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

PRELIMINARY AND INCOMPLETE
(Do not Quote)

August 29, 2003

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 encompasses 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 with a view to improving crisis prevention and mitigation policies. 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 cri-

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

sis indicators or composite measures of vulnerability to provide relatively quick warning signals of impending crises to trigger countervailing adjustments in macroeconomic policies. The goal of this paper, rather, is to enhance the effectiveness of annual or even less frequent assessments of crisis vulnerability. These assessments can be used to ameliorate crisis vulnerability by motivating structural reforms. The paper makes two contributions to the literature.

This paper is organized as follows. The theoretical and empirical literature on crisis intensity is reviewed in section two. The empirical methodology used in this paper is described in section three, and in section four we discuss the data and specification issues. Crisis intensity results are presented in section five. Section six concludes with a summary of the results and their policy implications.

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 Velasco, 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 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 and Patillo,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; Demirgüç-Kunt and Detragiache,1999a and b). 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. Demirgüç-Kunt and Detragiache (1999a) emphasize vulnerability to large capital inflows, bank deposit insurance, and the legal system. Demirgüç-Kunt et al., (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 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 et al., (2000) 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 Econometric Methodology

There are a number of fundamental characteristics of the problem. First, 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 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 have been advocated by a number of analysts. For example, Granger, King, and White (1995) advocate on 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 the assumption that one of the models considered is the true model.

The problem 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 Ω, in this instance crisis channel indicators. Given Ω, model selection proceeds by searching over the model space implied by Ω (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.[2][3] The principle advantages of this approach are that model selection 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 of this approach.[4] 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.[5][6]

In this study we evaluate the degree of model uncertainty using both information-theoretic and Bayesian methodologies. 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 model selection 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 as sample size increases, and assumed to lie within the candidate set of models. In this respect the 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[7]. Now elaborate with regards this particular problem for probability and intensity results ...

ADD: discussion from Bernardo and Smith and Durlauf re: Model-closed and Model-open inference on model uncertainty.

[2] An important caveat here is that the process of model selection is obviously conditional upon prior identification of an initial indicator set, Ω.

[3] 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.

[4] 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).

[5] 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

[6] 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.

[7] 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 The PolicyMakers Problem

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 define.

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. In the absence of any constraints the total number of possible models is 2^I. 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, $\boldsymbol{\omega}^{(1)} = (x_1^1, x_2^1, ..., x_{l_1}^1)'$ 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, or completely impotent, in selecting, for example, the appropriate indicators *within* $\omega^{(1)}$. Given a large set of indicators, $\mathcal{K}$, with the objective of identifying a smaller subset of indicators, say $\mathcal{K}_s$, and obtaining some measure of the unconditional importance of each indicator as a determinant of crisis. For now it is assumed that $\mathcal{K}_s$ is constant across countries and time. At a later point we can explore whether $\mathcal{K}$ is time dependent i.e. differs across pre-identified periods and/or differs in terms of emerging market versus industrialized economies.

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

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

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 $\{y_j\}$. We can 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:

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

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 length I vector of binary indicator variables which we use to index the 2^I distinct models in $\mathcal{M}$ and $\mathbf{1}$ 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

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

where $p(M_h)$ denotes the prior probability for model M_h. Note that in (1) all uncertainty over $\boldsymbol{\theta}_j = (\sigma_\varepsilon, \boldsymbol{\beta}_j, \alpha)'$ has been integrated out such that $l(\mathbf{y}|M_h)$ represents the *marginal* likelihood of model M_h, given by

$$l(\mathbf{y}|M_h) = \int l(\mathbf{y}|\boldsymbol{\theta}_h, M_h)p(\boldsymbol{\theta}_h|M_h)d\boldsymbol{\theta}_h. \tag{2}$$

$p(\boldsymbol{\theta}_h|M_h)$ denotes the prior for the parameters of model M_h. Using (1) the ratio of the posterior probability for model h and h', say $B_{hh'}$, is therefore given by

$$B_{hh'} = \frac{p(M_h|\mathbf{y})}{p(M_{h'}|\mathbf{y})} = \frac{l(\mathbf{y}|M_h)p(M_h)}{\underbrace{l(\mathbf{y}|M_{h'})}_{A}p(M_{h'})}. \tag{3}$$

Making the assumption of equal prior odds, the Bayes factor, given by the ratio A in (3), 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 $\boldsymbol{\theta}$; 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 (2). Below we examine how both of these requirements may be operationa;ised, but prior to this we first consider two alternative measures of (3) based on an asymptotic Bayesian approximation and an approximately unbiased estimator of the relative K-L distance.

Schwarz (1978) developed an approximation to a fully Bayesian approach to model uncertainty based upon asymptotic arguments. Predicated on the notion that the true unknown model can be specified and is contained within $\mathcal{M}$, a BIC approach to model uncertainty is consistent, in the sense that the dimension of the model is fixed, and that the probability of locating the model with the highest posterior probability approaches one as sample size increases.[8] Assuming uniform priors over $\mathcal{M}$, and vague priors over $\boldsymbol{\theta}_h$, an asymptotic approximation to the log of the posterior odds for models M_h and $M_{h'}$ (say $B^S_{hh'}$) provided by the Schwarz criterion, is given by

$$B^S_{hh'} = \log\left(\frac{l(\hat{\boldsymbol{\theta}}_h|\mathbf{y}, M_h)}{l(\hat{\boldsymbol{\theta}}_{h'}|\mathbf{y}, M_{h'})}\right) - c \tag{4}$$

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

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

The Bayesian Information Criterion (BIC) is equal to minus twice the Schwarz criterion.

Akaike (1973) proposed an alternative method of model selection based upon the concept of expected information distance. In contrast to the Schwarz criterion, Akaike's Information Criterion (AIC) is based upon the notion that truth cannot be represented in a model form, with $\mathcal{M}$ containing a number of *approximations* to the truth. In this respect AIC should be 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 $I(\mathfrak{m}^*, h)$. Akaike's information criterion (AIC) for M_h is generally written as

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

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]

Akaike (1983) demonstrated that

$$l(M_h|\mathbf{y}) \propto \exp(-\frac{1}{2}AIC_h) = cl(\widehat{\boldsymbol{\theta}}|\mathbf{y}, M_h)e^{-k_h}, \tag{8}$$

where c is an arbitrary constant, and suggests $\exp(-\frac{1}{2}AIC_h)$ as being the *relative* likelihood (or probability) of the model given the data. Once again assuming a uniform prior over all models in $\mathcal{M}$, the use of a simple normalisation facilitates an approximation[10] to the

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

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

[10] Note approximation but not in the sense of a limiting relationship between AIC and fully Bayesian approach to model selection as in ()

posterior probability for M_h, namely

$$p_{AIC}(M_h|\mathbf{y}) = \frac{\exp(-\frac{1}{2}AIC_h)}{\sum_{i=1}^{I}\exp(-\frac{1}{2}AIC_i)}, \tag{9}$$

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

$$B_{hh'}^{AIC} = \frac{\exp(-\frac{1}{2}AIC_h)}{\exp(-\frac{1}{2}AIC_{h'})}. \tag{10}$$

We note that for both the AIC and BIC selection criteria we may also consider a set of non-uniform *prior* probabilities over $\mathcal{M}$.[11] For example, extending (9) the posterior probability for M_h is given by

$$p_{AIC}(M_h|\mathbf{y}) = \frac{\exp(-\frac{1}{2}AIC_h)p(M_h)}{\sum_{i=1}^{I}\exp(-\frac{1}{2}AIC_m)p(M_i)} \tag{11}$$

4 Specification of Prior Distributions

The principal unknown parameters in the above model are $\boldsymbol{\beta}_\gamma$, $\boldsymbol{\gamma}$, α and σ_ε. In addition there are several unknown hyperparameters and any 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 we can gain the advantage of properly incorporating prior uncertainty about unknowns. Appropriate specification of priors is thus a crucial part of the procedure.

In general posterior inference *conditional* on a given model is less sensitive to prior specifications than in cases where model uncertainty is incorporated. (see, for example, 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.[12] For any 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. 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

[11] 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.

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

one or more parameters over model-invariant and model-specific parameters (see Jeffreys (1961)).

The chosen prior structure here may be factorised as follows

$$p(\boldsymbol{\beta}_\gamma, \boldsymbol{\gamma}, \alpha, \sigma_\epsilon, \kappa, \mathbf{X}, \boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X) = p(\boldsymbol{\beta}_\gamma|\boldsymbol{\gamma},\kappa, \mathbf{X})p(\boldsymbol{\gamma})p(\alpha)p(\sigma_\epsilon)p(\kappa)p(\mathbf{X}|\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X)p(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X), \tag{12}$$

where the covariates $\mathbf{X}$ are included directly within this prior framework thereby facilitating MCMC imputation of missing covariates. κ is a hyperparemeter used in the g-prior specification for $\boldsymbol{\beta}_\gamma$. For the sake of clarity in (12) we do not explicitly write out dependence of distributions on any *fixed* constant hyperparameters, 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

$$p(\boldsymbol{\beta}_\gamma|\boldsymbol{\gamma}, \kappa) = MVN_\gamma(\mathbf{0}, \boldsymbol{\Sigma}_\gamma), \tag{13}$$

where $\boldsymbol{\Sigma}_\gamma = \sigma_\varepsilon^2 \kappa(\mathbf{X}_\gamma' \mathbf{X}_\gamma)^{-1}$ for $\kappa > 0$, which 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 structure of the prior reflects an adjustment (by the factor κ) of the covariance structure in the likelihood, 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 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. We also note that George and Foster in comparing Bayesian methods for evaluating model uncertainty with approximations based on information criteria, indicate that AIC corresponds with setting $\kappa = 0.255$ and for BIC $\kappa = 1/n$. Further discussion in George and Foster (2000).

An alternative approach 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, see simulation results.

4.2 Prior for indicator vector γ

The prior for the variable selection vector γ is based on an independent Bernoulli assumption for each variable. Conditional upon the prior probability $l \in (0, 1)$ that a given variable is used as a predictor, this prior may be written as

$$p(\boldsymbol{\gamma}|l) = \prod_{i=1}^{I} l^{\gamma_i}(1-l)^{(1-\gamma_i)}. \tag{14}$$

As noted by George and Foster (2000), with κ and l fixed to set values, priors (13) and (14) have been used in Bayesian model selection problems (See for example Smith and Kohn, Fernandez, Ley, and Steel (2001b)). A uniform prior across all elements of the model space would imply that $l = 0.5$. However, 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 $p(\boldsymbol{\gamma}|D)$ 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 between analysts.

To make the model more robust to this kind of misspecification, we assign a Beta prior to l and marginalise, i.e.

$$p(l) = g(l|a,b) = l^{a-1}(1-l)^{b-1}/B(a,b), \quad a > 0,\ b > 0,\ 0 < l < 1, \tag{15}$$

with $B(a,b)$ denoting the complete beta function. Combining and (**??**) our beta-binomial prior $BB(I,a,b)$ is given by

$$\begin{aligned} p(\boldsymbol{\gamma}) &= \int_0^1 p(\boldsymbol{\gamma}|l)g(l|a,b)dl \\ &= \binom{I}{k} B(a+k, b+l-k)/B(a,b) \end{aligned} \tag{16}$$

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 (16) as

$$p(\boldsymbol{\gamma}) = \frac{\binom{I}{k} \prod_{i=0}^{k-1}(\pi + \tau i) \prod_{i=0}^{I-k-1}(1 - \pi + \tau i)}{\prod_{i=0}^{I-1}(1 + \tau i)} \tag{17}$$

(16) represents a Beta-Binomial(I,a,b) prior consisting of a mixture of binomial observations, k with 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 $\mathrm{Var}(l)$ goes to zero; this singularity is obviously consistent with simple binomial variation. Alternately, letting $a + b \to 0$, $\tau \to \infty$, and $\tau(1+\tau)^{-1} \to 1$. Given (), the conditional variance of $\sum \gamma_i$ goes to zero and the unconditional variance goes to $I^2\pi(1+\pi)$. See Palmquist...

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 the 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. In the particular case we consider in referring to competitive predictors ....

4.2.1 Model-Invariant Priors: σ_ε and α

We utilise a standard gamma conjugate prior distribution for σ_ε^2 given by

$$p(\sigma_\varepsilon^{-2}) = G(\sigma_\varepsilon^{-2} | c_\varepsilon, d_\varepsilon) \tag{18}$$

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. Here we 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.

4.3 Prior for covariates

There are many possibilities for the prior on covariates, including those which allow for a multiple dependent time series structure, and so on. 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 $\boldsymbol{\Sigma}_X = diag(\{\sigma_i^2\}_{i=1}^I)$.

5 Computational inference using MCMC

The prior framework described above facilitates the computation of posterior distributions. Specifically, 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, and both will be used in the construction of an efficient MCMC sampling algorithm. 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 for κ and the parameters for the Beta-binomial model distribution, τ and π.

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, $p(\boldsymbol{\beta}_\gamma, \boldsymbol{\gamma}, \alpha, \sigma_\epsilon, \kappa, \mathbf{X}_m, \boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X|\mathbf{D})$. 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 $p(\boldsymbol{\gamma}|D)$.[13] These sequences can be used to determine a range of models with high probability, and estimates of their probability.

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, in contrast with the schemes of Fernandez, Ley, and Steel (2001b), for example, where MCMC is used more for efficient sampling of a high-dimensional model space than for marginalisation of nuisance parameters and missing data. There are many 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, sampling $\boldsymbol{\gamma}$ from its reduced conditional distribution (with $\beta_{\boldsymbol{\gamma}}$, α and σ_ε analytically marginalised), and then imputing (β, 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, Kuo and Mallick,Fernandez, Ley, and Steel (2001b)) 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 the authors are able to utilise standard MCMC procedures by developing a prior model structure based on a mixture of two distributions. As opposed to allowing explanatory variables to be 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 Summary of algorithm

A single sweep of the full MCMC algorithm can be summarised as follows:

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

Discuss initialisation of MCMC i.e. setting missing data to means for first iteration. Also make sure we allow for any hierarchical structure in priors.

1. Indicators γ_i, $i = 1, 2, ..., I$ are sampled in turn with replacement from the reduced conditional distribution $p(\gamma_i|\boldsymbol{\gamma}_{/i}, \mathbf{y}, \mathbf{X}, \boldsymbol{\kappa})$, where $\boldsymbol{\gamma}_{/i} = (\gamma_1, ..., \gamma_{i-1}, \gamma_{i+1}, ..., \gamma_I)$, using a Metropolis-Hastings update. Specifically, if the current state of γ_i is s then we propose a change to $1-s$ and accept this change according to the Metropolis-Hastings rule with probability

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

2. $(\boldsymbol{\beta}_\gamma, \alpha, \sigma_\varepsilon)$ are sampled as a block from their full conditional $p(\boldsymbol{\beta}_\gamma, \alpha, \sigma_\varepsilon|\boldsymbol{\gamma}, \kappa, \mathbf{y}, \mathbf{X})$ (Normal-inverted gamma distribution).

3. Impute missing covariates from their full conditional distribution,

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

   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 are given in section/appendix ...

6 Posterior conditional distributions for MCMC sampler

>>>Details of full/reduced conditionals here <<<

6.1 Missing Data

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 γ 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.[14] , 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 $p(\mathbf{X}_{m,\gamma}| \quad) \equiv N(\overline{\mathbf{X}}^*_{obs,\gamma}, \sigma_x^2 B^*)$ By combining prior and data information, the posterior mean $\overline{\mathbf{X}}^*_{m,\gamma}$ may be written

$$\overline{\mathbf{X}}^*_{m,\gamma} = B^* \times [(P_p \times \overline{\mathbf{X}}_{ob}) + (D_p \times \widehat{\mathbf{X}}_{m,\gamma})]. \quad (19)$$

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. Letting $P_p = 1/vec(\Sigma_{ob})$, $D_p = \boldsymbol{\beta}'_m \boldsymbol{\beta}_m / \sigma_\varepsilon^2$, and $\widehat{\mathbf{X}}_{m,\gamma} = y_m / \boldsymbol{\beta}_m - \mathbf{X}_{ob,\gamma}\boldsymbol{\beta}_{ob,\gamma}$, (19) may be written as

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

Now random draw ....

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

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

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

.....

Discussion of work of George and Foster (2000) demonstrate that by adopting a specific set of implicit hyperparameters the ordering of models based on competing Bayesian posterior probabilities will be identical to the ordering provided by AIC and BIC selection criteria. Using this calibration, the authors propose an empirical Bayes selection criteria using hyperparameter *estimates* as opposed to the fixed choices... AIC, BIC are characterised by fixed dimensionality penalties... *Elaborate.*

6.2 MCMC Diagnostics

Correlation between histogram frequencies used to construct model probabilities and actual posterior probabilities. See Fernandez, Steele et al

6.2.1 Search over Models for BIC, AIC (incomplete)

Both the AIC and BIC approximation do not require MCMC for the parameters. Namely, 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. 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.

BIC approximation: the probability of moving from M^* to j is

$$\min\{1, \exp(BIC_j - BIC_{M^*})\}$$

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

In comparing Bayes factors with Akaike weights a number of observations are possible. First, note that the A represents a ratio of marginal likelihoods and therefore is equivalent to a ratio of classical likelihoods *conditional* upon integrating out parameter uncertainty as in (2). Therefore, we can view the Bayes factors as representing a measure of model uncertainty for h relative to $h^{'}$ which incorporates parameter uncertainty.[15] In this respect we can think of Akaike Weights as some sort of asymptotic approximation to Bayes factor.[16] Discussion of where the proportionately factor comes from above.

7 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.

7.1 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

[15] 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.

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

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.[17]

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.[18]

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.

7.2 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).

8 Crisis Intensity Results

Crisis intensity results are presented in Tables 4–6. 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

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

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

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

together with the 5th and 95th quantiles. In addition, we report the posterior marginal probability of inclusion for each indicator, m, given by

$$P_{l|y} = \sum_{j=1}^{2^I} \mathbf{1}(\gamma_{l,M_j} = 1) \cdot P(M_j|\mathbf{y})$$

which simply sums the posterior model probabilities for each model in which the indicator appears

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 make the following observations. First, in cases where we have specified a prior inclusion probability, following Xavier Sala-Martin et al (2002), we may 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 inclusion probability (mip); (ii) a credible interval which excludes zero, but with a relatively high marginal inclusion probability (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.[19] Using a similar framework to evaluate the relative significance over individual indicators as conducted in Xavier Sala-Martin et al (2002), we will refer to indicators as strong (s) if $mip > pip$ and the credible interval[20] does not include zero, and as marginal (m) if $pip > mip$ and the credible interval[21]. The remaining indicators are classified as weak (s). Although we concede the abritrary nature of such a classification, it does facilitate comparison over the Bayesian, AIC, and BIC results. In all tables the final column indicates this classification.

In Table 4 we present the results from a full Bayesian analysis, based on the exposition in Section *. The values for the fixed hyperparameters are provided as a footnote to the

[19] Possible to overcome this with a particular prior specification if such collinearity could be anticipated.

[20] Note that since Xavier Sala-Martin et al (2002) utilise a Bayesian Averging of Classical Estimates (BACE) approach to model uncertainty, then the term credible interval does not directly apply.

[21] Note that since Xavier Sala-Martin et al (2002) utilise a Bayesian Averging of Classical Estimates (BACE) approach to model uncertainty, then the term credible interval does not directly apply.

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 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. 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.

ADD: Now compare Bayesian results with approximations AIC and BIC

9 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 and contagion 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, 1999b). 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.

Extensions: Add country/time heterogeneity dummies, both additive and multiplicative

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.

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

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,.

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.

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.

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.

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.

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.

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.

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.

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.

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
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	

Chart 1, Number of Crises, 1977-99

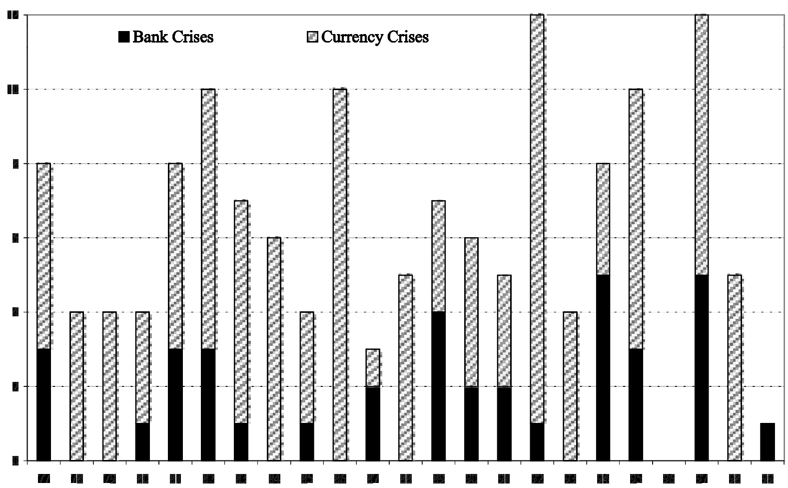

Figure 1:

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)
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%Quantile	95%Quantile	Post. Prob. I	Class
Legal Environment					
Rule of Law	0.13900	0.11466	0.81732	0.06823	m
Antidirector rights	0.47123	0.47516	1.1347	0.05783	m
Rat AccSt	-0.01813	-0.002653	0.03983	0.60567	w
Banking Sector					
4 year change in pc to GDP	-0.034356	-0.03477	-0.00501	0.09496	*m*
Domestic Credit to GDP	-2.9944	-3.1822	-0.41923	0.14180	*m*
Broad money to GDP	-3.8049	-4.3448	-0.53689	0.21170	*s*
External Sector					
Nbank L BIS	0.11758	0.11561	0.17061	0.13073	*m*
Imports to GDP	0.15230	0.15101	0.18625	0.68020	*s*
90day LIBOR	0.096881	0.05136	1.5882	0.04676	*m*
Net Nb BR	-0.04669	-0.0478	0.15889	0.05406	*w*
Current account to GDP	0.58405	0.60811	0.86657	0.23480	*s*
Trade Balance to GDP	0.050394	0.07810	0.25467	0.05793	*m*
Terms of Trade	-0.00176	-0.0024	0.09117	0.04620	*w*
Annual Average CPI Inflation	0.00164	0.00164	0.00334	0.05273	*m*
Income Development	0.08607	0.17730	1.4808	0.06990	*m*
Corporate Sector					
Total debt to Common Equit.	-2.7728	-2.3515	-1.2186	0.29567	*s*
Equit. to Total Capital	6.9354	7.3166	13.762	0.09933	*m*
Current Ratio	9.8604	9.8846	11.904	0.81953	*s*
Working Capital/Total Capital	-1.8168	-2.0555	0.69159	0.07286	*w*
Long term debt to common equit.	-0.37116	-1.0118	2.6578	0.08726	*w*
Quick Ratio	6.8484	7.5883	11.026	0.18490	*m*
Long term Debt. to Total Capital	1.7736	2.4034	5.5255	0.09856	*m*
Financial Breadth					
FBr2	1.4970	1.6747	3.0765	0.11840	*m*
Private Bond Mt. Cap/GDP	-0.00439	-0.00362	0.0438	0.04853	*w*
Foreign Exchange Liquidity					
Broad money	-1.4580	-1.6456	0.58066	0.11460	*w*
Change in pc flows to GDP	0.77229	0.77216	0.92549	0.96330	*s*
Contagion	-0.12661	-0.098186	0.14748	0.05026	*w*

Number of Monte Carlo Iterations (Burnin) 30000 (10000)
Average Posterior Model Size = 6.464
IG prior on σ_ε
$\alpha_\varepsilon/\beta_\varepsilon = le010/le - 010$
G-Prior for Covariance Matrix of beta.
g-factor = 0.0222222=(l/n)

Table 5: Akaike Small Sample Adjustment

	Posterior mean	5% Quantile	95% Quantile	Post. Prob. I	Class
Legal Environment					
Rule of Law	0.38940	-0.27157	0.9979	0.09513	*w*
Antidirector rights	0.56948	-0.32101	1.2683	0.08843	*w*
Rat AccSt	-0.036122	-0.19342	0.1233	0.16333	*w*
Banking Sector					
4 year change in pc to GDP	-0.039158	-0.0667	-0.01052	0.14213	*m*
Domestic Credit to GDP	-2.6782	-5.0319	0.13528	0.14810	*w*
Broad money to GDP	-4.1744	-5.8593	-1.3187	0.27220	*s*
External Sector					
Nbank L BIS	0.12812	0.07490	0.19509	0.15740	*m*
Imports to GDP	0.15327	0.10510	0.18803	0.63687	*s*
90day LIBOR	0.55810	-0.5718	1.7528	0.07143	*w*
Net Nb BR	-0.1379	-0.30458	-0.0053	0.09033	*m*
Current account to GDP	0.65756	0.46087	0.90581	0.49887	*s*
Trade Balance to GDP	-0.06867	-0.38687	0.23471	0.09566	*w*
Terms of Trade	-0.01875	-0.10045	0.07900	0.06846	*w*
Annual Average CPI Inflation	0.001686	0.00019	0.00387	0.08246	*m*
Income Development	-0.45617	-1.5797	0.62851	0.07516	*w*
Corporate Sector					
Total debt to Common Equit.	-3.6207	-6.8617	-1.3606	0.32877	*s*
Equit. to Total Capital	9.8298	2.9502	19.989	0.18117	*m*
Current Ratio	9.4731	7.3409	11.107	0.84797	*s*
Working Capital/Total Capital	-1.4680	-3.3821	1.5583	0.09310	*w*
Long term debt to common equit.	0.14470	-2.2165	4.275	0.11337	*w*
Quick Ratio	5.9681	1.0795	10.214	0.18483	*m*
Long term Debt. to Total Capital	3.3923	-1.3359	6.3765	0.17580	*w*
Financial Breadth					
FBR2	1.3218	0.26563	2.8322	0.13217	*m*
Private Bond Mt. Cap/GDP	0.00531	-0.04887	0.0522	0.07166	*w*
Foreign Exchange Liquidity					
Broad money	-0.55623	-1.7662	1.0745	0.08723	*w*
Change in pc flows to GDP	0.78615	0.60309	0.93011	0.97800	*s*
Contagion	-0.20490	-0.51548	-0.0051	0.07726	*w*

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% Quantile	95% Quantile	Post. Prob I	
Legal Environment					
Rule of Law	0.38421	-0.31719	1.0257	0.065033	w
Antidirector rights	0.55606	-0.29595	1.2399	0.056967	w
Rat AccSt	-0.047191	-0.19796	0.12095	0.12333	w
Banking Sector					
4 year change in pc to GDP	-0.043376	-0.073349	-0.013706	0.11037	m
Domestic Credit to GDP	-2.9941	-4.9828	-0.33467	0.12910	m
Broad money to GDP	-4.3020	-5.7380	-1.8073	0.23507	s
External Sector					
Nbank L BIS	0.12646	0.071667	0.19335	0.10147	m
Imports to GDP	0.16062	0.11210	0.18955	0.60543	s
90day LIBOR	0.63773	-0.52899	1.8750	0.048633	w
Net Nb BR	-0.13356	-0.30044	-0.00534	0.059167	m
Current account to GDP	0.66853	0.47574	0.88609	0.41133	s
Trade Balance to GDP	-0.00078	-0.37076	0.25797	0.065733	w
Terms of Trade	-0.02489	-0.10619	0.083637	0.045600	w
Annual Average CPI Inflation	0.001754	0.00028	0.00399	0.053200	m
Income Development	-0.49445	-1.6509	0.66056	0.053600	w
Corporate Sector					
Total debt to Common Equit.	-3.2948	-6.6033	-1.3975	0.24653	s
Equit. to Total Capital	9.6819	3.1092	19.676	0.13780	m
Current Ratio	9.5474	7.5349	11.125	0.79617	s
Working Capital/Total Capital	-1.5659	-3.3786	1.4265	0.063033	w
Long term debt to common equit.	-0.25875	-2.1230	3.3838	0.08233	w
Quick Ratio	7.0082	1.6501	10.625	0.19130	m
Long term Debt. to Total Capital	2.9468	-1.4450	6.1555	0.11260	
Financial Breadth					
FBR2	1.3988	0.26629	2.9768	0.09290	m
Private Bond Mt. Cap/GDP	0.003254	-0.05559	0.051812	0.04850	w
Foreign Exchange Liquidity					
Broad money	-0.70258	-1.7552	1.0103	0.05860	w
Change in pc flows to GDP	0.78866	0.60330	0.93091	0.95467	m
Contagion	-0.21499	-0.52472	-0.00807	0.05166	m

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

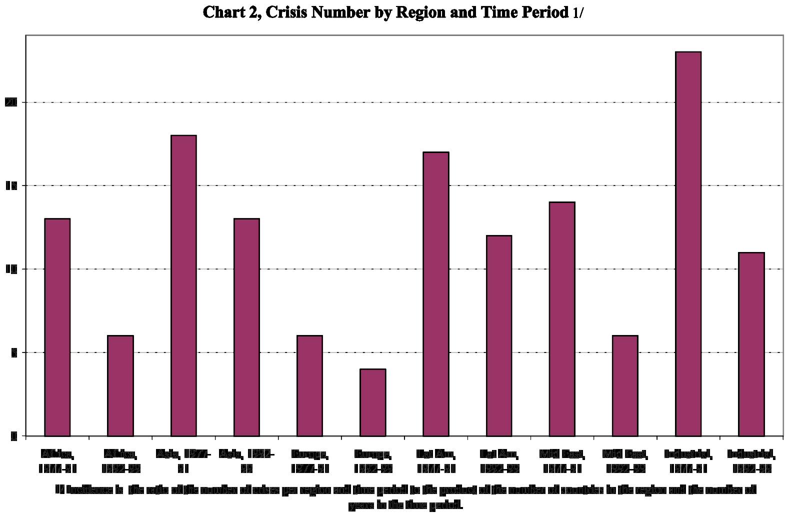

Figure 2:

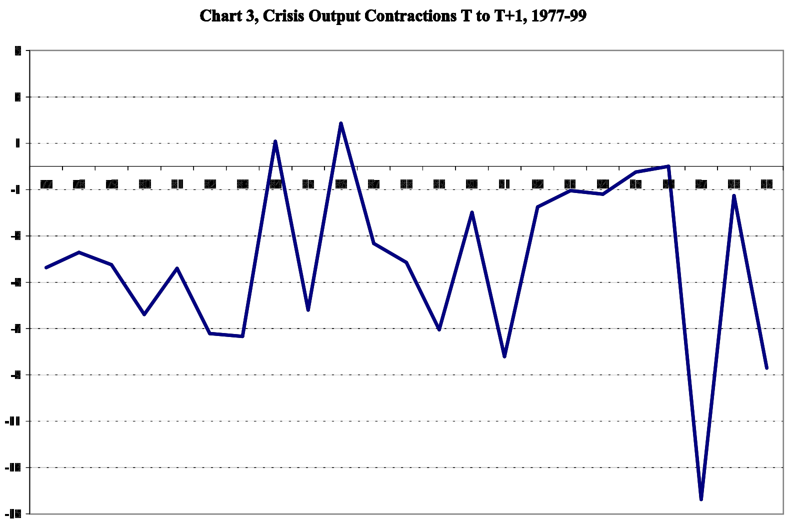

Figure 3: