Assessing the Impact of Private Sector Balance Sheets on Financial Crises: A Comparison of Bayesian and Information-Theoretic Measures of Model Uncertainty

Simon J. Godsill
Department of Engineering
University of Cambridge

Mark R. Stone
International Monetary Fund
Monetary and Exchange Affairs
Washington, DC

Melvyn Weeks*
Faculty of Economics and Politics
University of Cambridge

April 14 2004

Abstract

This paper examines the intensity of financial crises during the 1990s with a view to informing crisis prevention and mitigation policies. We compare the performance of a full Bayesian and an information-theoretic approach in addressing the econometric problems posed by the lack of a unifying theoretical model, a large number of crisis indicators, and a number of additional data shortfalls. The results indicate that the probability and intensity of financial crises are heightened by corporate illiquidity and leverage, a lack of nonbank sources of financing, excessive domestic relative to foreign currency liquidity and a cutoff of capital inflows. The implications are that policy measures aimed at improving the operation and monitoring of the corporate and nonbank financial sectors could reduce crisis vulnerability.

JEL Classification: B41; C21; C23; C51; C82; F33
Acknowledgements: The views expressed in this Working Paper are those of the author(s) and do not necessarily represent those of the IMF or IMF policy.

Author's email address: mw217@econ.cam.ac.uk

*Monetary and Exchange Affairs Department (MAE), IMF, and Faculty of Economics and Politics, University of Cambridge, respectively.

1 Introduction

Our starting point is the real-life situation of a policymaker aiming to identify and collect economic data, evaluate competing models of the intensity of financial crisis, and make policy decisions with a view to preventing and mitigating financial crises. The policymaker may be interpreted as either the IMF or the World Bank aiming to determine which crisis indicators to employ in their new role of assessing financial vulnerability. The tools available are a set of multiple, overlapping theories of financial crises emphasizing different channels (e.g., foreign exchange liquidity, bad banks) and a large set of economic data that encompass potentially useful indicators of crisis shocks and channels, but may be costly to collect. In this context, it seems sensible for the policymaker to extract useful crisis indicators from the data by imposing priors based on the literature, choosing indicators that explain the intensity of historical financial crises, and paying the costs of collecting these data. Uncertainty over which policy to recommend follows from a number of sources of uncertainty, including theory and measurement uncertainty. In this study we assume that the policy maker wants to evaluate policies *unconditionally* with respect to a potentially large number of alternate models of financial crisis intensity.

The assessment of post-crisis dynamics involves estimation of the *intensity* of a crisis in terms of its impact on the real sector. Intensity can be thought of as the distance that the economy travels from the pre-crisis equilibrium measured along the output dimension. This definition is useful for policy because governments care most about the welfare costs of financial crises, and welfare costs have a higher correlation with real GDP than with financial sector indicators. In addition, accurate financial indicators of crisis intensity are problematic, especially indicators meant to capture aggregate bank distress. Empirically, crisis intensity is gauged by the change in real GDP relative to the pre-crisis trend, conditional on the occurrence of a crisis.

This paper examines the intensity of financial crises during the 1990s. The motivation is the new mandate for the IMF and World Bank to undertake comprehensive assessments of the vulnerability of the financial sectors of member countries.[1] We address a number of fundamental problems posed by empirical analysis of financial crises. These problems are the lack of a single "true" underlying model, the combination of a large number of candidate indicators (many of which represent different measures of the same underlying construct), small sample size, and missing data. Our methodology recognises that in many instances the evolving body of theory and available crisis data do not support a single model, and in this regard it is important to provide some measure of the degree of uncertainty surrounding the process of indicator selection. To do this we calculate a set of data-based weights which we use as a metric to evaluate the degree of support for both specific models and individual indicators. We demonstrate that these weights are easy to calculate and contrast them with alternative measures of model uncertainty.

We note that this paper takes a different tack than the high frequency early warning system (EWS) literature (Berg, Borensztein, Milesi-Ferretti, and Pattillo (1999) and Mulder, Perrelli, and Rocha (2001)). An increasing number of studies are developing EWS's, typically with monthly data. These EWS's aim to identify a small number of leading crisis indicators or composite measures of vulnerability to provide relatively quick warn-

[1] The joint World Bank-IMF Financial Sector Assessment Program (FSAP) was introduced in May 1999 to assess financial system soundness in member countries.

ing signals of impending crises to trigger countervailing adjustments in macroeconomic policies. In extending the earlier work of Weeks and Stone (2001), the goal of this paper, rather, is to examine the issue of model uncertainty as applied to understanding the determinants of the intensity of financial crises. Specifically we compare findings based upon a fully Bayesian approach with two *approximations* : one the Akaike information criterion, based upon a measure of the distance between a true and competing model; and an approach which is a large sample approximation to a Bayesian approach.

This paper is organized as follows. The theoretical and empirical literature on crisis intensity is reviewed in section two. We consider the nature of the policymakers problem in section three, and use this to framework to examine the basic properties of the different approaches. In section four present the form of our prior specifications, and section five presents an overview of the MCMC methodology for constructing the posterior quantities of interest. The data used for this study is introduced in section six, and in section seven ee present our results . Section eight concludes.

2 Review of the Theoretical Literature to Assessing Crises of the 1990s

This section reviews the theoretical and empirical financial crisis literature with a view to motivating the empirical model of crisis intensity. Ideally, the supporting theory for an econometric analysis provides a single or small number of conceptual models with testable hypotheses. However, the theoretical work on systemic financial crises is marked by a multiplicity of explanations and lack of a unifying framework. In particular, the literature is constantly in flux because financial crises themselves are, by definition, ever changing. This is especially true of empirical analysis of crisis intensity.

The early theoretical financial crisis literature focused on currency crises and can be summarized in terms of "generations" of models that emphasized, first, the abandonment of a fixed exchange rate regime owing to fiscal channels (Krugman (1979)), and, second, multiple equilibria were developed in response to the absence of apparent fiscal instability (Obstfeld (1994)). Another strand of the literature focussed on bank crises mostly stressing the interplay between bank balance sheet and balance of payments (Velasco (1987)). The relatively recent foreign exchange liquidity approach explicitly addresses crisis channels arising from a shortfall of foreign exchange liquidity (Chang and A. (1999)). Many of the more recent and successful theoretical models of crises are rooted in the emergence of a crisis collateral channel (Gertler, Gilchrist, and Natalucci (2000)). Kiyotaki and Moore (1997)) introduced a new more direct collateral channel emphasizing macroeconomic rigidities in the form of underdeveloped domestic financial sector and corporate and financial sector balance sheets. The dynamic interaction between credit limits and the prices of assets used for collateral is a powerful crisis channel (Caballero and Krishnamurthy (1999) and Caballero and Krishnamurthy (2000)).

Almost all empirical analysis of financial crises uses binary indicator dependent variables. Empirically, a currency crisis is typically defined to occur when a weighted average of the exchange rate, international reserves and in some cases interest rates passes a predefined threshold (Eichengreen, Rose, and Wyplosz (1996); Berg, Borensztein, Milesi-Ferretti, and Pattillo (1999)). Bank crises are almost always gauged with a binary indicator because they are difficult, if not impossible, to measure with a continuous indicator (Eichengreen and Rose (1998); Demirguc-Kunt and Detragiache (1999a) and Demirguc-Kunt and Detragiache (1999b)). There has been relatively little joint empirical analysis of currency and bank crises, probably reflecting the difficulty of defining the latter. A notable exception is (Kaminsky and Reinhart (1999)).

An increasing number of papers are applying standard econometric techniques to the subject of this paper, crisis-induced output contractions, or what we refer to as the intensity of crises. Eichengreen and Rose (1998) found that bank crises produce output growth declines of 2-3 percent compared with noncrisis countries, but last only about a year. Demirguc-Kunt and Detragiache (1999b) emphasize vulnerability to large capital inflows, bank deposit insurance, and the legal system. Demirguc-Kunt, Detragiache, and Gupta (2000) looked at the pattern post-bank crisis output contraction during bank crises over 1980-95 and found that they last only a year or two, even though credit growth recovers quite slowly. Kaminsky and Reinhart (1999) conclude that the output contraction from concurrent crises (8 percent below non-crisis periods) is more severe than for single crises. They found that financial liberalization and increased capital inflows set the stage for crises, and that they are preceded by recession, which is attributable to a mix of terms of trade shocks, an overvalued exchange rate, and rising credit costs. Stone (2000) looked at the impact of financial crisis on output via the corporate sector and concluded that crisis-induced output contractions are associated with high levels of corporate debt, openness, and exchange rate over appreciation. Bordo, Eichengreen, Klingebiel, and Martinez-Peria (2001) examined output contractions over the past 120 years and concluded that the probability of crisis has increased but intensity has not. They attribute the increased probability to capital mobility and financial safety nets. Hoggarth, Reis and Saporta (2001) estimated cumulative output losses during crisis periods of roughly 15-20 percent of GDP and found that output losses incurred during crises in developed countries are as high, or higher, on average, than those in emerging market economies. Hutchison and Neuberger (2001) conclude that severe currency crises in emerging markets reduce output by about 5-8 percent over a two-three year period, an impact two to four times larger than the average output loss in a developing economy. The large output costs are likely related to their dependence on private capital markets and abrupt reversals in capital inflows that in turn force substantial real-side adjustment. In sum, a consensus has by no means been reached on crisis intensity in the empirical literature.

3 The PolicyMakers Problem

(*Needs more on policy motivation - can we isolate some indicators as quasi policy instruments, or at least indicators to monitor...*)

The nature of the problem facing the policymaker can be described as follows. Based upon a dataset $\mathbf{d}$, the policymaker seeks to make inference on the determinants of financial crises. For a given model of crises, m, the policymaker faces uncertainty over β, the parameters of the model, which will determine the effects of any policy conducted on the model. Thus, if we ignore any effect of model uncertainty then we may let $q(\beta|\mathbf{d}, m)$, represent the density of β, conditional on data d and model m. However, given the manner in which policy is conducted it is more appropriate to consider the *unconditional* effects of model parameters, and moreover, in this particular instance a policymaker would like guidance on what indicators are important for making inference on the determinants of financial crisis. Letting $\mathcal{M}$ define a space of models, the object $q(\beta|\mathbf{d})$ permits unconditional inference by integrating out all aspects of model uncertainty. In addition we are also interested in a measure $q(\gamma_l|\mathbf{d})$, which by integrating over $\mathcal{M}$, allows unconditional inference on the relative importance of the l^{th} indicator, as measured by a quantity which represents the probability that, conditional on the data, indicator l is relevant for intensity of crisis.[2]

In this particular instance the primary source of uncertainty relates to the observation that the policy maker faces a large number of overlapping theoretical models with the implication that there is no single "true" underlying model; a corollary of this is that there are a large number of candidate indicators. The combination of a large number of indicators and small sample size is also likely to result in collinearity with a large number of candidate models that differ marginally. In this context the use of both information theoretic (IT) and Bayesian approaches to model uncertainty have been advocated by a number of analysts. For example, Granger, King, and White (1995) advocate an information theoretic (IT) approach over formal hypothesis testing when testing vague economic theories. In the face of the aforementioned problems, the restricted null model is conferred a favourable advantage, such that in the testing of crisis indicators, the direction of error is to erroneously conclude that these indicators have no explanatory power. Note also that standard likelihood ratio tests for nested models, and modified likelihood in the case of non-nested models (see, for example, Cox (1961) and Pesaran and Weeks (2000)), are not helpful given the focus upon binary comparisons and more importantly, the assumption that one of the models considered is the true model.

The existence of a large number of candidate indicators will generate a large number of models. The methodological approach used to address these problems is predicated upon the identification of a candidate set of variables, say $\mathcal{K}$, in this instance crisis channel indicators. Given $\mathcal{K}$, the space of models, here $\mathcal{M}$, has in the absence of any constraints, dimension 2^I. Model selection proceeds by searching over the model space implied by $\mathcal{K}$ (possibly in conjunction with constraints imposed by the policymaker) and evaluating the performance of different subsets of indicators which are not derivative of a general base model. In this respect there is *no path dependence* in the selection procedure, and the approach is valid for both nested and non-nested models.[3][4] The principle advantages of this approach are that the analyses of uncertainty model does not require the identification of a single general model as a starting point. In contrast, a general-to-specific methodology is founded upon the identification of an unrestricted (and possibly large) congruent base model, and proceeds by testing downwards. Although such an approach is well suited for certain policy applications which can reliably be based on a single and well defined economic model (such as estimating a small macroeconomic model for monetary policy), limitations in the empirical analysis of financial crises create problems for applications

[2] As Brock, Durlauf, and West (2003) point out, in between these two extremes exist various ad hoc robustness checks. For example, evaluating $q(\theta|\mathbf{d}, m)$ with respect to $q(\theta|\mathbf{d}, m')$ for m' a baseline model.

[3] An important caveat here is that the process of model selection is obviously conditional upon prior identification of an initial indicator set, $\mathcal{K}$.

[4] Burnham and Anderson (1998) provide an overview of the IT approach to model selection, and Pesaran and Weeks (2000) evaluate the IT and the GTS in the context of model selection.

of this approach.[5] Specifically, given the presence of a large number of similar empirical measures of crisis indicators it will be difficult to identify a base model, and conduct inference within it.[6][7]

In this study we evaluate the degree of model uncertainty using both information-theoretic and Bayesian methodologies. At the outset it is important to highlight that approaches to model uncertainty founded upon information theoretic principles are fundamentally different from the Bayesian approaches. Central here is the distinction between what Bernardo and Smith (2000) refer to as $\mathcal{M}$-closed and $\mathcal{M}$-open model spaces. Thus, whereas the AIC proceeds, using Kullback-Leibler distances, to either *select* a KL best model, or average over a set of models, by construction the model space is open in that the true model is not contained in $\mathcal{M}$. In this respect the interpretation of asymptotic arguments over $\mathcal{M}$ are quite different. Namely, whereas AIC based model selection is based on the notion that the best approximation to the true model will receive a weight of 1, Bayesian approaches assume that $\mathcal{M}$ is closed, and that an asymptotic weight of unity applies to the true model.

Our choice of Akaikes Information Criterion (AIC - see Akaike (1973)) over alternative information-theoretic measures depends critically upon the observation that AIC is founded upon the notion that a true model does not exist; and that the purpose of both model selection and model averaging is to find the best approximating model. This criteria is distinct from a number of alternate approaches, such as the Bayesian Information Criterion (BIC) derived by Schwarz (1978), and a fully Bayesian methodology. The premise behind the use of BIC is that the objective of model selection is to locate the true model, which is fixed in dimension as sample size increases, and assumed to lie within the candidate set of models. In this respect a particular notion of consistency applies in that as sample size increases, the probability of locating the true model approaches one. However, in both the biological and social sciences it is more appropriate to envisage that the best approximation to an unknown true model is more likely to increase in dimension as sample size increases.[8] Obviously this problem has parallels with the incidental parameter problem which underlies the inconsistency of the estimator of fixed effects in a panel data regression. Here, as sample size increases there are additional parameters to estimate, such that the fundamental properties required for the consistency of an estimator are violated. AIC estimates of relative (expected) Kullback Leibler distance lack a dimension-consistency property given the premis that as sample size increases the data is more able to support more complex models, and thereby offer the potential for locating a better approximate model. Moreover, a number of analysts including Burnham and Anderson (1998) have noted the tendency of dimension-consistent methods to select under-fitted models when sample size is less than very large.

[5] See Davidson and Hendry (1981) for a discussion of the limitations of GTSA, and for recent applications of GTSA see, for example, Krolzig and Hendry (2000) Campos and Ericsson (2000) and White (1999).

[6] Application of GTS in the absence of a single "true" underlying model would lead to path dependence in the order of test, which could erroneously lead to the omission or inclusion of indicators that could be useful for analysis

[7] Burnham and Anderson (1998) provide an overview of the IT to model selection, and Pesaran and Weeks (2000) evaluate both the IT and the GTS in the context of model selection.

[8] As Swanson and White (1995) note, one caveat here is that in using an IT approach it is difficult to assess the magnitude of the implicit type 1 error which underlies the selected model. For dimension consistent procedures such as BIC, size is asymptotically zero. However, this is not the case for AIC.

3.1 A Model Framework

The assessment of post-crisis dynamics involves estimation of the intensity of a crisis in terms of its impact on the real sector. The measurement of crisis intensity using the change in real GDP relative to the pre-crisis trend (conditional on the occurrence of a crisis) is relatively uncontroversial, and, more importantly readily observed. In this respect intensity can be thought of as the distance that the economy travels from the pre-crisis equilibrium measured along the output dimension. This definition is useful for policy because governments care most about the welfare costs of financial crises, and welfare costs have a higher correlation with real GDP than with financial sector indicators. However, the determinants of crises, and the nature of the crisis channels - i.e. the role of the external sector, collateral, financial breadth, and the legal environment, are varied and not so easy to either define or measure.

Let $\mathcal{K}$ denote the set of crisis intensity indicators, indexed by $i = 1, ..., I$. We denote the set of models by $\mathcal{M}$ with the h^{th} member given by M_h; the dimension of M_h is denoted k_h. Observed data is $\mathbf{y} = \{y_j\}$ where y_j denotes a measure of the intensity of crisis (i.e. GDP - trend), and $\mathbf{x}_j = \{x_{ij}\}$ is a $I \times 1$ vector representing the total set of covariates for crisis episodes $j = 1, ..., n$. It is also possible to partition $\mathcal{K}$ into J crisis channel indicator groups, say $\boldsymbol{\omega}^{(1)}, ..., \boldsymbol{\omega}^{(J)}$. For example, $\boldsymbol{\omega}^{(1)} = (x_1^1, x_2^1, ..., x_l^1)'$ might denote indicators of corporate balance sheet channels (e.g., total debt to common equity and the ratio of total debt to total assets) which are believed to be critical determinants of the intensity of crises episodes; we have, in certain cases, a large number of similar measurements of these constructs.

In such a situation the policymaker is faced with considerable uncertainty given that theory is weak, in selecting, for example, the appropriate indicators *within* $\boldsymbol{\omega}^{(1)}$. Given a large set of indicators, $\mathcal{K}$, the objective is to identify a smaller subset of indicators, say $\mathcal{K}_s$, and obtain some measure of the unconditional importance of each indicator as a determinant of crisis. In this respect we note that what we call model uncertainty is aligned with specification uncertainty, insofar as a relatively large number of competing theories generates an indicator space with a large dimension; and that conditional on a given theory, the existence of rival measures of a single construct, will generate measurement uncertainty.

Although at the outset we assume that $\mathcal{K}_s$ is constant across countries and time, we also incorporate what Brock, Durlauf, and West (2003) refer to as *heterogeneity* uncertainty, by exploring the extent to which the processes generating the intensity of financial crises, vary, both over time and across countries (emerging market versus industrialised). We note that initially we assume complete knowledge of potential time and cross-sectional breaks in this process.

We postulate that covariates x_{ij}, $i = 1, ..., I$ may affect the observed data through a linear regression. We model this as

$$y_j = \alpha + \sum_{i=1}^{I} \beta_i x_{ij} + \varepsilon_j, \tag{1}$$

where $\varepsilon_j \sim N(0, \sigma_\varepsilon^2)$. Now, suppose that we wish to make inference on the subset of variables which are important for predicting the crisis observations $\{y_j\}$. This will correspond in the linear regression model above to certain parameters β_i being identically equal to zero. Denote the subset of non-zero β_i parameters as $\beta_{\mathcal{K}_s}$. The principal task is then to estimate $\beta_{\mathcal{K}_s}$ for the dataset $\{y_j\}$ (*Is this true here...*). In this respect, we may reparameterise the problem in terms of indicators $\gamma_i \in \{0,1\}$, $i = 1, ..., I$, which determine whether a particular crisis indicator is relevant to the data. We can rewrite the reparameterised model in the following form:

$$y_j = \alpha + \sum_{\{i;\ \gamma_i = 1\}} \beta_i x_{ij} + \varepsilon_j,$$

where element γ_i is equal to one (zero) if β_i is included (excluded) from the model. In matrix-vector form we have:

$$\mathbf{y} = \alpha\mathbf{1} + \mathbf{X}_\gamma \boldsymbol{\beta}_\gamma + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\gamma} = \{\gamma_i\}$ denotes an I length vector of binary indicator variables which we use to index the 2^I distinct models in $\mathcal{M}$, $\mathbf{1}$ denotes a vector containing all ones. $\mathbf{X}_\gamma \subset \mathbf{X}$ is the design matrix with columns extracted from $\mathbf{X}$ for which $\gamma_i = 1; \boldsymbol{\beta}_\gamma$ is the corresponding vector of β_i parameters for which $\gamma_i = 1$.

3.2 Posterior Model Probabilities

Posterior model probabilities are given by

$$p(M_h|\mathbf{y}) = \frac{l(\mathbf{y}|M_h)p(M_h)}{\sum\limits_{j=1}^{2^I} l(\mathbf{y}|M_j)p(M_j)}, \tag{2}$$

where $p(M_h)$ denotes the prior probability for model M_h. Note that in (2) all uncertainty over $\boldsymbol{\theta}_j = (\sigma_\varepsilon, \boldsymbol{\beta}_h, \alpha)'$ has been integrated out such that $l(\mathbf{y}|M_h)$ represents the *marginal* likelihood of model M_h, given by

$$l(\mathbf{y}|M_h) = \int l(\mathbf{y}|\boldsymbol{\theta}_h, M_h)p(\boldsymbol{\theta}_h|M_h)d\boldsymbol{\theta}_h. \tag{3}$$

$p(\boldsymbol{\theta}_h|M_h)$ denotes the prior for the parameters of model M_h. Using (2) the ratio of the posterior probability for model h and h', say $B_{hh'}$, is therefore given by

$$B_{hh'} = \frac{p(M_h|\mathbf{y})}{p(M_{h'}|\mathbf{y})} = \frac{l(\mathbf{y}|M_h)p(M_h)}{\underbrace{l(\mathbf{y}|M_{h'})}_{A}\, p(M_{h'})}. \tag{4}$$

Making the assumption of equal prior odds, the Bayes factor, given by the ratio A in (4), is equal to the posterior odds.

3.2.1 Approximating Posterior Model Probabilities

Although Kass and Raftery (1995) argue for a full Bayesian approach to model uncertainty, the two fundamental challenges are: (i) the requirement of a full prior specification over elements of both $\boldsymbol{\theta}$ and $\mathcal{M}$; and (ii) for any given model the calculation of posterior probabilities and Bayes factors requires evaluation of integrals in both the numerator and denominator of (3). In section 4 we examine how both of these requirements may be operationalised. However prior to this we first consider two alternative measures of (4) based on an asymptotic Bayesian approximation and an approximately unbiased estimator of the relative Kullback-Leibler distance.

Akaike (1973) proposed a method of model selection based upon the concept of expected information distance. Akaike's Information Criterion (AIC) is based upon the notion that truth cannot be represented in a model form, such that $\mathcal{M}$ contains a number of *approximations* to the truth; with each approximation to a true, unknown, model f, represented by an estimate of the Kullback-Leibler distance $I(f, M_h)$. In this respect AIC is interpreted as an estimate of the expected relative, directed distance between an estimated model and the unknown truth that generated the data; the model in $\mathcal{M}$ with the smallest AIC is then considered to be "closest" to the truth, *relative* to the candidate set in $\mathcal{M}$. Akaike demonstrated that the penalised maximised log-likelihood for model M_h, say

$$l(\hat{\boldsymbol{\theta}}_h|\mathbf{y}, M_h) - k_h, \tag{5}$$

is a unbiased general estimator of $I(f, M_h)$. Akaike's information criterion (AIC) for M_h is generally written as

$$AIC = -2l(\widehat{\boldsymbol{\theta}}|\mathbf{y},M_h) + 2k_h. \tag{6}$$

There have been a number of refinements to Akaike's information criterion. In this study we use a variant based upon the work of Sigiura (1978), Hurvich and Tsai (1989), who developed a small-sample version of AIC given by

$$AIC_s = AIC + \frac{2k(k+1)}{n-k+1}. \tag{7}$$

It is obvious that AIC_s includes an additional bias correction which disappears as the ratio n/k increases.[9]

At this juncture it is important to note that there have been a number of extensions of AIC under the auspices of generalised information criterion (see, for example, Bhansali and Downham ()). In most of these variants the multiplicative factor 2 on the number of parameters in (6) is replaced by an alternative cosntant. However, it is important to emphasise that the AIC penalty term is not arbitrary but firmly rooted in an information theory approach to model selection where $\mathcal{M}$ is assumed open. Thus, wherease both AIC and AIC_s are estimates of KL information, an alternative class of criteria are constructed so as to be dimension consistent. Namely for a $\mathcal{M}$-closed model space, criteria are designed such that the probability of selecting the true model approaches one as sample size increases indefinitely.

The most popular dimension-consistent criterion is due to Schwarz (1978). The Schwarz information criterion (SIC) is based upon an approximation to a fully Bayesian approach to model uncertainty. Predicated on the notion that the true unknown model can be specified and is contained within $\mathcal{M}$, the Schwarz criterion is consistent, in the sense that

[9] See Burnham and Anderson (1998) for further details.

the dimension of the true model is fixed, and that the probability of locating the model with the highest posterior probability approaches one as sample size increases.[10]. The SIC criterion, also referred to as the Bayesian information criterion (BIC), is given by

$$BIC = -2l(\hat{\boldsymbol{\theta}}_h|\mathbf{y}, M_h) + k_h \cdot \log(n)$$

The approximation to a fully Bayesian approach is founded upon the assumption of uniform priors over $\mathcal{M}$ and vague priors over $\boldsymbol{\theta}_h$. Specifically the BIC approximation corresponds to a unit information prior on the model parameters, which is a multivariate normal prior with mean ... and covariance matrix equal to the expected information for one observation (see Raftery (1995) for further discussion)). An asymptotic approximation to the log of the posterior odds for models M_h and $M_{h'}$ (say $B^S_{hh'}$), is then given by

$$B^S_{hh'} = \log\left(\frac{l(\hat{\boldsymbol{\theta}}_h|\mathbf{y}, M_h)}{l(\hat{\boldsymbol{\theta}}_{h'}|\mathbf{y}, M_{h'})}\right) - c \tag{8}$$

where $\hat{\boldsymbol{\theta}}_h$ is the maximum likelihood estimator under M_h and $c = 1/2(k_h - k_{h'})\log(n)$. As $n \to \infty$ the quantity

$$\frac{B^S_{hh'} - \log(B_{hh'})}{\log(B_{hh'})} \to 0.$$

3.2.2 AIC and BIC Differences and Weights

In moving from a process of model selection to explicitly addressing the issue of model uncertainty, it is not the absolute size of information criterion (IC) statistics that matters but the relative values.[11] In particular differences between IC statistics in M are important. Consider the following statistic

$$\Delta_i^{AIC} = AIC_i - \min_{j\in M} AIC = E_{\widehat{\boldsymbol{\theta}}}[\widehat{I}(f, g_i)] - \min_{j\in M} E_{\widehat{\boldsymbol{\theta}}}[\widehat{I}(f, g_j)], \tag{9}$$

where Δ_i^{AIC} obviously facilitates direct comparison of AIC across members of M and is simple to interpret; as Δ_i^{AIC} increases, the less plausible is g_i as the KL best model. Since BIC is a dimension consistent criterion the comparable estimator of difference is given by

$$\Delta_i^{BIC} = BIC_i - \min_{j\in M} BIC \tag{10}$$

Akaike (1983) demonstrated advocates $\exp(-\frac{1}{2}\Delta_h)$ as being the *relative* likelihood (or probability) of the model given the data. Assuming a uniform prior over all models

[10] See Bozdogan (1987) for a review of alternative dimension cosnsitent criterion.

[11] This remark will also have implications for the often observed similarity between AIC and BIC results in empirical work.

in $\mathcal{M}$, the use of a simple normalisation facilitates an approximation , denoted here,[12] $p_{AIC}(M_h|\mathbf{y})$, to the posterior probability for M_h, namely

$$p_{AIC}(M_h|\mathbf{y}) = \frac{\exp(-\frac{1}{2}\Delta_h)}{\sum_{i=1}^{I} \exp(-\frac{1}{2}\Delta_i)}, \tag{11}$$

where $p_{AIC}(M_h|\mathbf{y})$, also referred to as "Akaike" weights, may be interpreted as the evidence in favour of model h as being the actual K-L best model in $\mathcal{M}$. The posterior odds for models h and h', immediately follows as

$$B_{hh'}^{AIC} = \frac{\exp(-\frac{1}{2}\Delta_h)}{\exp(-\frac{1}{2}\Delta_{h'})}. \tag{12}$$

We note that for both the AIC and BIC selection criteria we may also consider a set of non-uniform *prior* probabilities over $\mathcal{M}$.[13] For example, extending (11) the posterior probability for M_h may be approximated

$$p_{AIC}(M_h|\mathbf{y}) = \frac{\exp(-\frac{1}{2}\Delta_h)p(M_h)}{\sum_{i=1}^{I} \exp(-\frac{1}{2}\Delta_m)p(M_i)}. \tag{13}$$

In comparing Bayes factors with approximations to these factors using AIC and BIC a number of observations are possible. First, note that the A in (4) represents a ratio of marginal likelihoods and therefore is equivalent to a ratio of classical likelihoods *conditional* upon integrating out parameter uncertainty as in (3). In this sense the Bayes factors represents a measure of model uncertainty for h relative to $h^{'}$ which integrates out parameter uncertainty.[14] Although we may think of Akaike Weights as an approximation to the Bayes factor,[15] we may not refer to a convergence of $B_{hh'}^{AIC}$ to $B_{hh'}$ based upon the same limiting arguments which apply to the BIC approximation.

4 Specification of Prior Distributions

In general posterior inference *conditional* on a given model is less sensitive to prior specifications relative to instances where model uncertainty is incorporated (see, for example,

[12] Note approximation but not in the sense of a limiting relationship between AIC and fully Bayesian approach to model selection.

[13] Given that AIC and BIC are based upon fundamentally different principles, the notion of what consistitutes a prior distribution over the space of models is also different. Under the Bayesian approach (and the BIC approximation) $p(M_h)$ is the prior belief that model M_h is the *true* model. In contrast the use of an information-theoretic approach interprets (M_h) as the prior belief that M_h is the best *approximation* to an unknown true model.

[14] Poskitt and Tremayne (1983) show that the use of different information-theoretic criteria imply alternative priors over $\mathcal{M}$. For an application of Bayesian model selection to the linear regression model see Raftery, Madigan and Hoeting (1997). Kass and Raftery (1995) provide an excellent overview of the Bayesian approach to inference, including a useful discussion of model uncertainty.

[15] Kass and Raftery (1995) show this for the Schwarz criterion.

Kass and Raftery (1995)). As an extreme example, while it can often be argued that use of improper priors for *model-specific* inference is appropriate, their use can highly problematic in the case of model selection, owing to the Jeffreys-Lindley paradox.[16] For a given model, the attendant arbitrary multiplicative constants cancel in the formula for the posterior distribution of the model-specific parameters. This cancelling does not, in general, happen in the case of posterior inference on model uncertainty, and hence invalid model selection results will be obtained. For example, this applies to the calculation of Bayes factors - see discussion in Raftery comment and Raftery (1995)). An exception to this statement is where an improper prior appears for a parameter which is common (and of fixed dimensionality) in all models.This will be important when we make the distinction between prior specifications on one or more parameters over model-invariant and model-specific parameters (see Jeffreys (1961)).

The principal unknown parameters in (1) are $\boldsymbol{\beta}_\gamma$, $\boldsymbol{\gamma}$, α and σ_ε. In addition there are several unknown hyperparameters and a number of missing (unobserved) covariates x_{ij}. In the fully Bayesian MCMC framework all of these unknowns are treated jointly within one single procedure. In this way, and through the adoption of a partial hierarchical prior structure, we gain the advantage of properly incorporating prior uncertainty about unknowns and removing the dependence on fixed hyperparamters.

The chosen prior structure here may be factorised as follows

$$p(\boldsymbol{\beta}_\gamma, \boldsymbol{\gamma}, \alpha, \sigma_\epsilon, \kappa, \mathbf{X}, \boldsymbol{\mu}_X, \Sigma_X) = p(\boldsymbol{\beta}_\gamma|\boldsymbol{\gamma},\kappa, \mathbf{X})p(\boldsymbol{\gamma})p(\alpha)p(\sigma_\epsilon)p(\kappa)p(\mathbf{X}|\boldsymbol{\mu}_X, \Sigma_X)p(\boldsymbol{\mu}_X, \Sigma_X). \tag{14}$$

In formulating a prior model we need to identify a flexible class of priors for each component of (14), and in a number of cases, make decisions as to whether, to guarantee conjugacy, fix values of hyperparameters, or adopt a hiterarchical prior structure, and thereby remove the influence of a specific hyperparameter. .... Note that covariates $\mathbf{X}$ are included directly within this prior framework thereby facilitating MCMC imputation of missing covariates; $\boldsymbol{\mu}_X$=$\{\mu_i\}_{i=1}^I$and Σ_X=diag($\{\sigma_i^2\}_{i=1}^I$), denote, respectively, the vector of means for covariates and the covariance matrix. κ is a hyperparemeter used in the g-prior specification for $\boldsymbol{\beta}_\gamma$. For the sake of clarity in (14) we do not explicitly write out the dependence of distributions on any *fixed* constant hyperparameters,[17] although this dependence is of course assumed throughout.

We now consider the specific forms proposed for the individual terms in this factorisation.

4.1 Prior for regression parameters β_γ

Here we utilise a multivariate prior of the form

$$p(\boldsymbol{\beta}_\gamma|\boldsymbol{\gamma}, \kappa) = MVN_\gamma(\mathbf{0}, \Sigma_\gamma), \tag{15}$$

[16] Give examples see Kass and Raftery (1995).

[17] For example, $p(\kappa)$, the distribution of parameters in our g-prior is dependent upon two parameters, which is not indicated here. Further $p(\gamma)$, the joint distribution over the zero-one variable selection vector γ, is dependent upon choosing hyperparameters of the Beta distribution.

where $\Sigma_\gamma = \sigma_\varepsilon^2 \kappa (\mathbf{X}_\gamma' \mathbf{X}_\gamma)^{-1}$ for $\kappa > 0$ corresponds to the well known g-prior (see Zellner (1986)). The principal advantages of such a prior specification relate to convenience, specifically conjugacy, and the fact that the structure of the prior reflects an adjustment (by the factor κ) of the covariance structure in the likelihood[18], thereby avoiding a prior that will dominate likelihood information. A key issue here is the choice of the hyperparameter κ. Fernandez, Ley, and Steel (2001a) in focusing upon model uncertainty in the normal linear regression model, examine the consequences of different choices for κ. In an extensive set of simulation experiments κ was set as either a function of the sample size, n, a model-specific number of regressors, k_j, and the total number of available regressors, I. Based on both theoretical and empirical analysis the authors conclude that a choice $\kappa = 1/\max(n, I^2)$ will generate reasonable results. Although the use of a unit information prior for $\boldsymbol{\beta}_\gamma$ (is this equivalent to setting $\kappa = 1/n$?) has been criticised on the grounds that it is too conservative, by setting $\kappa = 1/n$, we have a useful reference point with which to compare other more informative prior[19]. George and Foster (2000) in comparing Bayesian methods for evaluating model uncertainty with approximations based on information criteria, note that AIC corresponds with setting $\kappa = 0.255$ and for BIC $\kappa = 1/n$.

An alternative approach[20] proposed here considers an appropriate prior for this hyperparameter so that κ can be simulated along with the other parameters in the MCMC scheme. This is likely to add robustness and flexibility to the model. We adopt a conjugate gamma prior for this parameter

$$p(\kappa) = G(\kappa | c_\kappa, d_\kappa),$$

where c_κ and d_κ are fixed hyperperameters which can be chosen informatively to satisfy certain properties.

4.2 Prior for indicator vector $\boldsymbol{\gamma}$

Our prior for the zero-one variable selection vector $\boldsymbol{\gamma}$ is based on an independent Bernoulli assumption for each variable, combined with a conjugate prior for the binomial proportion parameter. Conditional upon prior probability $l \in (0, 1)$ that any given variable is used as a predictor, this prior may be written as

$$p(\boldsymbol{\gamma}|l) = \prod_{i=1}^{I} l^{\gamma_i} (1-l)^{(1-\gamma_i)}. \tag{16}$$

As noted by George and Foster (2000), with κ and l fixed to set values, priors (15) and (16) have been used in Bayesian model selection problems (see for example Smith and Kohn (1996) and Fernandez, Ley, and Steel (2001b)). For example, a simple uniform prior

[18] Also guarantee that the prior specification is on the same scale as the likelihood - see Potter and Koop.

[19] Since logs of Bayes factors based on $\kappa = 1/n$ behave asymptotically like BIC, we should also observe that using this prior specification our results for BIC and fully Bayesian are similar ..See Kass and Wasserman (1995) for further discussion.

[20] Note that Knox, Stock and Watson (2002) are among a number of analysts who have adopted an empirical Bayes approach, in the sense of locating a value of κ that maximises the marginal likelihood.

across all elements of the model space may be imposed by setting $l = 0.5$, it is questionable whether the assignment of a uniform prior distribution over the space of possible models is appropriate when $\mathcal{M}$ is large. This follows since a uniform prior suggests that the analyst considers that the number of covariates should be large, and that the posterior distribution $p(\boldsymbol{\gamma}|\mathbf{d})$ will have high probability for models with $I/2$ nonzero coefficients. Sala-I-Martin, Doppelhofer, and Miller (2002) circumvent this problem by selecting a *prior mean model size*, say n_0, such that each variable has prior probability $l = n_0/I$ of being included. A disadvantage of such an approach is that the notion of what constitutes a reasonable prior model size may vary across analysts.

To make the model more robust to this kind of uncertainty, we keep combine the notion of a prior mean model size, with the assigment of a a Beta prior to l, namely

$$p(l) = g(l|a,b) = l^{a-1}(1-l)^{b-1}/B(a,b), \quad a>0,\ b>0,\ 0<l<1, \tag{17}$$

with $B(a,b)$ denoting the complete beta function. Combining (16) and (17), and marginalising with respect to l, our beta-binomial prior $BB(I,a,b)$ is given by

$$\begin{aligned} p(\boldsymbol{\gamma}) &= \int_0^1 p(\boldsymbol{\gamma}|l)g(l|a,b)dl \\ &= \binom{I}{k} B(a+k, b+l-k)/B(a,b), \end{aligned} \tag{18}$$

where $k = \sum \gamma_i$.

Based upon a reparametisation (see Prentice (1986)) for $\tau = (a+b)^{-1}$, and $\pi = a(a+b)^{-1}$ we can rewrite (18) as

$$p(\boldsymbol{\gamma}) = \frac{\binom{I}{k} \prod_{i=0}^{k-1} (\pi + \tau i) \prod_{i=0}^{I-k-1} (1 - \pi + \tau i)}{\prod_{i=0}^{I-1} (1 + \tau i)} \tag{19}$$

(18) represents a Beta-Binomial(I,a,b) prior consisting of a mixture of binomial observations, k, a common number of covariates, I, and a Beta distribution $B(a,b)$ placed on l. Note that our priors for each model in $\mathcal{M}$ will differ both as a result of model size and the assumed distribution over l. By varying the hyperparameters a,b we can examine the sensitivity of our results to prior information. For example, letting $a+b \to \infty$, $\tau \to 0$, then Var(l) goes to zero; this singularity is obviously consistent with simple binomial variation. Alternately, letting $a+b \to 0$, $\tau \to \infty$, and $\tau(1+\tau)^{-1} \to 1$.

Finally it is worth noting that despite integrating out the dependence of the model prior on a fixed model size hyperparameter l, the Bernoulli structure still remains insofar as we entertain model priors founded upon independent probabilities over all elements of $\boldsymbol{\gamma}$. In this respect we do not account for the potential lack of prior independence in the value of l across indicators for a *given model*. Chipman (1996) considers a number of examples where it is appropriate to build in dependence across covariates in formulating model priors. These include cases where an analyst considers the significance of polynomial terms, say x^h, or more generally an interaction effect $x \times z$, and may want to force the inclusion of all terms x^a $a < h$, and both x and z.

4.3 Model-Invariant Priors: σ_ε and α

Fernandez, Ley, and Steel (2001a) advocate the use of an improper non-informative prior for σ_ε given that it is very difficult to choose a value for the hyperparameters, and that for large values of this parameter it is possible for the prior to dominate sample information.

$$p(\sigma_\varepsilon^{-2}) = G(\sigma_\varepsilon^{-2}|c_\varepsilon, d_\varepsilon) \tag{20}$$

Here we utilise a standard gamma conjugate prior distribution for σ_ε^2 and thereby adopt a proper but very diffuse prior specification, with d_ε and c_ε set very close to zero. This represents close to a non-informative belief about the parameter while helping to avoid potential mis-convergence of the MCMC owing to truly improper priors.

Similarly α is assigned a diffuse but proper prior $N(0, \sigma_\alpha^2)$ with σ_α^2 set very large.

4.4 Prior for covariates

There are many possibilities for the prior on covariates. Here we adopt what we regard as the simplest workable scheme: each covariate x_{ij} is assumed to be randomly drawn from a distribution $N(\mu_i, \sigma_i^2)$, with diffuse priors assigned to the means and variances, i.e.

$$x_{ij} \sim N(\mu_i, \sigma_i^2),\ \mu_i \sim N(0, \sigma_\mu^2),\ \sigma_i^2 \sim IG(c_\mu, d_\mu)$$

with σ_μ^2 set very large and c_μ, d_μ both set very close to zero. We define the vector of means for covariates as $\boldsymbol{\mu}_X = \{\mu_i\}_{i=1}^I$ and the covariance matrix as $\Sigma_X = diag(\{\sigma_i^2\}_{i=1}^I)$.

5 Computational inference using MCMC

The prior framework described above facilitates the computation of posterior distributions. Under certain conditions the conditional distribution $p(\boldsymbol{\beta}_\gamma, \alpha, \sigma_\varepsilon|\boldsymbol{\gamma}, \kappa, \mathbf{y}, \mathbf{X})$ and the marginal likelihood $p(\mathbf{y}|\boldsymbol{\gamma}, \kappa, \mathbf{X})$ are available in closed form. However, in the full implementation of our model, which includes both sampling of hyperparameters and imputation of missing covariates, we are outside the realms of analytically tractable models. As already noted, inference in model selection problems can be sensitive to prior specifications, and in this regard we note that the prior specification above includes a number of key hyperparameters. Particularly important will be the specification of the hyperparameters c_κ, d_κ used in the g-prior specification for the regression parameters, and the parameters for the Beta-binomial model distribution, a and b.

5.1 MCMC Method

We utilise Markov Chain Monte Carlo (MCMC) methods to simulate from the marginal distribution $p(\boldsymbol{\gamma}|\mathbf{d})$ where $\mathbf{d} = (\mathbf{y}, \mathbf{X}_{obs})$ is the data plus observed covariates. This is achieved by constructing an irreducible aperiodic Markov chain whose stationary distribution is the joint posterior of all the unknowns, $p(\boldsymbol{\beta}_\gamma, \boldsymbol{\gamma}, \alpha, \sigma_\epsilon, \kappa, \mathbf{X}_m, \boldsymbol{\mu}_X, \Sigma_X|\mathbf{d})$. Such a Markov chain is readily constructed using standard procedures such as the Metropolis-Hastings sampler, the Gibbs sampler, or hybrids of the two. The MCMC scheme here is based around a Gibbs sampler, including some Metropolis-Hastings sub-steps, and adapted to allow for blocking and marginalisation of parameters wherever possible. Using Gibbs sampling on the space of models, $\mathcal{M}$, sequences of the form

$$\boldsymbol{\gamma}^{(1)}, \boldsymbol{\gamma}^{(2)}, ...$$

are generated which converge in distribution to $p(\boldsymbol{\gamma}|\mathbf{d})$.[21] These sequences can be used to determine a range of models with high probability, and estimates of their probability.

As noted in the full implementation of our model, we are outside the realms of analytically tractable models, and therefore utilise MCMC methods. In contrast, Bayesian approaches to model uncertainty adopted by Fernandez, Ley, and Steel (2001b), utilise MCMC more for efficient sampling of a high-dimensional model space than for marginalisation of nuisance parameters and missing data. This is also the case in the use of AIC and BIC and BIC approximations to fully Bayesian posterior inference. Both the AIC and BIC approximation do not require MCMC for the parameters. Namely, given that both these methods do not integrate out parameter uncertainty, the complete posterior $p(\mathbf{y}|M^*) = \int l(\mathbf{y}|M^*, \boldsymbol{\theta})d\boldsymbol{\theta}$ for M^* is not required but only the value of the maximised log-likelihood.[22]

There are options for efficient simulation in this general setting, using, for example reduced conditional distributions or full conditional distributions in the sampling steps. Here we adopt a simple scheme, where we sample $\boldsymbol{\gamma}$ from its reduced conditional distribution (with $\boldsymbol{\beta}_\gamma$, α and σ_ε analytically marginalised), and then imputing ($\boldsymbol{\beta}$, α, σ_ε) as a joint draw from its full conditional distribution. This facilitates simple Gibbs sampling steps for the remaining unknowns, $\mathbf{X}_m$ and κ, where $\mathbf{X}_m$ denotes the missing covariates. The whole scheme is observed empirically to converge very rapidly for the datasets we have tested. In terms of methodology, our scheme falls within the same general class as other MCMC variable selection schemes, such as those of Carter and Kohn (1996), Kuo and Mallick (1997) with the added novelty of explicit modelling and imputation of missing covariates and sampling of hyperparameters. Our scheme can be distinguished from the stochastic search variable selection methods of George and McCulloch (1995), in which contrast to the authors are able to utilise standard MCMC procedures by developing a prior model structure based on a mixture of two distributions. In the sample scheme adopted here where explanatory variables are switched in and out depending upon their relevance to the observed data, each element of $\boldsymbol{\beta}$ is modelled either as a component of a distribution with near singularity centred on zero, or from a distribution of plausible values. Our general scheme can also be viewed as a special version of the reversible jump algorithm of Green (1995), see discussion in **?**.

5.2 Missing Data (incomplete)

The following section describes how we deal with the problem of missing data. We let the $n \times k$ covariate matrix as $\mathbf{X} = \mathbf{X}_{ob} \cup \mathbf{X}_m$ where the subscripts ob (m) denote observed (missing) data. $\mathbf{X}_{ob,\gamma}$ denotes the set of observed data for the model based on a particular

[21] In certain instances the posterior distribution $p(\boldsymbol{\gamma}|\mathbf{d})$ may be quite flat, with a large number of competing models having high posterior probability. As noted by George and McCulloch (1995), such a situation can arise in the presence of a $\mathcal{M}$ with high collinearity among indicators.

[22] In fact after a best model is found (or a set of best models) it would be possible to do a full MCMC for parameters if required.

$\boldsymbol{\gamma}$ configuration. Within a Gibbs Sampling setup covariates are imputed as an additional step. Assuming a simple independent Gaussian prior model for the missing covariates, namely

$$p(\mathbf{X}_m) \equiv N(\overline{\mathbf{X}}_{ob}, \Sigma_{ob}),$$

where Σ_{ob} is a diagonal matrix with typical element σ_k^2,[23] missing data $\mathbf{X}^m = \left\{ x_{ij}^m \right\}$ are sampled from

$$\mathbf{X}_{m,\gamma} \sim p(\mathbf{X}_{m,\gamma} | \overline{\mathbf{X}}^*_{obs,\gamma}, \boldsymbol{\beta}_\gamma, \boldsymbol{\gamma}, y, \sigma_\varepsilon, c_0, d_0).$$

Assume $p(\mathbf{X}_{m,\gamma}| \quad) \equiv N(\overline{\mathbf{X}}^*_{ob,\gamma}, \sigma_x^2 B^*)$ By combining prior and data information, the posterior mean $\overline{\mathbf{X}}^*_{m,\gamma}$ may be written

$$\overline{\mathbf{X}}^*_{m,\gamma} = B^* \times [(P_p \times \overline{\mathbf{X}}_{ob,ob}) + (D_p \times \widehat{\mathbf{X}}_{m,\gamma})]. \tag{21}$$

where $B^* = (P_p + D_p)^{-1}$, with P_p (D_p), denoting, respectively, prior and data precision. $\widehat{\mathbf{X}}_{m,\gamma}$ denotes the imputed value from data[24]. Letting $P_p = 1/vec(\Sigma_{ob})$, $D_p = \boldsymbol{\beta}'_m \boldsymbol{\beta}_m / \sigma_\varepsilon^2$, and $\widehat{\mathbf{X}}_{m,\gamma} = y_m / \boldsymbol{\beta}_m - \mathbf{X}_{ob,\gamma} \boldsymbol{\beta}_{ob,\gamma}$, (21) may be written as

$$\overline{\mathbf{X}}^*_{m,\gamma} = (\boldsymbol{\beta}_m / \sigma_\varepsilon^2) \times y_m - \mathbf{X}_{ob,\gamma} \boldsymbol{\beta}_{ob,\gamma}$$

Also integrate the above with sample mean and variance hyperparameters for missing covariates:

$$p(\mu_i, \sigma_i^2 | X) = p(\mu_i, \sigma_i^2 | \mathbf{X}, \mathbf{y}, \beta_\gamma, \gamma, \sigma_\varepsilon)$$

5.3 Summary of algorithm

A single sweep of the full MCMC algorithm for the fully Bayesian approach can be summarised as follows:

1. Indicators γ_i, $i = 1, 2, ..., I$ are sampled in turn with replacement from the *reduced* conditional distribution $p(\gamma_i | \boldsymbol{\gamma}_{/i}, \mathbf{y}, \mathbf{X}, \kappa)$, where $\boldsymbol{\gamma}_{/i} = (\gamma_1, ..., \gamma_{i-1}, \gamma_{i+1}, ..., \gamma_I)$, using a Metropolis-Hastings update. Obviously in sampling from a reduced conditional distributions this implies that we have analytically marginalized the remaining components of this conditional.

   Specifically, if the current state of γ_i is given by s then we propose a change to $1 - s$ and accept this change according to the Metropolis-Hastings rule with probability

   $$\min \left(1, \frac{p(\gamma_i = (1-s) | \boldsymbol{\gamma}_{/i}, \mathbf{y}, \mathbf{X}, \kappa)}{p(\gamma_i = s | \boldsymbol{\gamma}_{/i}, \mathbf{y}, \mathbf{X}, \kappa)} \right).$$

[23] Currently we do not take draws from this prior distribution, but use fixed data values $\overline{X}_{ob}$, Σ_{ob}.

[24] Note: in operationalising this component of the MCMC algorithm we set $P_p = 0$ for AIC and BIC approximations to Bayesian approach.

2. $(\boldsymbol{\beta}_\gamma, \alpha, \sigma_\varepsilon)$ are sampled as a block from their full conditional $p(\boldsymbol{\beta}_\gamma, \alpha, \sigma_\varepsilon|\boldsymbol{\gamma}, \kappa, \mathbf{y}, \mathbf{X})$. We note that such block sampling follows from the the fact that $p(\boldsymbol{\beta}_\gamma, \alpha, \sigma_\varepsilon|.)$ has a Normal-inverted gamma distribution.

3. Impute missing covariates from their full conditional distribution

$$\mathbf{X}_m \sim p(\mathbf{X}_m|\mathbf{y}, \mathbf{X}_{obs}, \kappa, \boldsymbol{\beta}_\gamma, \alpha, \sigma_\varepsilon),$$

   where $\mathbf{X}_{obs}$ are the observed covariates.

4. Sample mean and variance hyperparameters for missing covariates:

$$p(\mu_i, \sigma_i^2|\mathbf{X}) = p(\mu_i, \sigma_i^2|\mathbf{X}, \mathbf{y}, \beta_\gamma, \gamma, \sigma_\varepsilon),\ i = 1, ..., I.$$

5. Sample κ from $p(\kappa|\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}_\gamma, \boldsymbol{\gamma}, \kappa, \sigma_\varepsilon)$

This procedure is repeated from a random initialisation until convergence is complete (burn-in time). Following burn-in, samples from the chain can be used for Monte Carlo inference about the model $M_m \in \mathcal{M}$ or any other parameter of interest in the model. Full details of the posterior distributions required and the sampling steps summarised above will be given in an appendix .

6 Data Issues

The data is comprised of a panel dataset with a large amount of information collected over the period 1992-1999 for 49 countries. The selection of both the sample of countries and the period over which we conduct inference entails a number of trade-offs. For example, a larger sample of countries can provide more precise inference, but only if the parameters are stable across countries. Moreover, the number of countries is limited because countries need to be of a certain size before they have full access to international capital markets and thus can become vulnerable to a financial crisis. This suggests a compromise. In this study we utilise data on 49 medium and large countries that have access to international capital markets. See Table 2.

The time period for the analysis is 1992-99. An earlier starting point would provide more data for inference, but would call into question the implicit assumption of parameter stability, since by their very nature the causes and dynamics of crises evolve through time. Further, the purpose of this paper is to inform forward-looking policies. Finally, as a practical matter, much of the key data used in the analysis is available only for the 1990s.

The primary objective of our study is to examine the determinants of the *intensity* of crisis. To do this we use a binary indicator equal to one if a banking or currency crisis has occurred to select crisis observations. Using this selection mechanism we have approximately 42 crisis observations. Note that we treat these observations as a sample from a population defined by the presence of a financial crisis. Thus, our sample, covering the period 1992-99 will include crisis data points such as UK 1992, Argentina 1995, Mexico 1995 etc.

6.1 Definition and Measurement of Crisis intensity

Crisis intensity is gauged by the change in real GDP relative to the pre-crisis trend conditional on the occurrence of a crisis. The output shortfalls for these episodes are measured as the percentage deviation of actual GDP from its trend; the trend is calculated using a Hodrick-Prescott filter with standard parameter settings. Since the measurement of crisis intensity involves a duration component, another key specification issue is what duration to choose in the absence of a complete empirical macroeconomic model that would control for all the factors influencing output (Hoggarth et al., 2001). Measuring intensity using output data for the year of the crisis would seem too short, and there is also the problem of crises that start late in the year. On the other hand, using data for say four or five years after the onset of crisis would most likely introduce additional extraneous shocks. An alternative approach would be to employ a variable duration based on the number of post-crisis years for which GDP remained below trend. One problem with this approach is that a variable duration can introduce extraneous shocks and, moreover, raises difficult problems of defining explanatory indicators, e.g., should averages of indicators be used, or at the beginning of the crisis. For these reasons, the shortfall of output from trend for the crisis year and the following year was used. This approach introduces an extra source of measurement error at the gain of consistent definitions, and interpretations of explanatory indicators.[25]

Financial crises caused output to contract by 4 percent on average across the entire sample (see Figure 3). The average impact on output of crisis events for the period 1977-99 has been negative, except for four years, mostly covering industrial country crises. Generally, the most severe crises occurred during the mid-1980s, 1997, and to a lesser extent during the early 1990s.[26]

Crisis intensity varies widely with an average annual range of some 14 percent. Indonesia in 1997-98 experienced the largest output contraction of 30 percent. On average, developing countries are hit harder in comparison to industrial countries, and with a wider range of the crisis impact is wider. An exception are the developing countries in Europe who experience a less adverse and even a positive impact on output during the 1990s (mostly EMU crisis observations) probably reflecting their ties to industrial countries which are less prone to financial crises. The Asian country crises during 1993-99 had the most deleterious impact on output across the regions and time periods. Most of the other episodes ordered in this way had average contractions in the range -4 to -2 percent range.

6.2 Crisis Channel Indicators

The review of the theoretical literature as well as practical experience, suggests that it is possible to classify our set of crisis indicators according to the following crisis channels: the external sector, banking sector, corporate sector, collateral, financial breadth, foreign exchange liquidity, and the legal environment (see Table 1).

[25] We note that estimation with alternative definitions of duration suggest that the results are robust with respect to the term of the duration.

[26] Note that the differential number of crises per year distorts these annual averages.

7 Crisis Intensity Results

Crisis intensity results are presented in Tables 4–9. As noted, crisis intensity is measured as the accumulated deviation from trend GDP during the crisis year and the following year. We also reiterate that crisis events include either currency crises and banking crises. For each indicator, we present the mean of posterior distribution of the parameters, conditional on the indicator being included, namely

$$E(\beta_l|\gamma) = \sum_{j=1}^{2^I} \mathbf{1}(\gamma_{l,M_j} = 1) \cdot p(M_j|\mathbf{y})\widehat{\boldsymbol{\beta}}_{jl},$$

together with the 5th and 95th quantiles.

In addition, we report the posterior marginal probability of inclusion for each indicator, l, given by

$$p(\gamma_l|\mathbf{y}) = \sum_{j=1}^{2^I} \mathbf{1}(\gamma_{l,M_j} = 1) \cdot p(M_j|\mathbf{y}), \tag{22}$$

which simply sums the posterior model probabilities for each model in which the indicator appears. This particular posterior quantity represents a measure of *unconditional* indicator uncertainty, and as a consequence is an important objects for policy decisions. Namely, in identifying the first stage of conducting policy as deciding which indicators to collect data for, a measure of the relative importance of each indicator as represented by a scale free probability will be useful. We note that in the case of the full Bayesian approach these estimates are completely unconditional after integrating out both model and parameter uncertainty. For both the BIC and AIC approximations the marginal posterior inclusion probabilities are constructed unconditionally with respect to the model space, but do not reflect an account of parameter uncertainty.

We present our results ordered by the crisis channels as listed in Table 3. Note that in doing this we are able to gauge both the absolute importance of a given indicator, and the relative importance for an indicator within a particular channel. This will be of particular importance in the case of crisis channels, such as the Corporate Sector, where there exist a relative large number of measures.

In interpreting the results we consider the following policy problem. A policymaker is interested in both the relative importance of a candidate set of indicators of crises intensity, and the impact of any given indicator which is not conditional on any particular specified model. Based upon this observation, we need to consider the information present in both the magnitude of posterior inclusion probabilities relative to any prior beliefs; and the estimated effect of any given indicator on crises intensity. Therefore, in cases where we have specified a prior inclusion probability, following Sala-I-Martin, Doppelhofer, and Miller (2002), we evaluate marginal significance across indicators relative to a prior value. For example, by setting the expected model size to six, the prior inclusion probability (*pip*) is $6/27 = 0.222$. Although, we note that in the full Bayesian approach we have removed dependence on this fixed hyperparameter by utilising a Beta-Binomial prior, we still utilise *pip* as an informal benchmark with which to evaluate the relative magnitude of posterior inclusion probabilities. Second, it is important to interpret the joint information contained in the posterior density of a given indicator and the marginal probability of inclusion. In this context we may identify the following cases: (i) a credible interval which excludes zero, but with a relatively low marginal posterior inclusion probability (*mip*); (ii) a credible interval which excludes zero, but with a relatively high *mip*; (iii) a credible interval which contains zero, but with a relatively low marginal inclusion probability (*mip*); and (iii) a credible interval which contains zero, but with a relatively high marginal inclusion probability (*mip*). Cases (i) and (ii) require no additional explanation. However, we note that (iii) is likely to occur if a model contributes to the fit of a model, but is liable to switch signs due to collinearity.[27] Using a similar framework to evaluate the relative significance over individual indicators as conducted in Sala-I-Martin, Doppelhofer, and Miller (2002), we will refer to indicators as strong (s) if $mip > pip$ and the credible interval[28] does not include zero, and as marginal (m) if $pip > mip$ and the credible interval does not contain zero. The remaining indicators are classified as weak (w). Although we concede the abritrary nature of such a classification, it does facilitate comparison over the Bayesian, AIC, and BIC results.

In Table 4 we present the results from a full Bayesian analysis. The values for the fixed hyperparameters are provided as a footnote to the table. Across the crisis channels we note a large number posterior densities with credible intervals excluding zero, but for which the posterior marginal probability of inclusion is less than the prior inclusion probability. The legal environment is also seen as a crucial cause of crisis even though this concept is also difficult to define. Poor governance—reflecting lax shareholder rights, opaque accounting, and weak law enforcement—undermined the resiliency of the private sector to external shocks. In this instance, our measures of the legal environment, the Rule of Law and Antidirector Rights are reasonably precisely estimated, but have marginal probabilities that a quite low. Similarly indicators of Banking Sector problems all present credible intervals which do not contain zero, but with posterior inclusion probabilities ranging from 13 to 21%. For the External Sector we observe two strong indicators in the form of the ratio of balance of the current account to GDP and imports to GDP. Note also that the categorical variable indicating the level of development, is classified as marginal, which suggests that the remaining indicators are capturing much of the variation in crisis intensity.

The results for the Corporate Sector indicators are particularly interesting. Note that in this case we have a relatively large number of candidate indicators measuring corporate leverage and corporate liquidity. In this regard we note that there exists considerable prior uncertainty over the appropriate indicator. Although both the ratio of equity to total capital, and the quick ratio are both marginal, the current ratio - the ratio of total current assets to total current liabilities is by far the most important indicator within this crisis channel.

The relatively recent foreign exchange liquidity approach explicitly addresses joint currency and bank crisis dynamics arising from a shortfall of foreign exchange liquidity.

[27] Note that it may be possible to overcome this with a particular prior specification if such collinearity could be anticipated.

[28] Note that since Sala-I-Martin, Doppelhofer, and Miller (2002) utilise a Bayesian Averging of Classical Estimates (BACE) approach to model uncertainty, then the term credible interval does not directly apply.

Liquidity is defined as the difference between potential short-term obligations in foreign currency and the amount of accessible foreign currency in the consolidated financial system. Of the remaining crisis channels and indicators therein, the most notable result is the importance of one measure of foreign exchange liquidity, the change in private capital flows. In this respect the importance of the cutoff to capital inflows suggests that the magnitude of the crisis triggering shock is crucial to the output consequences of a financial crisis.

In tables five and six we present the results from the use of two 'approximations' to a full Bayesian approach to model uncertainty: AIC which estimates the relative KL distance across a M-open model space; and BIC based upon an approximation to full Bayesian assuming a M-closed model space. In Table 9 we summarise our findings across these three approaches, focussing on the magnitudes of the estimated *mip* (as represented in (22)) and the classification system discussed above.

A number of key observations can be made:

1. with a few notable exceptions there exists a high degree of correspondence between the estimated *mip* across all three approaches

2. in the case of BIC and AIC, the general ranking Akaike *mip* > BIC *mip* confirms previous findings (see, for example, Burnham and Anderson (1998)).

3. when an indicator is classified as strong by any of the three approaches it is similarly classified by the other two; the difference in the estimated *mip* across the approaches is generally low.

4. for indicators classified as either marginal - m (*pip* > *mip* and credible interval excludes zero) or weak - w (*pip* > *mip* and credible interval contains zero), there is almost 100% correspondence between AIC and BIC in terms of the classification, with again the estimated *mip* based upon AIC exceeds that for BIC.

   In this instance although the estimated *mip* for BIC and Bayesian are very similar, the classification predicted by the Bayesian approach is in a number of cases different than BIC and AIC

5. the most significant difference in estimated *mip* values appears for the indicator - Rating on Accounting Standards. Although all three approaches classify this indciator as weak, the Bayes approach predicts a high *mip* (0.606), whereas the predictions based on AIC (0.163) and BIC (0.123) are much lower.

8 Conclusion

This paper examined the intensity of financial crises during the 1990s with a view to improving crisis prevention and mitigation policies. The motivation was the new mandate for the IMF and World Bank to undertake comprehensive assessments of the vulnerability of the financial sectors of member countries, as well as the need for national policymakers to formulate mitigation policies. This paper aimed to extend the financial crisis empirical literature to help inform these assessments and policies. The results for crisis intensity indicate that the magnitude of the crisis-triggering shock may matter as much as the underlying balance sheet dynamics. The cutoff of private capital inflows, corporate balance sheet indicators, and to a lesser extent imports to GDP are the most robust indicators.

Our results have implications for assessments of the vulnerability of the financial sector to crises, as well as for broader economic policies. The importance of the capital inflow and import/GDP indicators highlights the importance of external sector adjustment in shaping the response of output to a financial crisis (Krugman (1999)). Thus, in forming crisis mitigation policies, e.g., countercyclical monetary and fiscal policy responses, governments should pay careful attention to the magnitude of the crisis-triggering cutoff of private capital inflows. In contrast to other studies, the legal indicators enter marginally, suggesting their influence independent of the other indicators is minimal. Similarly, the banking and financial breadth indicator results are mixed, indicating that they may be conduits of corporate distress and liquidity constraints, rather than have an independent role in crisis vulnerability.

References

Akaike, H. (1973): "Information Theory and an Extension of the Maximum Likelihood Principle," in *Proceedings of the 2nd International Symposium on Information Theory*, ed. by N. Petrov, and F. Csadki, pp. 267–281. Akademiai Kiado, Budapest.

Berg, A., E. Borensztein, G. M. Milesi-Ferretti, and C. Pattillo (1999): "Anticipating Balance of Payments Crises - The Role of Early Warning Systems," IMF Occasional Paper No. 186.

Bordo, M., B. Eichengreen, D. Klingebiel, and M. S. Martinez-Peria (2001): "Is the Crisis Problem Growing More Severe," *Economic Policy*.

Bozdogan, H. (1987): "Model Selection and Akaike's Information Criterion (AIC): The General Theory and its Analytical Extensions," *Psychometrika*, 52, 345–370.

Brock, W. A., S. N. Durlauf, and K. D. West (2003): "Policy Evaluation in Uncertain Economic Environments," unpublished paper.

Burnham, K. P., and D. R. Anderson (1998): *Model Selection and Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, New York.

Caballero, R., and A. Krishnamurthy (1999): "Emerging Markets Crisis - An Asset Markets Perspective," IMF Working Paper 99/129, Washington.

——— (2000): "International and Domestic Collateral Constraints in a Model of Emerging Markets Crises," mimeo.

Campos, J., and N. R. Ericsson (2000): "Constructive Data Mining: Modeling Consumers' Expenditure in Venezuela," Board of Governors of the Federal Reserve System, International Finance Discussion Papers No. 663,.

Carter, C., and R. Kohn (1996): "Markov Chain Monte Carlo in Conditionally Gaussian State Space Models," *Biometrika*, pp. 589–601.

Chang, R., and V. A. (1999): "Liquidity Crises in Emerging Markets: Theory and Policy," NBER Working Paper No. 7272, July.

Chipman, H. (1996): "Bayesian Variable Selection with Related Predictors," *Canadian Journal of Statistics*, 24, 17–36.

Cox, D. (1961): "Tests of Separate Families of Hypothesis," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*.

Davidson, J. E. H., and D. F. Hendry (1981): "Interpreting Econometric Evidence: Consumers' Expenditure in the UK," *European Economic Review*, 16, 177–192, Reprinted in Hendry, D. F. (1993), Econometrics: Alchemy or Science? Oxford: Blackwell Publishers.

Demirguc-Kunt, A., and E. Detragiache (1999a): "The Determinants of Banking Crises: Evidence from Developing and Developed Countries," *IMF Staff papers*, 45(1), 81–109, Washington: International Monetary Fund.

——— (1999b): "Monitoring Banking Sector Fragility - A Multivariate Logit Approach," IMF Working Paper 99/147.

Demirguc-Kunt, A., E. Detragiache, and P. Gupta (2000): "Inside the Crises - An Empirical Analysis of Banking Systems in Distress," IMF Working Paper 00/156, October (Washington: International Monetary Fund).

Eichengreen, B., A. Rose, and C. Wyplosz (1996): "Contagious Currency Crises," NBER No. 5681, July.

Eichengreen, B., and A. K. Rose (1998): "Staying Afloat When the Wind Shifts: External Factors and Emerging-Market Banking Crises," NBER Working Paper No. W6370, January.

Fernandez, C., E. Ley, and M. F. J. Steel (2001a): "Benchmark Priors for Bayesian Model Averaging," *Journal of Econometrics*, 100, 381–427.

Fernandez, C., E. Ley, and M. F. J. Steel (2001b): "Model Uncertainty in Cross-Country Growth Regressions," *Journal of Applied Econometrics*, 16.

George, E. I., and D. P. Foster (2000): "Calibration and Empirical Bayes Variable Selection," *Biometrika*, 87(4), 731–747.

George, E. I., and R. E. McCulloch (1995): "Stochastic Search Variable Selection," in *Practical Markov Chain Monte Carlo in Practice*, ed. by W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, pp. 203–14. Chapman and Hall, London.

Gertler, M., S. Gilchrist, and F. M. Natalucci (2000): "External Constraints on Monetary Policy and the Financial Accelerator," mimeo.

Granger, C., M. L. King, and H. White (1995): "Comments on Testing Economic Theories and the Use of Model Selection Criteria," *Journal of Econometrics*, 67, 173–187.

Green, P. J. (1995): "Reversible Jump Markov-Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, 82, 711–732.

Hurvich, C. M., and C. L. Tsai (1989): "Regression and Time Series Model Selection in Small Samples," *Biometrika*, 44, 297–307.

Kaminsky, G. L., and C. M. Reinhart (1999): "The Twin Crises: The Causes of Banking and Balance-of-Payments Problems," *The American Economic Review*, 89(3), 473–500.

Kass, R. E., and A. E. Raftery (1995): "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–795.

Kass, R. E., and A. E. Raftery (1995): "Bayes Factors," *Journal of the American Statistical Association*, 90(430), 773–795.

Kiyotaki, B., and J. Moore (1997): "Credit Cycles," *Journal of Political Economy*.

Krolzig, H. M., and D. F. Hendry (2000): "Computer Automation of General-to-Specific Model Selection Procedures," Working Paper, Institute of Economics and Statistics, Oxford.

Krugman, P. (1979): "A Model of Balance-of-Payments Crises," *Journal of Money, Credit and Banking*, 11(3), 311–25.

——— (1999): "Analytical Afterthoughts on the Asian Crisis," web.mit.edu/krugman/www/MINICRIS.htm, September.

Kuo, L., and B. Mallick (1997): "Variable Selection for Regression Models," *Sankhya*.

Mulder, C., R. Perrelli, and M. Rocha (2001): "The Role of Corporate, Legal and Macro Balance Sheet Indicators in Crisis Detection and Prevention," mimeo, International Monetary Fund.

Obstfeld, M. (1994): "Logic of Currency Crises," National Bureau of Economic Resarch Working Paper Series (U.S.), No. 4640, February.

Pesaran, H., and M. Weeks (2000): "Non-Nested Hypothesis Tests," in *Theoretical Econometrics*, ed. by B. Baltagi. Basil Blackwell, Oxford.

Poskitt, D. S., and A. R. Tremayne (1983): "On the Posterior Odds of Time Series Models," *Biometrika*, 70, 157–62.

Prentice, R. L. (1986): "Binary Regression Using an Extended Beta-Binomial Distribution, with Discussion of Correlation Induced by Covariate Measurement Errors," *Journal of the American Statistical Association*, 81(394).

Sala-I-Martin, X., G. Doppelhofer, and R. Miller (2002): "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach," Faculty of Economics, University of Cambridge, UK.

Schwarz, G. (1978): "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464.

Sigiura, N. (1978): "Further Analysis of the Data by Akaike's Information Criterion and the Finite Corrections," *Communications in Statistics, Theory and Methods*, A7, 13–26.

Smith, M., and R. Kohn (1996): "Nonparametric Regression Using Bayesian Variable Selection," *Journal Econometrics*, 75, 317–44.

Stone, M. (2000): "The Corporate Sector Dynamics of Systemic Financial Crises," IMF Working Paper No. 00/114 (Washington: International Monetary Fund).

Swanson, N. R., and H. White (1995): "A Model-Selection Approach to Assessing the Information in the Term Structure Using Linear Models and Artificial Neural Networks," *Journal of Business and Economic Statistics*, 13(3), 265–275.

Velasco, A. (1987): "Financial Crises and Balance of Payments Crises: A Simple Model of the Southern Cone Experience," *Journal of Development Economics*, 27(1-2), 263–83.

WEEKS, M., AND M. STONE (2001): "Systemic Financial Crises, Balance Sheets, and Model Uncertainity," Working Paper No. 01/162, International Monetary Fund, Washington DC.

WHITE, H. (1999): "A Reality Check for Data Snooping," mimeo, Department of Economics, University of California at San Diego, La Jolla, California.

ZELLNER, A. (1986): "On Assessing Prior Distributions and Bayesian Regression Analysis with G-Prior Distributions," in *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finietti*, ed. by P. K. Goel, and A. Zellner, pp. 233–43. North-Holland, Amsterdam.

Data Sources:

- Beck, Thorsten, Asli Memirguc-Kunt and Ross Levine (1999), “A new database on financial development and structure”, World Bank Policy Research Working Paper No. 2146.
- La Porta, Rafael, Florencio Lopez-de-Silane, Andrei Shleifer, and Robert W. Vishny (1996), “Law and Finance”, NBER Working Paper No. 5661, July.
- BIS, 2000, Joint BIS-IMF-OECD-World Bank Statistics on External Debt, http://www.bis.org/publ/index.htm.

Table 1: Data Sources

Indicators	*Sources*
External Indicators	
Imports to GDP	WEO
Real Effective Exchange Rate, deviation from HP trend	IFS
LIBOR	WEO
Net liabilities of nonbanks resident in BIS reporting countries to GDP	BIS
Change in Current Account to GDP	WEO
Banking Sector Indicators	
4-Year Change in Private Credit to GDP	IFS/WEO
Domestic credit to GDP	IFS
Broad money to GDP	IFS
Corporate Sector Indicators	
Total debt to Common equity	Worldscope
Equity to Total capital	Worldscope
Current ratio: total current assets/total current liabilities	Worldscope
Working capital to Total capital	Worldscope
Quick ratio: Cash & equivalents + receivable net/total current liabilities	Worldscope
Total debt to Total assets	Worldscope
Financial Breadth Indicators	
Financial Breadth 1: ratio of outstanding bonds (national corporations) to bank credit	WEO/Beck et. al.
Financial Breadth 2: Ratio of outstanding bonds (national corporations) + stock	WEO/Beck et. al.
Private bond market capitalisation to GDP	WEO/Beck et. al.
LT Debt to Common equity	Beck. et. al.
LT Debt to Total capital	Worldscope
Foreign Exchange Indicators	
Broad money/International reserves	IFS
Change in capital flows to GDP	WEO

List of Indicators continued

Indicators	Sources
Legal Environment Indicators	
Antidirectors Rights	La Porta
Rule of Law	La Porta
-Rating on Accounting Standards (Rat AccSt)	
Other Indicators	
Annual Average CPI Inflation	IFS
Income development indicator	Beck. et al.
Real GDP/Hodrick_Prescott trends	IFS
Real interest rate	IFS
Contagion	-

Table 2: List of Countries

Argentina	Hungary	Poland
Australia	India	Portugal
Austria	Indonesia	Russia
Belgium	Ireland	Singapore
Brazil	Israel	South Africa
Canada	Italy	Spain
Chile	Japan	Sri Lanka
China	Jordan	Sweden
Colombia	Korea, Rep.	Switzerland
Czech Republic	Malaysia	Thailand
Denmark	Mexico	Turkey
Egypt, Arab. Rep.	Netherlands	United Kingdom
Finland	New Zealand	United States
France	Norway	Venezuela
Germany	Pakistan	Zimbabwe
Greece	Peru	
Hong Kong	Phillippines	

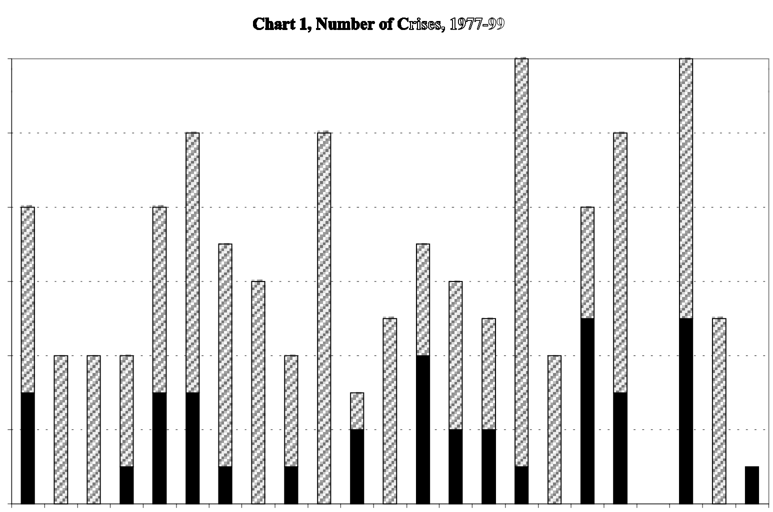

Figure 1:

Table 3: Crisis Channel Indicator Groups

External Sector
Imports to GDP (Imp/GDP) +
Real effective exchange rate, deviation from trend (REER) +
LIBOR +
Net liabilities of nonbanks to banks residents in BIS reporting countries to GDP (Ext Liabs) +
Current account balance to GDP (CA/GDP) +
Banking sector
Leading four year change in private credit to GDP (Cred Gr Cred/GDP) +
Domestic credit to GDP (Cred/GDP) + or -
Broad money to GDP (Mon/GDP) -
Corporate sector
Total debt to common equity (TotDCE) +
Equity to total capital (EqTC) -
Current ratio - Ratio of total current assets to total current liabilities (CurrR) -
Working capital to total capital (WCapTC) -
Long-term debt to common equity (LTDCE) +
Quick ratio - Ratio of cash and equivalent plus net receivables to total current liabilities (Quick R)
Long-term debt to total capital (LTDTC) +
Financial Breadth
Ratio of outstanding bonds plus stock market capitalisation to bank credit (FBr2) -
Private bond market capitalisation to GDP (PBM_gdp) -
Foreign Exchange Liquidity
Broad money/International reserves (Broad$) +
Change in capital inflows (ChPC/GDP) -
Legal Environment
Rule of Law (ROL) -
Antidirectors Rights (AntiDirR) -
Rating on Accounting Standards (Rat AccSt)
Other
Annual average CPI inflation (CPI Inf) +
Income development (1 = high income,..., 4 = low income) (IdevI) -
Real GDP deviation from trend (R_GDPhpt) +

Note : Signs indicate hypothesised impact of each indicator on crisis probability (the impact on crisis intensity is of the opposite sign).

Table 4: Full Bayesian

	Posterior mean	5% Q	95% Q	PPIncl.	% missing	Class
Legal Environment						
Rule of Law	0.140	0.115	0.817	0.068	0.156	m
Antidirector rights	0.471	0.475	1.135	0.058	0.156	m
Rating on Accounting Standards	-0.018	-0.003	0.040	0.606	0.289	w
Banking Sector						
4 year change in pc to GDP	-0.034	-0.035	-0.005	0.095	0.044	m
Domestic Credit to GDP	-2.994	-3.182	-0.419	0.142	0.000	m
Broad money to GDP	-3.805	-4.345	-0.537	0.212	0.000	s
External Sector						
Nbank L BIS	0.118	0.116	0.171	0.131	0.022	m
Imports to GDP	0.152	0.151	0.186	0.680	0.000	s
90day LIBOR	0.097	0.051	1.588	0.047	0.000	m
Net Nb BR	-0.047	-0.048	0.159	0.054	0.000	w
Current account to GDP	0.584	0.608	0.867	0.235	0.000	s
Trade Balance to GDP	0.050	0.078	0.255	0.058	0.000	m
Terms of Trade	-0.002	-0.002	0.091	0.046	0.000	w
Annual Average CPI Inflation	0.002	0.002	0.003	0.053	0.000	m
Income Development	0.086	0.177	1.481	0.070	0.000	m
Corporate Sector						
Total debt to Common Equity	-2.773	-2.351	-1.219	0.296	0.000	s
Equity to Total Capital	6.935	7.317	13.762	0.099	0.000	m
Current Ratio	9.860	9.885	11.904	0.820	0.000	s
Working Capital/Total Capital	-1.817	-2.056	0.692	0.073	0.022	w
Long term debt to common equity	-0.371	-1.012	2.658	0.087	0.000	w
Quick Ratio	6.848	7.588	11.026	0.185	0.000	m
Long term Debt. to Total Capital	1.774	2.403	5.526	0.099	0.000	m
Financial Breadth						
FBr2	1.497	1.675	3.077	0.118	0.022	m
Private Bond Mt. Cap/GDP	-0.004	-0.004	0.044	0.049	0.000	w
Foreign Exchange Liquidity						
Broad money	-1.458	-1.646	0.581	0.115	0.000	w
Change in pc flows to GDP	0.772	0.772	0.925	0.963	0.000	s
Contagion	-0.127	-0.098	0.147	0.050	0.000	w

Notes:
Number of Monte Carlo Iterations (Burnin) 30000 (10000)
Average Posterior Model Size = 6.464
IG prior on $\sigma_\varepsilon : \alpha_\varepsilon/\beta_\varepsilon = 1e-010/1e-010$
G-Prior for Covariance Matrix of beta.
g-prior type parameter (κ) = 0.0222222=(l/n)

Table 5: Akaike Small Sample Adjustment

	Posterior mean	5% Q	95% Q	PPIncl.	% missing	Class
Legal Environment						
Rule of Law	0.389	-0.272	0.998	0.095	0.156	w
Antidirector rights	0.569	-0.321	1.268	0.088	0.156	w
Rating on Accounting Standards	-0.036	-0.193	0.123	0.163	0.289	w
Banking Sector						
4 year change in pc to GDP	-0.039	-0.067	-0.011	0.142	0.044	m
Domestic Credit to GDP	-2.678	-5.032	0.135	0.148	0.000	w
Broad money to GDP	-4.174	-5.859	-1.319	0.272	0.000	s
External Sector						
Nbank L BIS	0.128	0.075	0.195	0.157	0.022	m
Imports to GDP	0.153	0.105	0.188	0.637	0.000	s
90day LIBOR	0.558	-0.572	1.753	0.071	0.000	w
Net Nb BR	-0.138	-0.305	-0.005	0.090	0.000	m
Current account to GDP	0.658	0.461	0.906	0.499	0.000	s
Trade Balance to GDP	-0.069	-0.387	0.235	0.096	0.000	w
Terms of Trade	-0.019	-0.100	0.079	0.068	0.000	w
Annual Average CPI Inflation	0.002	0.000	0.004	0.082	0.000	m
Income Development	-0.456	-1.580	0.629	0.075	0.000	w
Corporate Sector						
Total debt to Common Equity	-3.621	-6.862	-1.361	0.329	0.000	s
Equity to Total Capital	9.830	2.950	19.989	0.181	0.000	m
Current Ratio	9.473	7.341	11.107	0.848	0.000	s
Working Capital/Total Capital	-1.468	-3.382	1.558	0.093	0.022	w
Long term debt to common equity	0.145	-2.217	4.275	0.113	0.000	w
Quick Ratio	5.968	1.080	10.214	0.185	0.000	m
Long term Debt. to Total Capital	3.392	-1.336	6.377	0.176	0.000	w
Financial Breadth						
FBR2	1.322	0.266	2.832	0.132	0.022	m
Private Bond Mt. Cap/GDP	0.005	-0.049	0.052	0.072	0.000	w
Foreign Exchange Liquidity						
Broad money	-0.556	-1.766	1.075	0.087	0.000	w
Change in pc flows to GDP	0.786	0.603	0.930	0.978	0.000	s
Contagion	-0.205	-0.515	-0.005	0.077	0.000	w

Notes:
Number of Monte Carlo Iterations (Burnin) 30000 (10000)
Prior Model Size = 6
Average Model Size = 6.754

Table 6: Bayesian Information Criterion

	Posterior mean	5% Q	95% Q	PPIncl.	% missing	Class
Legal Environment						
Rule of Law	0.384	-0.317	1.026	0.065	0.156	*w*
Antidirector rights	0.556	-0.296	1.240	0.057	0.156	*w*
Rating on Accounting Standards	-0.047	-0.198	0.121	0.123	0.289	*w*
Banking Sector						
4 year change in pc to GDP	-0.043	-0.073	-0.014	0.110	0.044	*m*
Domestic Credit to GDP	-2.994	-4.983	-0.335	0.129	0.000	*m*
Broad money to GDP	-4.302	-5.738	-1.807	0.235	0.000	*s*
External Sector						
Nbank L BIS	0.126	0.072	0.193	0.101	0.022	*m*
Imports to GDP	0.161	0.112	0.190	0.605	0.000	*s*
90day LIBOR	0.638	-0.529	1.875	0.049	0.000	*w*
Net Nb BR	-0.134	-0.300	-0.005	0.059	0.000	*m*
Current account to GDP	0.669	0.476	0.886	0.411	0.000	*s*
Trade Balance to GDP	-0.001	-0.371	0.258	0.066	0.000	*w*
Terms of Trade	-0.025	-0.106	0.084	0.046	0.000	*w*
Annual Average CPI Inflation	0.002	0.000	0.004	0.053	0.000	*m*
Income Development	-0.494	-1.651	0.661	0.054	0.000	*w*
Corporate Sector						
Total debt to Common Equity	-3.295	-6.603	-1.398	0.247	0.000	*s*
Equity to Total Capital	9.682	3.109	19.676	0.138	0.000	*m*
Current Ratio	9.547	7.5349	11.125	0.796	0.000	*s*
Working Capital/Total Capital	-1.566	-3.375	1.427	0.063	0.022	*w*
Long term debt to common equity	-0.259	-2.123	3.384	0.082	0.000	*w*
Quick Ratio	7.008	1.650	10.625	0.191	0.000	*m*
Long term Debt. to Total Capital	2.947	-1.445	6.156	0.113	0.000	w
Financial Breadth						
FBR2	1.399	0.266	2.977	0.093	0.022	*m*
Private Bond Mt. Cap/GDP	0.003	-0.056	0.052	0.049	0.000	*w*
Foreign Exchange Liquidity						
Broad money	-0.703	-1.755	1.010	0.059	0.000	*w*
Change in pc flows to GDP	0.789	0.603	0.931	0.955	0.000	*s*
Contagion	-0.215	-0.525	-0.008	0.052	0.000	*m*

Notes:
Number of Monte Carlo Iterations (Burnin) 30000 (10000)
Prior Model Size = 6
Average Model Size = 6.041

Table 7: Full Bayesian with Time and Country Effects

	Posterior mean	5% Q	95% Q	PPIncl	% missing	Class
Legal Environment						
Rule of Law	0.163	0.182	0.767	0.057	0.156	*m*
Antidirector rights	0.503	0.523	1.094	0.057	0.156	*m*
Rating on Accounting Standards	-0.022	-0.006	0.048	0.552	0.289	*w*
Banking Sector						
4 year change in pc to GDP	-0.038	-0.037	-0.012	0.084	0.044	*m*
Domestic Credit to GDP	-3.149	-3.451	-0.362	0.129	0.000	*m*
Broad money to GDP	-4.115	-4.621	-1.499	0.227	0.000	*s*
External Sector						
Nbank L BIS	0.106	0.107	0.163	0.107	0.022	*m*
Imports to GDP	0.155	0.156	0.189	0.631	0.000	*s*
90day LIBOR	-0.220	-0.174	0.217	0.051	0.000	*m*
Net Nb BR	-0.035	-0.017	0.124	0.072	0.000	*w*
Current account to GDP	0.612	0.647	0.843	0.240	0.000	*s*
Trade Balance to GDP	0.043	0.096	0.287	0.046	0.000	*m*
Terms of Trade	-0.029	-0.028	0.049	0.049	0.000	*w*
Annual Average CPI Inflation	0.002	0.002	0.003	0.058	0.000	*m*
Income Development	0.232	0.403	1.520	0.064	0.000	*m*
Corporate Sector						
Total debt to Common Equity	-2.822	-2.280	-1.161	0.254	0.000	*s*
Equity to Total Capital	8.109	7.939	17.198	0.113	0.000	*m*
Current Ratio	9.748	9.828	11.754	0.775	0.000	*s*
Working Capital/Total Capital	-1.602	-2.004	1.891	0.067	0.022	*w*
Long term debt to common equity	-0.112	-0.696	2.857	0.087	0.000	*w*
Quick Ratio	6.779	7.374	11.573	0.184	0.000	*m*
Long term Debt. to Total Capital	2.231	3.010	5.830	0.087	0.000	*m*
Financial Breadth						
FBR2	1.770	1.828	3.530	0.116	0.022	*m*
Private Bond Mt. Cap/GDP	-0.002	-0.005	0.044	0.057	0.000	*w*
Foreign Exchange Liquidity						
Broad money	-1.372	-1.572	0.040	0.100	0.000	*w*
Change in pc flows to GDP	0.764	0.766	0.919	0.951	0.000	*s*
Contagion	-0.111	-0.093	0.158	0.056	0.000	*w*
Industrial Countries	1.427	1.566	3.540	0.057	0.000	*m*
Year 92-95	3.357	2.971	6.535	0.083	0.000	*m*

Notes:
Number of Monte Carlo Iterations (Burnin) 30000 (10000)
Average Posterior Model Size =5.7663
Prior Model Size = 6
IG prior on $\sigma_\varepsilon : \alpha_\varepsilon/\beta_\varepsilon = 1e-010/1e-010$
G-Prior for Covariance Matrix of beta.
g-prior type parameter (κ) = 0.0222222=(1/n)

Table 8: Akaike Small Sample Adjustment with Time and Country Effects

	Posterior mean	5% Q	95% Q	PPIncl	% missing	Class
Legal Environment						
Rule of Law	0.292	0.347	0.769	0.069	0.156	*m*
Antidirector rights	0.640	0.647	1.484	0.070	0.156	*m*
Rating on Accounting Standards	-0.143	-0.170	0.082	0.301	0.289	*w*
Banking Sector						
4 year change in pc to GDP	-0.041	-0.041	-0.012	0.108	0.044	*m*
Domestic Credit to GDP	-3.061	-3.750	-0.135	0.100	0.000	*m*
Broad money to GDP	-4.284	-4.781	-1.487	0.228	0.000	*s*
External Sector						
Nbank L BIS	0.110	0.110	0.173	0.091	0.022	*m*
Imports to GDP	0.160	0.162	0.194	0.557	0.000	*s*
90day LIBOR	-0.153	-0.132	0.630	0.053	0.000	*m*
Net Nb BR	-0.070	-0.104	0.197	0.055	0.000	*w*
Current account to GDP	0.674	0.713	0.911	0.366	0.000	*s*
Trade Balance to GDP	0.025	0.068	0.275	0.058	0.000	*m*
Terms of Trade	-0.016	0.019	0.140	0.035	0.000	*w*
Annual Average CPI Inflation	0.002	0.002	0.004	0.053	0.000	*m*
Income Development	-0.306	-0.325	0.837	0.041	0.000	*m*
Corporate Sector						
Total debt to Common Equity	-3.213	-2.277	-1.154	0.211	0.000	*s*
Equity to Total Capital	9.391	9.247	17.981	0.121	0.000	*m*
Current Ratio	9.801	9.750	12.464	0.780	0.000	*s*
Working Capital/Total Capital	-1.805	-2.210	0.189	0.065	0.022	*w*
Long term debt to common equity	0.547	-0.578	4.889	0.084	0.000	*w*
Quick Ratio	6.514	7.169	10.241	0.168	0.000	*m*
Long term Debt. to Total Capital	2.603	3.467	6.396	0.085	0.000	*m*
Financial Breadth						
FBR2	1.864	1.710	3.488	0.141	0.022	*m*
Private Bond Mt. Cap/GDP	0.004	0.003	0.055	0.049	0.000	*w*
Foreign Exchange Liquidity						
Broad money	-0.555	0.620	1.266	0.066	0.000	*w*
Change in pc flows to GDP	0.784	0.787	0.934	0.964	0.000	*s*
Contagion	-0.214	0.156	0.074	0.054	0.000	*w*
Industrial Countries	2.016	2.183	4.536	0.062	0.000	*m*
Year 92-95	3.514	3.671	6.172	0.093	0.000	*m*

Notes:
Number of Monte Carlo Iterations (Burnin) 30000 (0)
Average Posterior Model Size =5.4972
Prior Model Size = 6

Table 9: A Comparison of Posterior Probabilities of Indicator Inclusion

	Full Bayesian		Akaike Small Sample Adjustment		Bayesian Information Criterion		% Missing
	PPIncl.	Class	PPIncl	Class	PPIncl.	Class	
Legal Environment							
Rule of Law	0.068	*m*	0.095	*w*	0.065	*w*	0.156
Antidirector rights	0.058	*m*	0.088	*w*	0.057	*w*	0.156
Rating on Accounting Standards	0.606	*w*	0.163	*w*	0.123	*w*	0.289
Banking Sector							
4 year change in pc to GDP	0.095	*m*	0.142	*m*	0.110	*m*	0.044
Domestic Credit to GDP	0.142	*m*	0.148	*w*	0.129	*m*	0.000
Broad money to GDP	0.212	*s*	0.272	*s*	0.235	*s*	0.000
External Sector							
Nbank L BIS	0.131	*m*	0.157	*m*	0.101	*m*	0.022
Imports to GDP	0.680	*s*	0.637	*s*	0.605	*s*	0.000
90day LIBOR	0.047	*m*	0.071	*w*	0.049	*w*	0.000
Net Nb BR	0.054	*w*	0.090	*m*	0.059	*m*	0.000
Current account to GDP	0.235	*s*	0.499	*s*	0.411	*s*	0.000
Trade Balance to GDP	0.058	*m*	0.096	*w*	0.066	*w*	0.000
Terms of Trade	0.046	*w*	0.068	*w*	0.046	*w*	0.000
Annual Average CPI Inflation	0.053	*m*	0.082	*m*	0.053	*m*	0.000
Income Development	0.070	*m*	0.075	*w*	0.054	*w*	0.000
Corporate Sector							
Total debt to Common Equity	0.296	*s*	0.329	*s*	0.247	*s*	0.000
Equity to Total Capital	0.099	*m*	0.181	*m*	0.138	*m*	0.000
Current Ratio	0.820	*s*	0.848	*s*	0.796	*s*	0.000
Working Capital/Total Capital	0.073	*w*	0.093	*w*	0.063	*w*	0.022
Long term debt to common equity	0.087	*w*	0.113	*w*	0.082	*w*	0.000
Quick Ratio	0.185	*m*	0.185	*m*	0.191	*m*	0.000
Long term Debt. to Total Capital	0.099	*m*	0.176	*w*	0.113	*w*	0.000
Financial Breadth							
FBr2	0.118	*m*	0.132	*m*	0.093	*m*	0.022
Private Bond Mt. Cap/GDP	0.049	*w*	0.072	*w*	0.049	*w*	0.000
Foreign Exchange Liquidity							
Broad money	0.115	*w*	0.087	*w*	0.059	*w*	0.000
Change in pc flows to GDP	0.963	*s*	0.978	*s*	0.955	*s*	0.000
Contagion	0.050	*w*	0.077	*w*	0.052	*m*	0.000

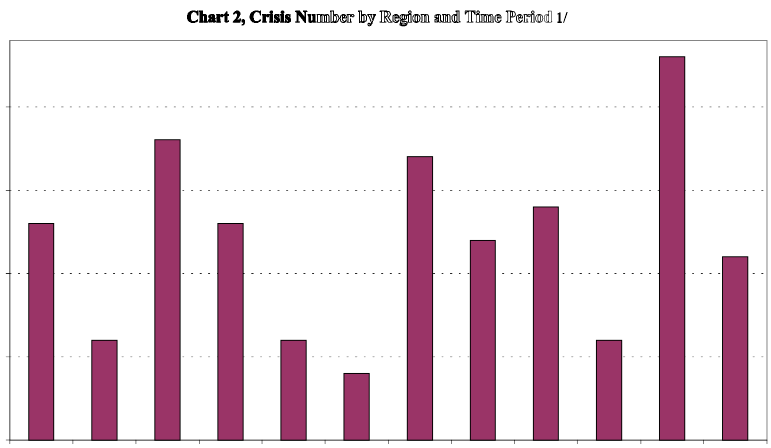

Figure 2:

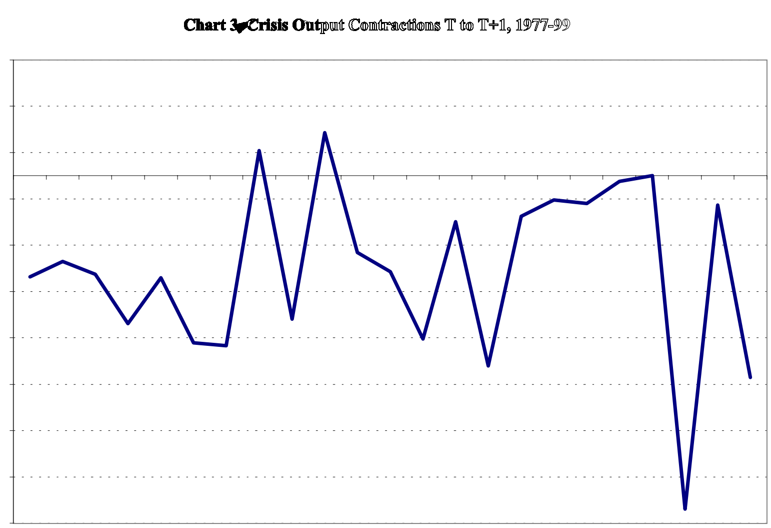

Figure 3: