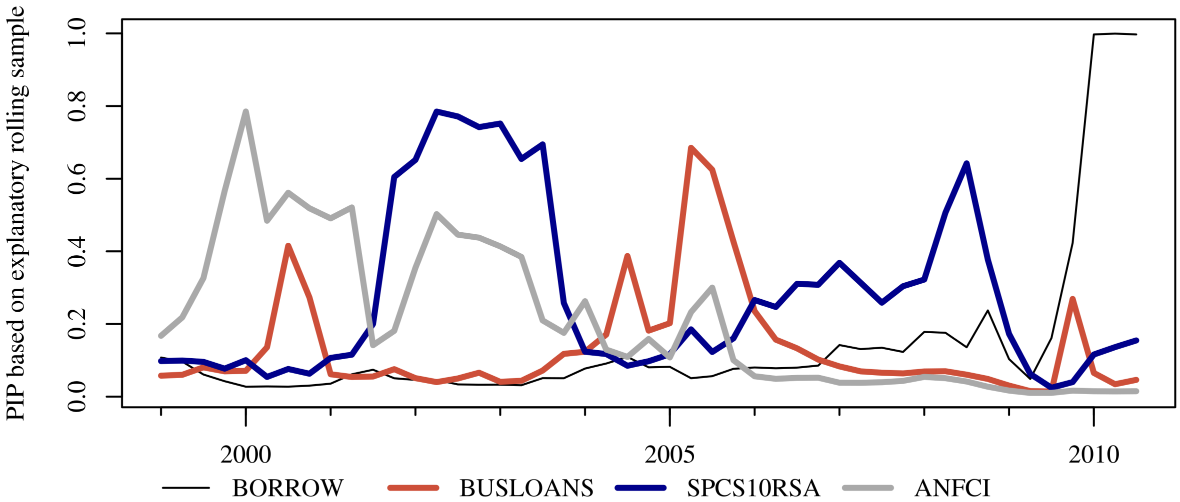


Figure 4.9: *Rolling posterior inclusion probabilities (PIPs) of financial indicators (2):* PIP of four financial indicators over rolling evaluation of the 'full PCMA' framework. Dates on the abscissa denote the end-point of each 40-quarter rolling sample used for estimating the PIPs.

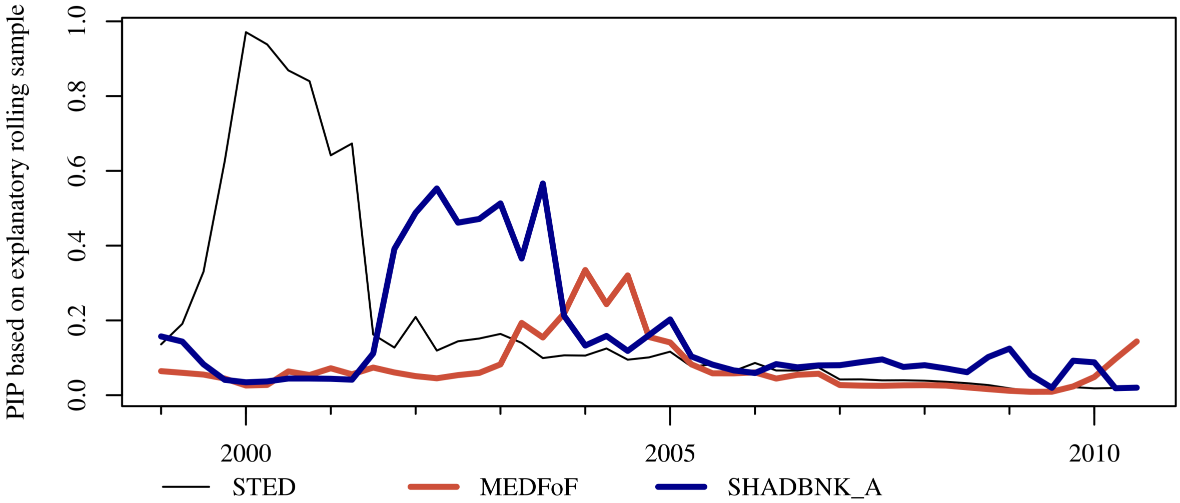


Figure 4.10: *Rolling posterior inclusion probabilities (PIPs) of financial indicators (3):* PIP of three financial indicators over rolling evaluation of the 'full PCMA' framework. Dates on the abscissa denote the end-point of each 40-quarter rolling sample used for estimating the PIPs.

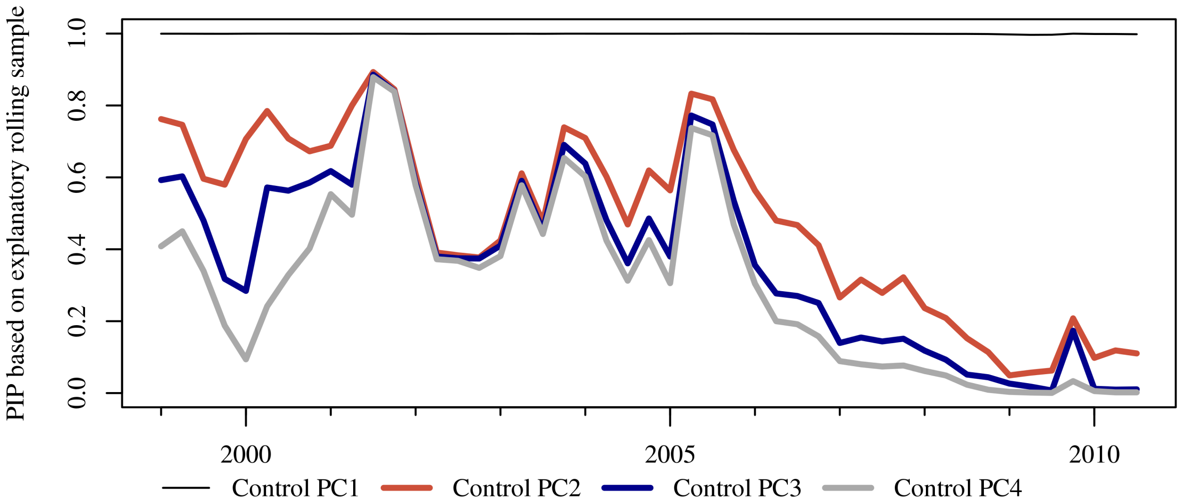


Figure 4.11: *Rolling posterior inclusion probabilities (PIPs) of macroeconomic principal components* from rolling evaluation of the 'full PCMA' framework. Dates on the abscissa denote the end-point of each 40-quarter rolling sample used for estimating the PIPs.

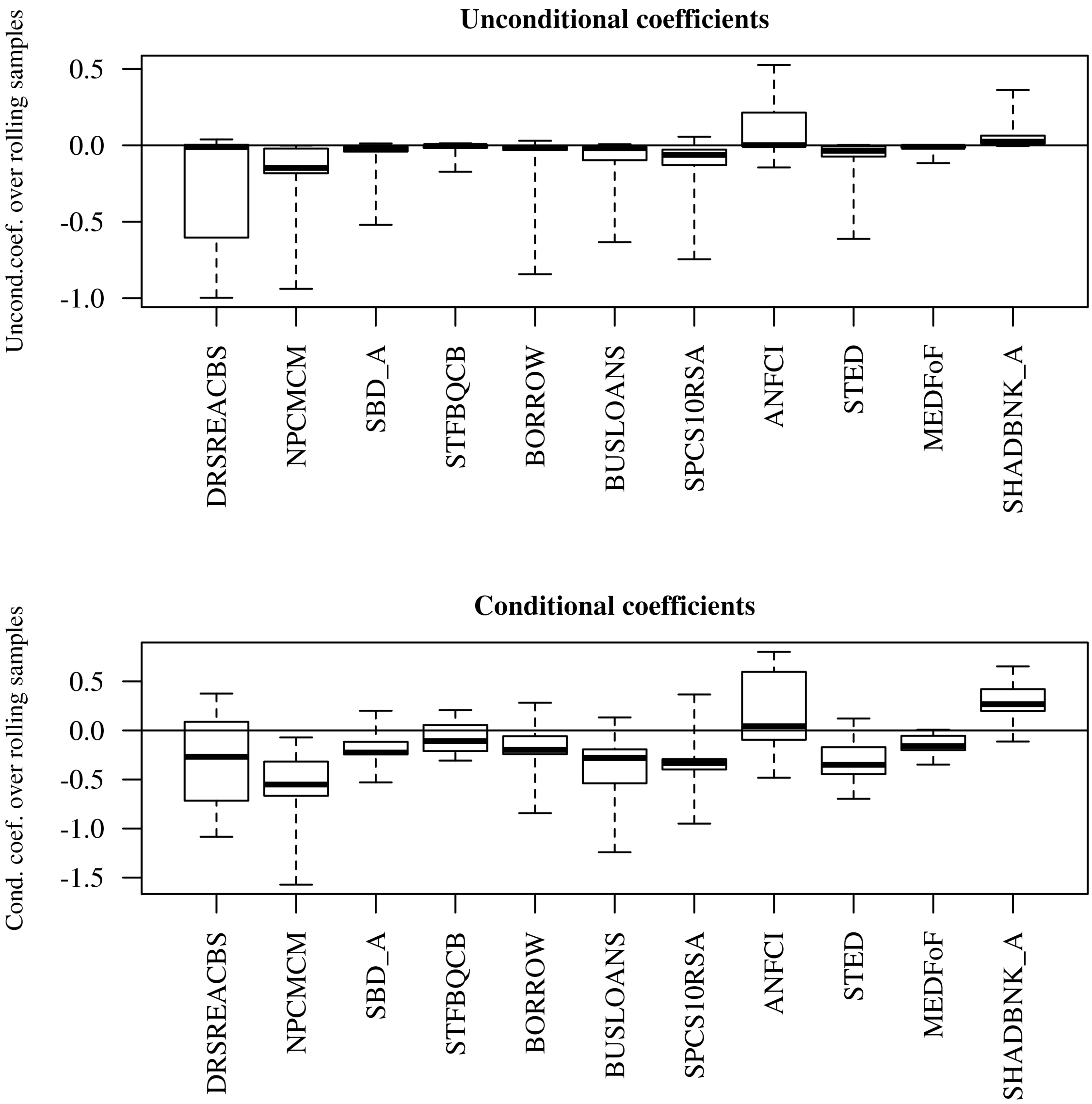


Figure 4.12: *Rolling coefficients of financial indicators:* Upper panel shows unconditional coefficients (averaged over all models), while lower panel shows coefficients conditional on inclusion (averaged over models where this coefficient is not restricted to zero). Both panels show posterior expected values of standardized coefficients of financial indicators over rolling evaluation of the 'full PCMA' framework (target variable evaluation sample: 2000Q1-2011Q3, rolling window: 40 quarters). Central bar in boxplot corresponds to median coefficient over entire sample, the box bounds to the 25th and 75th percentile, while whiskers display minimum and maximum coefficient over the entire sample.

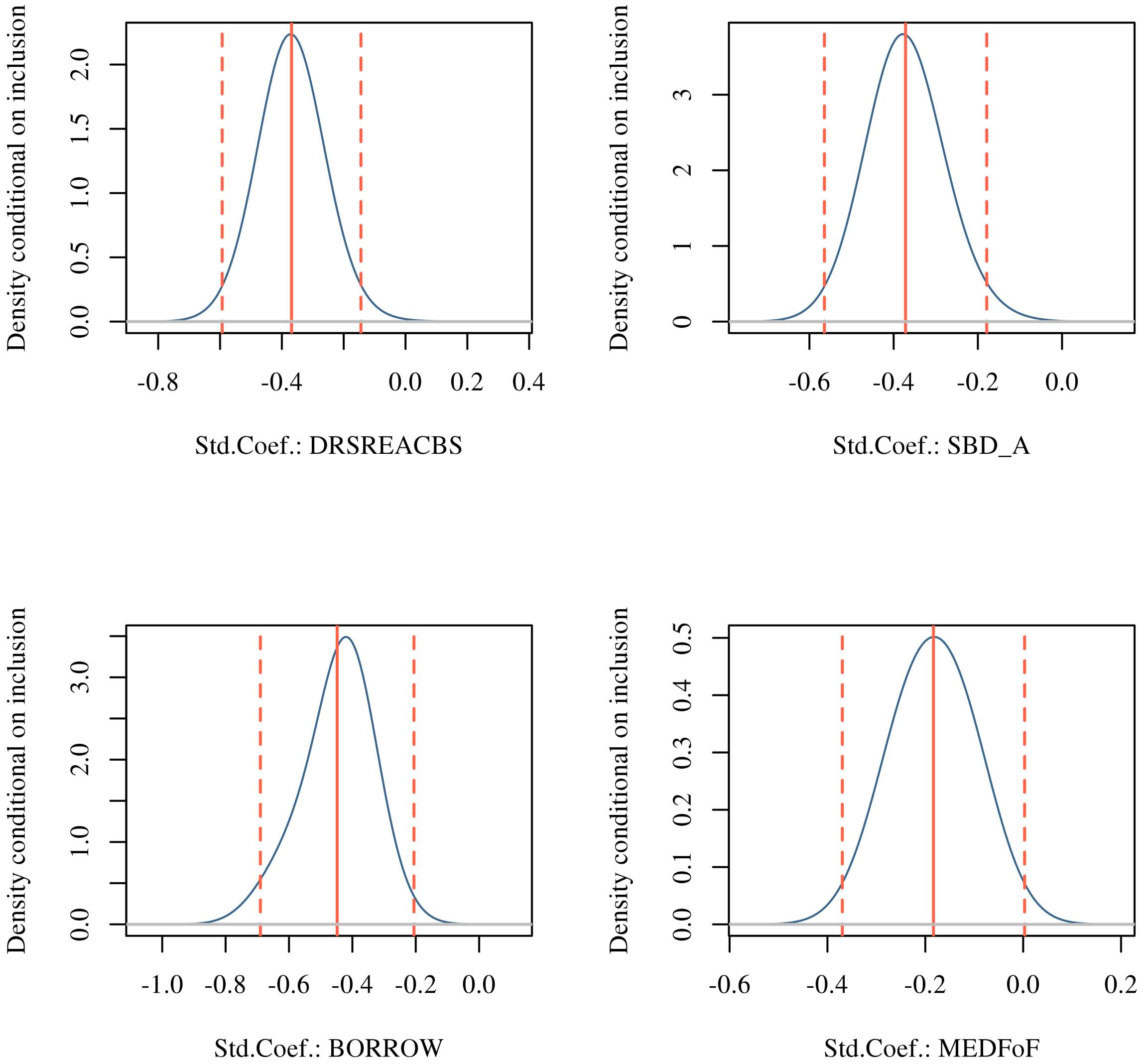


Figure 4.13: *Posterior density of coefficients four financial indicators* from 'full PC-MA', in-sample: Mean (conditional) posterior standardized coefficients of financial indicators from in-sample estimation of the 'full PCMA' framework over sample 1989Q1-2011Q3. Solid red line corresponds to the posterior expected value of standardized coefficients, conditional on inclusion of the variable. Dotted bounds correspond to two times the posterior expected value plus/minus two times the posterior standard deviation of the standardized coefficient density.

4.B Tables

Table 4.3: Dynamic factor model, in-sample results ('DFM')

	Std.Coef.	t-statistic	p-value
(Intercept)	0.03	5.16	0.00
LAGY	0.25	1.13	0.26
LAGQ	-0.46	-2.37	0.02
Control PC1	-0.55	-2.95	0.00
Control PC2	0.27	2.11	0.04
Control PC3	0.27	2.58	0.01
Control PC4	0.56	5.66	0.00
Control PC5	0.11	1.28	0.20
Adj. R2-squared	0.38		

Notes: Estimates from a dynamic factor model on sample 1989Q1-2011Q3, estimating GDP growth for four quarters ahead (in-sample) via OLS. Principal components are denoted as 'Control PC', with a number conforming to their rank as principal components of a macroeconomic control data set. Variables 'LAGY' corresponds to lagged annual GDP growth, while 'LAGQ' is most recent quarterly GDP growth. Column 'Std.Coef' denotes standardized coefficients, and 'p-value' the prob-value of the respective (ordinary) t-statistic.

Table 4.4: Dynamic factor model with alternative financial indicators, in-sample results ('PCAR')

	Std.Coef.	t-statistic	p-value
(Intercept)	0.05	6.61	0.00
LAGY	-0.38	-1.84	0.07
LAGQ	-0.28	-1.88	0.06
DRSREACBS	-0.24	-1.53	0.13
NPCMCM	-0.12	-1.08	0.28
SBD_A	-0.29	-3.08	0.00
STFBQCB	-0.07	-0.67	0.50
BORROW	-0.38	-3.32	0.00
BUSLOANS	-0.52	-2.44	0.02
SPCS10RSA	-0.06	-0.51	0.61
ANFCI	0.31	2.85	0.01
STED	-0.12	-0.75	0.45
MEDFoF	-0.09	-1.04	0.30
SHADBNK_A	0.44	3.04	0.00
Control PC1	-0.65	-3.74	0.00
Control PC2	-0.06	-0.49	0.62
Control PC3	0.01	0.06	0.95
Control PC4	0.36	3.50	0.00
Control PC5	0.06	0.71	0.48
Adj. R2-squared	0.68		

Notes: Estimates from a dynamic factor model augmented with additional financial regressors, estimating GDP growth for four quarters ahead (in-sample) via OLS on sample 1989Q1-2011Q3. Principal components are denoted as 'Control PC', with a number conforming to their rank as principal components of a macroeconomic control data set. Variables 'LAGY' corresponds to lagged annual GDP growth, while 'LAGQ' is most recent quarterly GDP growth. The remaining variables are financial indicators (cf. Table 4.2 for an overview). Column 'Std.Coef' denotes standardized coefficients, and 'p-value' the prob-value of the respective (ordinary) t-statistic.

Table 4.5: PC-MA with alternative financial indicators, in-sample results ('full PC-MA')

Variable	Post.Incl.Prob.	Coef (Post.exp.val.)	Coef. St.Dev.	Cond.Sign Certainty (+)
Control PC1	1.00	-0.25	0.16	0.00
BORROW	0.98	-0.44	0.14	0.00
SBD_A	0.88	-0.32	0.15	0.00
Control PC2	0.87	0.05	0.10	0.81
Control PC3	0.86	0.10	0.10	0.97
Control PC4	0.85	0.26	0.14	1.00
DRSREACBS	0.59	-0.22	0.20	0.00
Control PC5	0.35	0.03	0.07	1.00
SHADBNK_A	0.27	0.07	0.14	1.00
Control PC6	0.21	0.02	0.05	0.99
LAGQ	0.19	-0.06	0.14	0.01
Control PC7	0.17	-0.02	0.05	0.00
Control PC8	0.14	0.00	0.03	0.76
Control PC9	0.14	0.00	0.02	0.82
Control PC10	0.14	0.00	0.03	0.85
Control PC11	0.14	-0.01	0.03	0.08
Control PC12	0.14	0.03	0.08	1.00
MEDFoF	0.14	-0.03	0.07	0.00
Control PC13	0.13	-0.00	0.02	0.09
Control PC14	0.13	-0.00	0.02	0.00
Control PC15	0.13	0.01	0.03	1.00
Control PC16	0.13	0.03	0.07	1.00
Control PC17	0.13	-0.02	0.07	0.00
ANFCI	0.10	0.02	0.07	0.98
STED	0.08	-0.02	0.07	0.01
Control PC18	0.08	-0.00	0.02	0.01
LAGY	0.07	-0.02	0.09	0.01
Control PC19	0.07	-0.00	0.02	0.00
Control PC20	0.07	0.00	0.02	1.00
Control PC21	0.07	0.01	0.03	1.00
Control PC22	0.06	-0.00	0.01	0.13
Control PC23	0.06	-0.01	0.04	0.00
SPCS10RSA	0.04	0.00	0.04	0.79
BUSLOANS	0.04	-0.01	0.05	0.30
STFBQCB	0.04	-0.00	0.03	0.26
NPCMCM	0.03	-0.00	0.02	0.04

Notes: Posterior estimates from PC-MA on sample 1989Q1-2011Q3, estimating GDP growth for four quarters ahead (in-sample), based on macroeconomic control indicators, lags and alternative financial indicators. Fully Bayesian estimation was used, with a PCMA-group prior (see Section 3) as model prior, and the hyper-g prior framework (Feldkircher and Zeugner, 2009) as coefficient prior. Variables are divided into 'Control PC', i.e. ranked principal components based on a control data set. Variables 'LAGY' corresponds to lagged annual GDP growth, while 'LAGQ' is most recent quarterly GDP growth. The remaining variables are financial indicators (cf. Table 4.2 for an overview). '*Post.Incl.Prob.*' denotes the posterior inclusion probability (PIP) per variable. '*Coef (Post.exp.val)*' details the posterior expected value ('estimate') for standardized coefficients. Note that these coefficients are 'unconditional', i.e. averaged over models that include and do not include the variable.[1] '*Coef. St.Dev.*' displays the unconditional posterior standard deviation of these coefficients. '*Cond.Sign Certainty (+)*' denotes the share of models where the coefficient was found to be positive, conditional on inclusion (i.e., sign certainty for a negative coefficient corresponds to one minus this indicator).

Table 4.6: PC-MA with just macroeconomic control factors, in-sample results ('PC-MA macro')

Variable	Post.Incl.Prob.	Coef (Post.exp.val.)	Coef. St.Dev.	Cond.Sign Certainty (+)
Control PC1	1.00	-0.51	0.20	0.00
Control PC2	1.00	0.09	0.10	0.80
Control PC3	1.00	0.14	0.08	1.00
Control PC4	1.00	0.42	0.09	1.00
Control PC5	0.97	0.10	0.07	1.00
Control PC6	0.96	0.17	0.08	1.00
Control PC7	0.92	-0.06	0.08	0.01
Control PC8	0.91	0.07	0.08	1.00
Control PC9	0.91	0.01	0.06	1.00
Control PC10	0.91	0.00	0.07	0.37
Control PC11	0.91	-0.02	0.06	0.11
Control PC12	0.91	0.22	0.10	1.00
Control PC13	0.87	-0.01	0.06	0.35
Control PC14	0.87	-0.02	0.06	0.00
Control PC15	0.87	0.04	0.06	1.00
Control PC16	0.87	0.16	0.09	1.00
Control PC17	0.86	-0.17	0.09	0.00
Control PC18	0.56	-0.03	0.05	0.00
Control PC19	0.49	-0.03	0.05	0.00
Control PC20	0.44	0.01	0.05	1.00
Control PC21	0.42	0.05	0.08	1.00
LAGQ	0.39	-0.14	0.22	0.00
Control PC22	0.34	0.00	0.04	0.60
Control PC23	0.32	-0.02	0.05	0.00
LAGY	0.27	-0.08	0.17	0.04

Notes: Posterior estimates from PC-MA on sample 1989Q1-2011Q3, estimating GDP growth for four quarters ahead (in-sample), based on macroeconomic control indicators and lags (framework 'PC-MA macro'). Fully Bayesian estimation was used, with a PCMA-group prior (see Section 3) as model prior, and the hyper-g prior framework (Feldkircher and Zeugner, 2009) as coefficient prior. Variables are divided into 'Control PC', i.e. ranked principal components based on a control data set. Variables 'LAGY' corresponds to lagged annual GDP growth, while 'LAGQ' is most recent quarterly GDP growth. '*Post.Incl.Prob.*' denotes the posterior inclusion probability (PIP) per variable. '*Coef (Post.exp.val)*' details the posterior expected value ('estimate') for standardized coefficients. Note that these coefficients are 'unconditional', i.e. averaged over models that include and do not include the variable.[1] '*Coef. St.Dev.*' displays the unconditional posterior standard deviation of these coefficients. '*Cond.Sign Certainty (+)*' denotes the share of models where the coefficient was found to be positive, conditional on inclusion (i.e., sign certainty for a negative coefficient corresponds to one minus this indicator).

Table 4.7: PC-MA with principal components extracted from macroeconomic controls and financial indicators, in-sample results ('PC-MA macrofin'

Variable	Post.Incl.Prob.	Coef (Post.exp.val.)	Coef. St.Dev.	Cond.Sign Certainty (+)
Control PC1	1.00	0.61	0.19	1.00
Control PC2	1.00	0.09	0.11	0.97
Control PC3	1.00	0.09	0.07	1.00
Control PC4	1.00	0.41	0.07	1.00
Control PC5	1.00	-0.24	0.07	0.00
Control PC6	1.00	0.22	0.06	1.00
Control PC7	0.99	0.03	0.06	1.00
Control PC8	0.99	-0.06	0.07	0.00
Control PC9	0.99	0.07	0.07	1.00
Control PC10	0.99	0.01	0.07	0.59
Control PC11	0.99	-0.05	0.06	0.00
Control PC12	0.99	-0.16	0.06	0.00
Control PC13	0.99	-0.26	0.07	0.00
Control PC14	0.97	0.04	0.06	1.00
Control PC15	0.97	-0.18	0.07	0.00
Control PC16	0.94	0.12	0.07	1.00
Control PC17	0.90	0.13	0.07	1.00
LAGY	0.58	-0.18	0.21	0.01
LAGQ	0.45	-0.12	0.19	0.01
Control PC18	0.39	0.00	0.04	0.61
Control PC19	0.26	0.01	0.04	1.00
Control PC20	0.20	-0.01	0.03	0.00
Control PC21	0.19	0.02	0.06	1.00
Control PC22	0.07	-0.00	0.02	0.00
Control PC23	0.06	0.00	0.01	1.00
Control PC24	0.05	0.00	0.01	1.00
Control PC25	0.05	-0.00	0.02	0.00
Control PC26	0.05	-0.01	0.03	0.00
Control PC27	0.05	-0.00	0.01	0.00
Control PC28	0.04	-0.00	0.03	0.00
Control PC29	0.03	0.00	0.02	1.00
Control PC30	0.02	-0.00	0.01	0.00
Control PC31	0.01	-0.00	0.01	0.00
Control PC32	0.00	-0.00	0.00	0.00
Control PC33	0.00	-0.00	0.00	0.00
Control PC34	0.00	-0.00	0.00	0.47

Notes: Posterior estimates from PC-MA on sample 1989Q1-2011Q3, estimating GDP growth for four quarters ahead (in-sample), based on macroeconomic control indicators and lags (framework 'PC-MA macrofin.'). Fully Bayesian estimation was used, with a PCMA-group prior (see Section 3) as model prior, and the hyper-g prior framework (Feldkircher and Zeugner, 2009) as coefficient prior. Variables are divided into 'Control PC', i.e. ranked principal components based on a control data set. Variables 'LAGY' corresponds to lagged annual GDP growth, while 'LAGQ' is most recent quarterly GDP growth. '*Post.Incl.Prob.*' denotes the posterior inclusion probability (PIP) per variable. '*Coef (Post.exp.val)*' details the posterior expected value ('estimate') for standardized coefficients. Note that these coefficients are 'unconditional', i.e. averaged over models that include and do not include the variable.[1] '*Coef. St.Dev.*' displays the unconditional posterior standard deviation of these coefficients. '*Cond.Sign Certainty (+)*' denotes the share of models where the coefficient was found to be positive, conditional on inclusion (i.e., sign certainty for a negative coefficient corresponds to one minus this indicator).

RMSE over sample:	DFM	PC-MA macro	PC-MA macrofin.	PCAR	BMA	full PC-MA	FMA
2000Q1-2008Q2	1.488	1.426	1.418	2.356	2.097	1.546	9.813
2000Q1-2011Q3	2.233	2.364	2.294	4.891	4.814	4.437	13.324

Table 4.8: Root mean squared errors for rolling out-of sample forecasts of annual GDP growth over 2000Q1-2011Q3. For model definitions, cf. descriptions of Figures 4.6 and 4.7. Model 'FMA' is the frequentist model averaging (S-BIC) equivalent of 'BMA'.

4.C Technical Appendix

Generalizing the PC-MA prior to size-based model priors

The adjusted PC-prior proposed in section 4.3.2 may be combined with any model prior on the focus variables, and reconciled with requirement (A) from section 4.3.2 that defines the prior importance of a variable group with respect to all other variable groups and the null model. Let $f(\mathcal{M}_\gamma)$ denote the model prior for model $\mathcal{M}_\gamma$ with $\gamma \in 2^K,\ K = \sum_l K_{C,l}$ under outright model averaging. Moreover , let $\mathcal{M}_{\gamma \notin l}$ denote models that do not include any variable out of group l. Then, the conceptual equivalent of requirement (A) is

$$\Theta_l = (1 - \sum_{\mathcal{M}_{\gamma \notin l}} f(\mathcal{M}_{\gamma \notin l}))$$

For a general model prior, this restriction might be numerically difficult to evaluate, but it is readily available in case of model priors that only depend on model size k, such as the one proposed by Ley and Steel (2009), or the binomial model prior evoked in section 4.3.2. Denote this model-size prior by $f(k_\gamma, \sum_l K_{C,l})$, only dependent on model $\mathcal{M}_\gamma$'s parameter size k_γ and the total number of variables. Then, restriction (A) boils down to:

$$\Theta_l = 1 - \sum_{m=0}^{K-K_l} \binom{K_l}{m} f(m, K)$$

Note that restriction (A) generally implies a shifting/skewing of the prior model size distribution, but it leaves the relative prior weights unchanged among models that only include variables from singleton groups (i.e. only focus variables). The expression for overall prior model size (11) holds under general size-based priors. The prior expected number of PCs of group l to be included is defined as $E(k_{\Lambda,l}) = \Theta_l \frac{\sum_{i=1}^{r_l} \lambda_{i,l}^{\alpha_l}}{\lambda_{1,l}^{\alpha_l}}$. Note that this number is isomorphic to the hyperparameter α_l. This result obtains since $\lim_{\alpha_l \to \infty} \frac{\lambda_{1,l}^{\alpha_l} - \lambda_{1,2}^{\alpha_l}}{\lambda_{1,l}^{\alpha_l}} = 1$ and $\lim_{\alpha_l \to 0} \frac{\lambda_{r,l}^{\alpha_l}}{\lambda_{1,l}^{\alpha_l}} = 1$, as well as $\frac{\partial E(k_{\Lambda,l})}{\partial \alpha_l} < 0 \forall a$. While there is no closed-form solution for the inverse function $\alpha_l(E(k_{\Lambda,l}))$, it may be nonetheless advisable to specify the degree of concentration in terms of the prior expected number of included PCs $E(k_{\Lambda,l})$.

References

Abramowitz, M. and Stegun, I. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, volume 10. National Bureau of Standards, Applied Mathematics Series. 10, 22

Adrian, T., Moench, E., and Shin, H. (2010). Financial intermediation, asset prices, and macroeconomic dynamics. Staff Reports 422, Federal Reserve Bank of New York. 55

Adrian, T. and Shin, H. (2010). Financial intermediaries and monetary economics, in: Friedman, B. In Friedman, B. and Woodford, M., editors, *Handbook of monetary economics*, volume 3. Elsevier, Amsterdam. 55, 107, 110, 112, 114

Aikman, D. and Paustian, M. (2006). Bank capital, asset prices and monetary policy. Working Paper 305, Bank of England. 55, 64

Alessi, L. and Detken, C. (2011). Quasi real time early warning indicators for costly asset price boom/bust cycles: A role for global liquidity. *European Journal of Political Economy*, 27(3):520–533. 101

Altunbas, A., Gambacorta, L., and Marques-Ibanzes, D. (2010). An empirical assessment of the risk-taking channel. mimeo. 108, 110

Angeloni, I. and Faia, E. (2009). A tale of two policies: prudential regulation and monetary policy with fragile banks. Working paper, University of Frankfurt. 55

Antipa, P., Mengus, E., and Mojon, B. (2010). Would macroprudential policies have prevented the great recession? Working paper, Banque de France. 55

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221. 92

Barbieri, M. M. and Berger, J. O. (2003). Optimal Predictive Model Selection. *The Annals of Statistics*, 32(3):870–897. 18

Barrell, R., Davis, P., Karim, D., and Liadze, I. (2010). Bank regulation, property prices and early warning systems for banking crises in OECD countries. Discussion Paper 330, NIESR. 108, 110

Bean, C., Larsen, J., and Nikolov, K. (2002). Financial Frictions and the Monetary Transmission Mechanism: Theory, Evidence and Policy Implications. Working Paper 113, ECB. 108

Bernanke, B. (1993). Credit in the Macroeconomy. *Federal Reserve Bank of New York Quarterly Review*, 18:50–58. 108

Brown, P., Vannucci, M., and Fearn, T. (1998). Multivariate Bayesian Variable Selection and Prediction. *Journal of the Royal Statistical Society B*, 60:627–641. 5

Chipman, H., George, E., and McCulloch, R. (2001). *The Practical Implementation of Bayesian Model Selection*. Institute of Mathematical Statistics Lecture Notes-Monograph Series Vol. 38. Beachwood, Ohio. 4

Ciccone, A. and Jarociński, M. (2010). Determinants of Economic Growth: Will Data Tell? *American Economic Journal: Macroeconomics*, 2:222–46. 3, 6, 17, 18, 44, 45, 46, 47, 48, 49, 92

Claeskens, G. and Hjort, N. (2008). *Model Selection and Model Averaging*. Cambridge University Press. 6, 94, 95, 103, 104

Clark, T. and West, K. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138:291–311. 61

Clyde, M., Desimone, H., and Parmigiani, G. (1996). Prediction Via Orthogonalized Model Mixing. *Journal of the American Statistical Association*, 91(435). 93

Clyde, M., Ghosh, J., and Littman, M. (2010). Bayesian adaptive sampling for variable selection and model averaging. *Journal of Computational and Graphical Statistics*, 20(1):80–101. 92

Crespo Cuaresma, J. (2010). How different is Africa? A comment on Masanjala and Papageorgiou (2008). *Journal of Applied Econometrics*, 26(6):1048–1050. 45

Crespo Cuaresma, J. and Doppelhofer, G. (2007). Nonlinearities in Cross-Country Growth Regressions: A Bayesian Averaging of Thresholds (BAT) Approach. *Journal of Macroeconomics*, 29:541–554. 17

Cui, W. and George, E. (2008). Empirical Bayes vs. fully Bayes variable selection. *Journal of Statistical Planning and Inference*, 138:4:888–900. 2

Cúrdia, V. and Woodford, M. (2009). Credit frictions and optimal monetary policy. Working paper, New York Fed and Columbia University. 55

De Mol, C., Giannone, D., and Reichlin, L. (2008). Forecasting using a large number of predictors: Is bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, 146(2):318–328. 93, 105, 107

De Walque, G., Pierrard, O., and Rouabah, A. (2010). Financial (in) stability, supervision and liquidity injections: A dynamic general equilibrium approach. *The Economic Journal*, 120(549):1234–1261. 55

Devereux, M. and Yetman, J. (2010). Leverage constraints and the international transmission of shocks. *Journal of Money, Credit and Banking*, 42:71–105. 54, 64

Dib, A. (2010). Banks, credit market frictions, and business cycles. Working Paper 2010-24, Bank of Canada. 55

Eicher, T. S., Papageorgiou, C., and Raftery, A. E. (2011). Default priors and predictive performance in bayesian model averaging, with application to growth determinants. *Journal of Applied Econometrics*, 26(1):30–55. 2, 6, 17

Eklund, J. and Karlsson, S. (2007). Forecast Combination and Model Averaging using Predictive Measures. *Econometric Reviews*, 26:329–362. 22

Faber, K. and Kowalski, B. (1997). Modification of Malinowski's F-test for abstract factor analysis applied to the quail root II data set. *Journal of Chemometrics*, 11:53–72. 97

Faust, J., Gilchrist, S., Wright, J., and Zakrajšek, E. (2011). Credit spreads as predictors of real-time economic activity: a bayesian model-averaging approach. *NBER Working Paper*, 16725. 59, 69

Feldkircher, M. and Zeugner, S. (2009). Benchmark Priors Revisited: On Adaptive Shrinkage and the Supermodel Effect in Bayesian Model Averaging. Working Paper 09/202, IMF. 46, 47, 68, 103, 104, 111, 127, 128, 129

Feldkircher, M. and Zeugner, S. (2012). The Impact of Data Revisions on the Robustness of Growth Determinants – A Note on 'Determinants of Economic Growth. Will Data Tell?'. *Journal of Applied Econometrics*, (forthcoming). 17, 18, 106

Fernández, C., Ley, E., and Steel, M. F. (2001a). Benchmark Priors for Bayesian Model Averaging. *Journal of Econometrics*, 100:381–427. 2, 4, 5, 6, 9, 11, 12, 13, 14, 15, 18, 20, 23, 25, 45, 46, 47, 48, 68, 92, 95, 102, 103, 111

Fernández, C., Ley, E., and Steel, M. F. (2001b). Model Uncertainty in Cross-Country Growth Regressions. *Journal of Applied Econometrics*, 16:563–576. 13, 17

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and Statistics*, 82(4):540–554. 92

Foster, D. P. and George, E. I. (1994). The Risk Inflation Criterion for Multiple Regression. *The Annals of Statistics*, 22:1947–1975. 6

Frale, C. and Veredas, D. (2009). A monthly volatility index for the US real economy. Discussion Paper 2008-08, ECARES. 64

Galati, G. and Moessner, R. (2010). Macroprudential policy – a literature review. Working Paper 267, De Nederlandsche Bank. 55

Gelfand, A. and Dey, D. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(3):501–514. 24

Gelman, A., Carlin, J. B., Stern, S. H., and Rubin, B. D. (1995). *Bayesian Data Analysis*. Chapman & Hall. 14

George, E. (2010). Dilution priors: Compensating for model space redundancy. In Berger, J., Cai, T., and Johnstone, I., editors, *Borrowing Strength: Theory Powering Applications–A Festschrift for Lawrence D. Brown. IMS Collections*, pages 158–165. 69

George, E. and Foster, D. (2000). Calibration and empirical Bayes variable selection. *Biometrika*, 87(4):731–747. 12

Gerali, A., Neri, S., Sessa, L., and Signoretti, F. (2010). Credit and Banking in a DSGE Model of the Euro Area. *Journal of Money, Credit and Banking*, 42:107–141. 55

Gertler, M. and Kiyotaki, N. (2010). Financial intermediation and credit policy in business cycle analysis. Working paper, New York University. 55

Geweke, J. (1977). The Dynamic Factor Analysis of Economic Time Series. In Aigner, D. and A.S. Goldberger, A., editors, *Latent Variables in Socio-Economic Models*. Amsterdam: North-Holland. 111

Giannone, D., Lenza, M., and Primiceri, G. (2012). Prior selection for vector authoregressions. *CEPR Discussion Paper*, 8755. 3

Giannone, D., Lenza, M., and Reichlin, L. (2008). Explaining the Great Moderation: it is not the shocks. *Journal of the European Economic Association*, 6:621–633. 64

Giannone, D., Lenza, M., and Reichlin, L. (2010). Market freedom and the global recession. *IMF Economic Review*, 59(1):111–135. 102

Giannone, D., Reichlin, L., , and Sala, L. (2005). Monetary policy in real time. *NBER Macroeconomics Annual 2004*, 19:161–200. 57

Gilchrist, S. and Zakrajsek, E. (2010). Credit spreads and business cycle fluctuations. mimeo. 108, 110

Guptar, A. K. and Nagar, D. K. (2000). *Matrix Variate Distributions*. Chapman & Hall, CRC, Monographs and Surveys in Pure and Applied Mathematics, 104. 22

Gürkaynak, R., Sack, B., and Wright, J. (2007). The U.S. treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 54:2291–2304. 67

Hansen, B. (2007). Least squares model averaging. *Econometrica*, 75(4):1175–1189. 94

Hansen, M. and Yu, B. (2001). Model selection and the principle of minimum description length. *Journal of the American Statistical Association*, 96(454):746–774. 12

Hatzius, J., Hooper, P., Mishkin, F., Schoenholtz, K., and Watson, M. (2010). Financial condition indexes: a fresh look after the financial crisis. Working Paper 16150, NBER. 108, 110

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science*, 14, No. 4:382–417. 2, 4, 94, 95

Hubrich, K. and West, K. (2010). Forecast evaluation of small nested model sets. *Journal of Applied Econometrics*, 25:574–594. 61

Huizinga, H. and Laeven, L. (2009). Accounting discretion and banks during a financial crisis. Working Paper 09/207, IMF. 56

Int Veld, J., Raciborski, R., Ratto, M., and Roeger, W. (2011). The recent boom-bust cycles: the relative contribution of capital flows, credit supply and asset bubbles. *European Economic Review*, 55:386–406. 55

Johnson, S., Larson, W., Papageorgiou, C., and Subramanian, A. (2009). Is Newer Better? Penn World Table Revisions and Their Impact on Growth Estimates. Working Paper 191, Center for Global Development. 17, 19, 20

Kass, R. and Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, 90:928–934. 6

Kilian, L. and Vigfusson, R. (2009). Pitfalls in estimating asymmetric effects of energy price shocks. Discussion paper, CEPR. 63

Kollmann, R., Enders, Z., and Müller, G. (2011). Global banking and international business cycles. *European Economic Review*, 55:407–426. 55, 64

Kollmann, R. and Malherbe, F. (2011). Financial contagion. Discussion Paper 2011-01, ECORE. 54, 64

Kollmann, R. and Zeugner, S. (2012). Leverage as a predictor for real activity and volatility. *Journal of Economic Dynamics and Control*. 108, 110

Koop, G. and Potter, S. (2003). Forecasting in Large Macroeconomic Panels Using Bayesian Model Averaging . *FRB NY Staff Report*, 163. 13

Koop, G. and Potter, S. (2004). Forecasting in dynamic factor models using bayesian model averaging. *The Econometrics Journal*, 7(2):550–565. 93

Korobilis, D. (2009). Assessing the transmission of monetary policy shocks using dynamic factor models. Working Paper 35-09, Rimini Centre for Economic Analysis. 108, 110

Kraay, A. and Tawara, N. (2010). Can disaggregated indicators identify governance reform priorities? *World Bank Policy Research Working Paper*, 5254. 99

Kritchman, S. and Nadler, B. (2008). Determining the number of components in a factor model from limited noisy data. *Chemometrics and Intelligent Laboratory Systems*, 94:19–32. 97

Krugman, P. (2009). The international finance multiplier. Working paper, Princeton University. 54, 64

Laud, P. and Ibrahim, J. (1995). Predictive model selection. *Journal of the Royal Statistical Society*, Series B 57:247–262. 14

Lawley, D. and Maxwell, A. (1963). *Factor Analysis as a Statistical Method*. Butterworths. 97

Leamer, E. (1978). *Specification searches: ad hoc inference with nonexperimental data*. John Wiley and Sons, New York. 94

Ley, E. and Steel, M. (2011a). Mixtures of g-priors for bayesian model averaging with economic application. Policy Research Working Paper 5732, World Bank. 68

Ley, E. and Steel, M. F. (2009). On the Effect of Prior Assumptions in Bayesian Model Averaging with Applications to Growth Regressions. *Journal of Applied Econometrics*, 24(4):651–674. 5, 18, 25, 92, 95, 101, 111, 131

Ley, E. and Steel, M. F. (2011b). Mixtures of g-priors for bayesian model averaging with economic applications. Policy Research Working Paper 5732, World Bank. 2, 9, 47, 95

Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g Priors for Bayesian Variable Selection. *Journal of the American Statistical Association*, 103:410–423. 2, 3, 5, 9, 10, 12, 16, 23, 24, 47, 95, 103

Magnus, J., Powell, O., and Prüfer, P. (2010). A comparison of two model averaging techniques with an application to growth empirics. *Journal of Econometrics*, 154(2):139–153. 93, 101, 103, 104, 111

Malinowski, E. (1989). Statistical F-tests for abstract factor analysis and target testing. *Journal of Chemometrics*, 3:49–60. 97

Masanjala, W. H. and Papageorgiou, C. (2008). Rough and Lonely Road to Prosperity: A Reexamination of the Sources of Growth in Africa Using Bayesian Model Averaging. *Journal of Applied Econometrics*, 23:671–682. 13, 45

Meh, C. and Moran, K. (2010). The role of bank capital in the propagation of shocks. *Journal of Economic Dynamics and Control*, 34:555–576. 55, 64

Raftery, A. E. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology*, 25:111–163. 2, 4

Sala-i-Martin, X. (1997). I Just Ran 2 Million Regressions. *American Economic Review*, 87:178–183. 17

Sala-i-Martin, X., Doppelhofer, G., and Miller, R. I. (2004). Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach. *American Economic Review*, 94:813–835. 5, 8, 17, 18, 44, 45, 59, 69, 92, 94, 95, 99, 101, 102, 104, 105

Schott, J. (2006). A high-dimensional test for the equality of the smallest eigenvalues of a covariance matrix. *Journal of Multivariate Analysis*, 97:827–843. 97

Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics*, 6(2):461–464. 6

Stock, J. and Watson, M. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179. 56, 57, 93, 107, 115

Stock, J. and Watson, M. (2005). Implications of Dynamic Factor Models for VAR Analysis. Working Paper 1467, NBER. 107

Stock, J. and Watson, M. (2008). Forecasting in dynamic factor models subject to structural instability. In Castle, J. and Shepherd, N., editors, *The Methodology and Practice of Econometrics, A Festschrift in Honor of Prof. David Hendry*. Oxford: Oxford University Press. 57

Stock, J. and Watson, M. (2011). Generalized shrinkage methods for forecasting using many predictors. mimeo, Harvard University. 93, 108, 109, 111

Stock, J. H. and Watson, M. W. (2006). Forecasting with many predictors. In Elliott, G., Granger, C., and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 1, pages 515–554. Elsevier. 59, 92, 105, 107

Strachan, R. and van Dijk, H. (2004). Exceptions to Bartlett's Paradox. Keele economic research papers. 2

Van den Heuvel, S. (2008). The welfare cost of bank capital requirements. *Journal of Monetary Economics*, 55:298320. 55

von Peter, G. (2009). Asset prices and banking distress: a macroeconomic approach. *Journal of Financial Stability*, 5:298–319. 55

Wagner, M. and Hlouskova, J. (2010). Growth regressions, principal components and frequentist model averaging. mimeo. 93, 103

White, H. (2000). A reality check for data snooping. *Econometrica*, 68:1097–1126. 112

Yuan, Z. and Yang, Y. (2005). Combining Linear Regression Models: When and How? *Journal of the American Statistical Association*, 100:1202–1214. 101, 103, 104

Zellner, A. (1986). *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, chapter On Assessing Prior Distributions and Bayesian Regression Analysis with g-Prior Distributions. North-Holland: Amsterdam. 2, 4, 95

Zellner, A. (2008). Comments on Mixtures of g-priors for Bayesian Variable Selection by F. Liang, R. Paulo, G. Molina, MA Clyde and JO Berger. mimeo, University of Chicago. 9