Essays on Bayesian and Classical Econometrics with Small Samples

Marek Jarociński
Universitat Pompeu Fabra
Barcelona

Advisor: Albert Marcet

April 24, 2006

Dipòsit legal: B.34199-2006
ISBN: 84-690-0320-8

Keywords: Monetary Policy Transmission, Structural Vector Autoregression, Bayesian Estimation, Exchangeable Prior, Small-Sample Bias, Bias Correction, Bayesian Model Averaging, Growth Empirics

Contents

Acknowledgements

Above all, I would like to thank my advisor Albert Marcet for his invaluable advice, guidance and support. Further, I am greatful to Antonio Ciccone for the opportunity to work with him, which had important positive externalities for all my research. My work hugely benefited from feedback and conversations with many people, especially Fabio Canova, Marco Del Negro, Jordi Gali, Luca Gambetti, Bartosz Maćkowiak, Ricardo Nunes, Philip Sauré, Gustavo Solorzano and participants of workshops at the Universitat Pompeu Fabra. My work for CASE - Center for Social and Economic Research, Warsaw, was stimulating, by exposing me to the applied perspective on economic research. Finally, I would like to thank CEFIR, Moscow for hospitality during part of this work. All errors are the sole responsibility of the author.

Barcelona, April 2006 Marek Jarociński

All computations were performed in Ox version 3.40 (see: Doornik, J.A. (2002), Object-Oriented Matrix Programming Using Ox, 3rd ed. London: Timberlake Consultants Press and Oxford: www.doornik.com).

Foreword

Small available samples are a common problem in applied econometrics. It is therefore important to be able extract the most information from the available sources as efficiently as possible. One promising method is the use of Bayesian analysis in which all available non-sample information is introduced in the form of prior distributions. However, specifying - as well as justifying - the parameters of such distributions present significant challenges. Furthermore, understanding the relationships and differences between Bayesian and Classical estimators is important for the practitioner, and often yields interesting theoretical insights.

Macroeconomic data are particularly scarce in industrial countries, and even more so in the rest of the world. The three articles below deal with the estimation on such data, in situations where, lacking additional prior information, a researcher risks misinterpreting the evidence implied by data. Vector Autoregressions (VARs) are the subject of the first two papers and the third deals with cross-country growth regressions.

VARs are particularly data-demanding, since they contain many parameters, and, being time series models, rely on asymptotic theory. Yet, they are used routinely, and often on quite short datasets. Inference from their OLS estimates is marred by two problems: random noise and a systematic small-sample bias. The first two papers (one of them joint with Albert Marcet) demonstrate how to deal with these difficulties. The first adapts the exchangeable prior specification due to Gelman, Carlin, Stern and Rubin (1995) to the VAR context. The second proposes a new prior - on initial growth rates of modeled variables.

To explain why the prior on growth rates improves the small-sample inference in time series, the second paper reinterprets and reconciles contradictory views of Bayesian and Classical econometricians on this issue. This is achieved by indicating what is unreasonable in the flat prior in the time-series context (which is a new point, unrelated to the existing critique of Phillips 1991) and which priors are implicit in classical analyses of time series. As a by-product, we find estimators for the AR(1) model which, in terms of Mean Squared Error, outperform available alternatives in a relevant range of parameters (which we illustrate in a Monte Carlo study).

All three papers, but especially the first and third, contain novel empirical findings. The first provides new results about the effects of monetary shocks in Eastern European countries which, due to short data series, have received little attention to date.

The third paper (joint with Antonio Ciccone) contributes to the empirical growth literature, where data shortage and data quality are key problems. It is an empirical work, focusing on the issues raised by model uncertainty and model selection in a small sample, in the context of measurement error in the income data. It updates our knowledge about robust determinants of growth with new data. At the same time, however, it raises the problem of sensitivity of empirical results to measurement errors, studying it by Monte Carlo and on multiple available datasets. It uses and compares both classical and Bayesian approaches. While Bayesian approaches are found to be more robust to measurement error, the quality of the available data imposes binding limits on what can be learned from agnostic cross-country growth empirical studies. Therefore, with imperfect income data, informative priors are necessary to resolve model uncertainty.

Further work is planned to study the forecasting performance of VARs with priors on growth rates. The effect is somewhat similar to that of the popular Minnesota prior, but we argue that it is less ad hoc and is interesting if it provides an edge in forecasting. Another direction is to apply VARs with the proposed priors to other problems where short data are particularly constraining. Among the most interesting applications is studying the effects of shocks in the euro area countries after the introduction of the common currency.

Chapter 1

Responses to Monetary Policy Shocks in the East and the West of Europe: A Comparison

Abstract: This paper compares responses to monetary shocks in the EMU countries (in the pre-EMU sample) and in the New Member States (NMS) from Central Europe. The small-sample problem, especially acute for the NMS, is mitigated by using a Bayesian estimation procedure which combines information across countries. A novel identification scheme for small open economies is used. The estimated responses are quite similar across regions, but there is some evidence of more lagged, but ultimately stronger price responses in the NMS economies. This contradicts the common belief that monetary policy is less effective in post-transition economies, because of their lower financial development. NMS also have a probably lower sacrifice ratio, which is consistent with the predictions of both the imperfect information model of Lucas (1973) and the New-Keynesian model of Ball et al. (1988).

1.1 Introduction

Prior to the creation of the Economic and Monetary Union (EMU), much research was devoted to the question of possible heterogeneity of responses to monetary shocks in the prospective member countries. The question was motivated by significant differences in some structural characteristics of the EMU economies and their possible implications for monetary transmission. If responses to monetary shocks are significantly heterogeneous, and the reasons of this heterogeneity do not disappear in the monetary union, conducting common monetary policy will be politically difficult (Dornbusch et al., 1998):

The burden of disinflation will fall disproportionately on some countries, while other will have to accept higher than average inflation.[1]

Examples of papers discussing the impact of structural characteristics of the European economies on their monetary transmission are Dornbusch et al. (1998), Guiso et al. (1999), Mihov (2001), Ehrmann et al. (2003). These papers first look for indicators of interest sensitivity of output, size, health and structure of the banking sector, stock market capitalization and other, and relate them, by theoretical reasoning, to the strength of monetary transmission. The results of this type of analysis are often ambiguous, as different characteristics sometimes have conflicting implications, and their relative quantitative importance is unclear. The ultimate judgment has to come from macroeconomic data, usually analyzed with a Structural VAR technique. Papers in this line or research naturally fall into two categories: those that find significant and interpretable differences among the examined countries, and those that don't. Examples of the first group are Mihov (2001) and Ramaswamy and Slok (1998). Kieler and Saarenheimo (1998) and Ehrmann et al. (2003)/Mojon and Peersman (2001), among others, find that whatever asymmetries in monetary transmission might exist among EU countries, they are not strong enough to be robustly detected in the available data.

Now research along similar lines is being extended to the New Member States (NMS) from the Central and Eastern Europe, which joined the European Union in 2004, and which are legally obliged to adopt the euro some time afterwards. Examples are Anzuini and Levy (2004), Elbourne and de Haan (2005) and Creel and Levasseur (2005). Ganev et al. (2002), in addition to their own empirical analysis, review numerous studies conducted in the NMS central banks. Most authors agree that, mainly because of the small size of the financial markets, monetary policy in transition countries should have little effect, although its effectiveness is likely to be increasing with time, as the market economies in the region become more mature.

What has been missing so far, is an explicit comparison of responses to monetary shocks across the two regions: Central-Eastern and the Western Europe. Such comparison can be expected to be more meaningful and interesting than intra-regional comparisons performed so far, as the structural differences between these regions dwarf those within them. This paper fills this gap, by using a novel econometric technique, which allows to robustly estimate responses for both regions despite short data series and in a unified framework.

[1] This problem is independent from the long debated question of whether current and potential future EMU member countries constitute an optimal currency area.

The comparison yields interesting results: First, monetary shocks in the NMS are associated with larger movements of the interest rate. Second, in spite of the structural differences between the regions, we find no support for the relative ineffectiveness of monetary policy in the NMS. Responses of output and prices are broadly similar, once controlling for the size of the interest rate shock. If anything, price responses are more lagged, but later more vigorous in the NMS and in the medium term these countries may be facing a more favorable sacrifice ratio.

We conclude that, on the one hand, the structural weakness of monetary transmission in the NMS is quantitatively less important than widely believed, possibly because it is compensated by a relative strength of the exchange rate channel, by more limited access to foreign financial markets and other discussed factors. On the other hand, prices appear to be more responsive, once aggregate demand is affected. This finding is consistent with greater volatility of aggregate demand and higher average inflation in the NMS, through well known mechanisms proposed in Lucas (1973) and Ball et al. (1988).

As regards the implications of these results for the adoption of the euro by the NMS, the common limitation of this and previous studies is that they are subject to the Lucas critique: responses to monetary shocks are likely to change after a further EMU expansion. Nevertheless, as argued in the above quoted papers, the empirical results based on past data provide stylized facts, which are a reasonable departure point for further speculations. Taken at face value, results of this paper downplay the structural weakness of monetary transmission as an argument against further EMU expansion.

The principal obstacle in the study of the former communist countries are the short available data series. To mitigate this problem, we perform a Bayesian estimation with the prior (called 'exchangeable prior'), which conveys the intuition that parameters of VAR models for individual countries are similar across the region, since all economies in the region are special cases of the same underlying economic model. This prior results in estimates which are shrunk towards a common mean. The Bayesian setup used here allows to formulate the problem in such a way, that both the degree of shrinking, and the weights of countries in the common regional mean are endogenous and optimal for the sample at hand. This guarantees the most efficient use of the scarce available data.

Applications of the estimation with the exchangeable prior in economics include Zellner and Hong (1989) and Canova and Marcet (1995). The former find that the exchangeable prior improves the out of sample forecasting ability in time series models, which has been also exploited in the forecasting time-varying VARs of Canova and Ciccarelli (2004). This finding suggests, that it should also increase reliability of a structural analysis. However, it has not been used for structural VARs, except for Canova (2005), which uses a different technique of working with a similar prior. The present paper adapts to VARs the formulation of Gelman et al. (1995), called Hierarchical Linear Model, which allows to avoid the specification of a subjective prior about the degree of similarity between units, but instead determines it solely from the data.

The second methodological novelty of this paper is the proposed identification of monetary shocks in a VAR. They are identified by assuming that they influence output and prices with at least one month lag, and that they involve a negative comovement of interest rate and exchange rate innovations. This is a combination of standard zero restrictions with more recently proposed sign restrictions.[2]

The structure of the paper is following: Section 2 discusses estimation of reduced form VARs for a panel of countries as a Hierarchical Linear Model, Section 3 describes identification of monetary shocks, Section 4 presents results and Section 5 contains conclusions. Details about data and estimation are in the appendix.

1.2 Estimation

VAR models contain many parameters, and their estimation with short samples, such as those available for the post-communist countries, results in wide error bands and point estimates which are very sensitive to small changes in sample or specification. The strategy employed here to obtain more robust results, is to analyze whole regions (first EMU, then NMS) jointly and exploit the intuition, that parameters of VAR models for individual countries are similar across the region, since all economies in the region are special cases of the same underlying economic model. However, we need to stop short of assuming that all slope coefficients are the same across countries and performing a standard panel estimation. This assumption would only be an approximation, and in dynamic model (such as a VAR) it could seriously distort the results (Pesaran and Smith (1995) show that it results in the inconsistency of the estimator).

The estimation procedure is Bayesian, and uses the Hierarchical Linear

[2] Other sign restrictions have been used in Faust (1998), Uhlig (2001), Canova and De Nicoló (2002) and Kieler and Saarenheimo (1998). Unlike in the mentioned papers, the combination of zero and sign restrictions enables one to find the desired factorizations of the error variance matrix analytically, avoiding a numerical search procedure.

Model of Gelman et al. (1995).[3] The idea of similarity is specified as a Normal prior for each country's coefficients, which is centered at the mean which is common for all the region (an exchangeable prior). This prior causes the coefficients to be shrunk towards the common mean. The second stage of the hierarchy consists of the 'hyperprior' about the prior parameters: common mean and the variance of country coefficients around the common mean ('hypervariance'). The Hierarchical Linear Model allows the priors in the second stage of the hierarchy to be noninformative, and therefore the posterior common mean and hypervariance are determined optimally only from the data. Intuitively, more different and more tightly estimated country coefficients increase the posterior probability of large values of the hypervariance. When country coefficients are more similar, or if they differ, but have larger error bounds, hypervariance is more likely to be smaller. Country models which are more tightly estimated receive more weight in the posterior common mean, relative to countries whose estimates are imprecise.

Below, we first distinguish between parameters which are likely to be similar across countries, and those which need not be. So, we apply the exchangeable prior to parameters determining dynamic interrelationships between the endogenous variables, and reactions to common exogenous variables. We specify a noninformative prior for parameters of exogenous variables which are not present for all countries, and for constant terms, which implies that we have country 'fixed effects'.

The following two subsections specify the above prior in the panel VAR setup: first the overall framework, and then the parametrization of the hypervariance. The computation of the posterior is explained in the appendix.

1.2.1 Panel of VARs as a Hierarchical Linear Model

In what follows, vectors are denoted by lowercase, matrices by uppercase bold symbols, $i = 1 \dots I$ denotes countries, $j = 1 \dots J$ denotes endogenous variables in a VAR, $k = 1 \dots K$ denotes the common right-hand-side variables in the reduced form VAR, $l = 1 \dots L$ denotes lags, $m = 1 \dots M_i$ denotes country specific exogenous variables in the VAR, $t = 1 \dots T_i$ denotes time periods.

For each country in the panel we consider a reduced form VAR model of

[3] Classical estimators for heterogeneous panels exist, but are much less efficient: Monte Carlo study in Hsiao et al. (1999) shows that in small samples they perform worse than a variant of Bayesian estimator with the exchangeable prior.

the form:

$$\boldsymbol{y}_{it} = \sum_{l=1}^{L} \boldsymbol{B}'_{il}\boldsymbol{y}_{i(t-l)} + \boldsymbol{\Delta}'_i\boldsymbol{w}_t + \boldsymbol{\Gamma}'_i\boldsymbol{z}_{it} + \boldsymbol{u}_{it} \tag{1.1}$$

$\boldsymbol{y}_{it}$ is a vector of J endogenous variables and $\boldsymbol{w}_t$ is a vector of those exogenous variables which are common across countries. We will specify an exchangeable prior about the coefficients of $\boldsymbol{y}_{i(t-l)}$ and $\boldsymbol{w}_t$. The prior will be uninformative for variables in $\boldsymbol{z}_{it}$, which include country specific constant terms and variables which are included for some, but not all countries. Vector $\boldsymbol{u}_{it}$ contains VAR innovations which are i.i.d. $N(0, \boldsymbol{\Sigma}_i)$.

We gather the variables to which the exchangeable prior applies in a vector $\boldsymbol{x}_{it} = \left[\boldsymbol{y}'_{i(t-1)} \cdots \boldsymbol{y}'_{i(t-L)}, \boldsymbol{w}'_t\right]'$. Stacking vertically $\boldsymbol{y}'_{it}, \boldsymbol{x}'_{it}, \boldsymbol{w}'_t$ for all t we obtain the model in terms of data matrices:

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{B}_i + \boldsymbol{Z}_i\boldsymbol{\Gamma}_i + \boldsymbol{U}_i \tag{1.2}$$

where $\boldsymbol{Y}_i$ and $\boldsymbol{U}_i$ are $T_i \times J$, $\boldsymbol{X}_i$ are $T_i \times K$, $\boldsymbol{B}_i$ are $K \times J$, $\boldsymbol{Z}_i$ are $T_i \times M_i$ and $\boldsymbol{\Gamma}_i$ are $M_i \times J$. We have: $K = JL + W$, where W is the length of the $\boldsymbol{w}_t$ vector. The coefficient matrix $\boldsymbol{B}_i$ is related to coefficients of (1.1) by: $\boldsymbol{B}_i = [\boldsymbol{B}'_{i1}, \ldots, \boldsymbol{B}'_{iL}, \boldsymbol{\Delta}']'$.

Let $\boldsymbol{y}_i = \operatorname{vec}\boldsymbol{Y}_i, \boldsymbol{\beta}_i = \operatorname{vec}\boldsymbol{B}_i, \boldsymbol{\gamma}_i = \operatorname{vec}\boldsymbol{\Gamma}_i$.

The statistical model generating the data is assumed to be following:

Likelihood for country i:

$$p(\boldsymbol{y}_i|\boldsymbol{\beta}_i, \boldsymbol{\gamma}_i, \Sigma_i) = \mathrm{N}((\boldsymbol{I}_J \otimes \boldsymbol{X}_i)\boldsymbol{\beta}_i + (\boldsymbol{I}_J \otimes \boldsymbol{Z}_i)\boldsymbol{\gamma}_i, \boldsymbol{\Sigma}_i \otimes \boldsymbol{I}_{T_i}) \tag{1.3}$$

Country coefficients on the variables in $\boldsymbol{X}_i$ are assumed to be drawn from a normal distribution with a common mean $\bar{\boldsymbol{\beta}}$:

$$p(\boldsymbol{\beta}_i|\bar{\boldsymbol{\beta}}, \lambda, \boldsymbol{L}_i) = \mathrm{N}(\bar{\boldsymbol{\beta}}, \lambda\boldsymbol{L}_i) \tag{1.4}$$

where λ is an overall prior tightness parameter and $\boldsymbol{L}_i$ is a known, fixed matrix whose construction is discussed below.

Prior for $\bar{\boldsymbol{\beta}}$ and $\boldsymbol{\gamma}_i$ is uninformative, uniform on the real line:

$$p(\bar{\boldsymbol{\beta}}) \propto p(\boldsymbol{\gamma}_i) \propto 1 \tag{1.5}$$

Alternatively, one could use some informative prior for $\bar{\boldsymbol{\beta}}$, e.g. the Minnesota prior, but, as discussed in Gelman et al. (1995), this is not necessary for the estimation problem to be well posed, and, in order to keep things simple, we do not pursue this possibility here. We also use the standard diffuse prior for the error variance:

$$p(\boldsymbol{\Sigma}_i) \propto |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}(J+1)} \tag{1.6}$$

Finally, the prior for the overall tightness parameter λ is:

$$p(\lambda|s,v) = \mathrm{IG}_2(s,v) \propto \lambda^{-\frac{v+2}{2}} \exp\left(-\frac{1}{2}\frac{s}{\lambda}\right) \tag{1.7}$$

where IG_2 denotes the inverted gamma-2 distribution, while s and v are known parameters. For $s > 0$ and $v > 0$ this is a proper, informative prior while $s = 0$ and $v = -2$ results in an improper, noninformative prior.

The model in (1.3)-(1.7) defines the structure advocated in the introduction: the countries' dynamic models of variables in $\boldsymbol{Y}_i$ (and possibly some exogenous controls in $\boldsymbol{W}$) are special cases of the unknown underlying model defined by $\bar{\boldsymbol{\beta}}$. Variables in $\boldsymbol{Z}_i$ are those, for which the exchangeable prior would not be reasonable, primarily the country specific constant terms.

The functional form of the prior: combination of normal, uniform, inverted gamma and a degenerate inverted Wishart (for $\boldsymbol{\Sigma}_i$) densities is standard, motivated by computational convenience, so that the prior is conditionally conjugate. The posterior density of the parameters of the model is computed from the Bayes theorem, as a normalized product of the likelihood and the prior. The conditional conjugacy of the prior means that all conditional posterior densities are also normal, inverted gamma and inverted Wishart, which enables convenient numerical analysis of the posterior with the Gibbs sampler.[4]

1.2.2 Specification of the prior variance

The parametrization of the prior variance for $\boldsymbol{\beta}_i$ is inspired by prior variances in Litterman (1986) and Sims and Zha (1998): it is assumed to be diagonal, with the terms of the form:

$$\frac{\lambda\sigma_{ij}^2}{\sigma_{ik}^2} \tag{1.8}$$

The ratio of variances reflects the scaling of the variables. As in the above papers, σ^2s are computed as error variances from univariate autoregressions of the variables in question. Therefore, the $\boldsymbol{L}_i$ is computed as:

$$\boldsymbol{L}_i = \mathrm{diag}(\sigma_{ij}^2) \otimes \mathrm{diag}(\frac{1}{\sigma_{ik}^2}) \tag{1.9}$$

The parameter λ determines the overall tightness of the exchangeable prior. $\lambda = 0$ results in full pooling of information across countries and implies a panel VAR estimation, where all country VAR models are assumed to be

[4] See appendix for details. Gibbs sampler for a similar univariate problem is discussed in more detail in Gelman et al. (1995).

identical. On the other hand, as λ grows, country models are allowed to differ more, and become similar to the respective single country estimates. Since the value of λ is unknown, a (possibly uninformative) prior distribution is assumed for it and a posterior distribution is obtained. The reported results are integrated over this posterior distribution. If the posterior inferences conditional on particular values of λ differ in an economically meaningful way, odds ratios for alternative ranges of λ can be computed.

The use of noninformative priors carries the danger of obtaining an improper posterior and rendering the whole problem ill-defined. It is known, that in a hierarchical linear model like (1.3)-(1.7) the use of the usual noninformative prior for a variance parameter:

$$p(\lambda) \propto \frac{1}{\lambda} \tag{1.10}$$

(which obtains when $s = 0$ and $v = 0$) results in an improper posterior (see Hobert and Casella, 1996; Gelman et al., 1995). In this case, the marginal posterior for λ behaves like $1/\lambda$ close to the origin and is not integrable. However, Theorem 1 in Hobert and Casella (1996, p.1464), proved in a similar setup, suggests that the posterior is proper when $s = 0$ and $v = -2$, which corresponds to:

$$p(\lambda) \propto 1 \tag{1.11}$$

1.3 Identification of monetary shocks

Identification of the structural model assumes a small open economy with exchange rates flexible enough to react immediately to monetary policy, and monetary policy reacting immediately to the movements of the exchange rate. This assumption remains valid also in presence of managed exchange rates with target bands, like the ERM in the European Union, and some arrangements in the NMS, as long as the rate is not effectively fixed. It is a known empirical regularity (confirmed in the robustness analysis for this paper) that for countries other than the USA, identification schemes that do not allow for immediate response of the exchange rate to the interest rate, and vice versa, produce a 'price puzzle', i.e. an initially positive response of prices to monetary tightening (for more on this subject see e.g. Kim and Roubini (2000)).

The endogenous variables in the VARs are: output, consumer prices, short term interest rate and the exchange rate in national currency units per foreign currency unit, all measured at monthly frequency. No money aggregate is included: It is assumed that the central banks target short term interest rates, and adjust monetary aggregates consistently with this objective. In this setup, interest rates reflect only money supply decisions, and fluctuations of monetary aggregates carry additionally information about money demand. Since identification of money demand is beyond scope of this paper, we conserve degrees of freedom and do not include money aggregates in the specification.

In order not to confuse domestic monetary shocks with the central banks' responses to external developments, specifications include several foreign variables which are treated as exogenous. World developments are captured by current and lagged US Federal Funds Rate, oil and commodity prices. The status of Germany as both regions' locomotive is reflected by including, for all countries, current and lagged German interest rate and two lags German industrial production.

As usually in the identified VAR literature, it is assumed that structural shocks are orthogonal, and thus the covariance matrix of the VAR residuals conveys information about the coefficients of the contemporaneous relationships between endogenous variables. The relationship between the vector of structural shocks $\boldsymbol{v}_{it}$ and the vector of VAR innovations $\boldsymbol{u}_{it}$ is following:

$$\boldsymbol{G}_i \boldsymbol{v}_{it} = \boldsymbol{u}_{it} \tag{1.12}$$

where $\text{var}(\boldsymbol{v}_{it}) = \boldsymbol{I}_J$ (identity matrix of order J) and $\text{var}(\boldsymbol{u}_{it}) = \boldsymbol{\Sigma}_i = \boldsymbol{G}_i \boldsymbol{G}_i'$. Therefore, the identification involves finding a factorization $\boldsymbol{G}_i$ of the residual covariance matrix that complies with the identifying restrictions.

The identification restrictions[5] adopted to pin down the monetary shock are following:

1. Output and prices do not respond immediately to the monetary policy shock

2. The monetary policy shock is the one which involves a negative co-movement of the interest rate and the exchange rate on impact, i.e. interest rate rise is accompanied by exchange rate appreciation.

The remaining shocks are not identified and the triangular form of the upper left block of the matrix reflects a normalization, which has no effect on the impulse responses to the monetary policy shock. The identification restric-

[5]Restrictions are imposed here on the impulse responses in the first period (inverse of the matrix of structural coefficients), and not on the structural coefficients.

tions are summarized in the scheme below:

$$\begin{pmatrix} + & 0 & 0 & 0 \\ \bullet & + & 0 & 0 \\ \bullet & \bullet & + & + \\ \bullet & \bullet & - & + \end{pmatrix} \begin{pmatrix} v_{it1} \\ v_{it2} \\ \widehat{v_{it3}} \\ v_{it4} \end{pmatrix} = \begin{pmatrix} u_{it1} \\ u_{it2} \\ u_{it3} \\ u_{it4} \end{pmatrix} \begin{matrix} \leftarrow \text{output innovation} \\ \leftarrow \text{price innovation} \\ \leftarrow \text{interest rate innovation} \\ \leftarrow \text{exchange rate innovation} \end{matrix} \tag{1.13}$$

where + denote coefficients that are constrained to be positive, 0 - zero restrictions and $\bullet$ - unconstrained coefficients. $\widehat{v_{it3}}$ is the monetary policy shock. Factorizations satisfying (1.13) are obtained by rotating the bottom-right block of the Choleski factor of the residual covariance matrix - see the appendix for details.

The zero restrictions applied here are standard ones, used in Leeper et al. (1996), Kim (1999), Kim and Roubini (2000) and other papers. The strategy followed in these papers is to impose some more zero restrictions: to assume that monetary authorities don't react immediately to output and price developments. Then the model becomes (over)identified and can be estimated by maximum likelihood. It is then expected, that resulting estimates of the relationships between interest rate and exchange rate are 'reasonable', i.e. as in assumption 2 above. When multiple maxima of the likelihood function exist, the ones satisfying the implicit sign restriction are chosen, which is justified by a Bayesian reasoning (Kim, 1999, footnote 13, p.395).

However, the zero restrictions on the authorities' response to output and prices may not hold exactly: While it is true that official data on output and prices are compiled with delay, it could be argued that the central bankers can have access to quick business community surveys, price surveys, and certainly are able to identify quickly major shocks, like big strikes or floods. The sign restrictions proposed here are an attractive alternative.

1.4 Data and Samples

We analyze two panels of countries: five euro-area countries (EMU5) and four New Member States (NMS4) from the Central and Eastern Europe.[6] The goal of the paper is to compare the two regions within one unified framework. On the one hand, we want to make the analysis representative and include as many countries as possible. On the other hand, to ensure comparability of the results across regions, we want to maintain the same specifications and identification schemes. This leads to the choice of the assumption of open

[6] Romania and Bulgaria were considered as potential members of this group, although strictly speaking they are only expected to join the EU in 2007. On the other hand, we did not consider Malta and Cyprus, which are both economically and geographically distinct.

economy with a flexible (i.e. not fixed) exchange rate, which maximizes the sizes of both panels.

The EMU5 panel consists of Finland, France, Germany, Italy and Spain. Austria, Belgium and Netherlands were excluded because of their quasi-fixed exchange rate against the D-Mark[7], and Ireland because of the lack of monthly CPI data. In Greece, interbank interest rates are not available before 1998, only the Central bank rate is reported. Portugal proved to be an outlier because of dramatic swings of its interest rates in the wake of the 1992 European Monetary System crisis. Therefore, both countries are considered only in robustness checks.

The NMS4 panel consists of Czech Republic, Hungary, Poland and Slovenia. Bulgaria and the Baltic countries were excluded because they had currency boards. In Romania and Slovakia market interest rates vary widely and, for much of the sample, independently of the central bank interest rates. This suggests that the standard model of monetary management, which underlies this analysis, where the central bank manages market interest rates by setting its instrument interest rate, has not been firmly in place. Another possibility is that these countries experienced big shocks to money demand which were not accommodated by the central bank. In either case, the identification of monetary policy shocks adopted here might not be appropriate. Therefore, Slovakia and Romania are considered only in robustness checks.

The sample periods for the NMS countries span the second half of 1990-s up to second half of 2004 and differ for each country, depending on when the post-transition exchange rate control was relaxed. The information on the chronology of the exchange rate regimes was taken from Anzuini and Levy (2004), Table 8 and from Ganev et al. (2002), Section 3 and Table 1. The details about the samples are in the appendix. For the euro-area countries we consider samples of similar lengths as for the NMS countries, covering second half of the 1980's up to 1998 (the start of the EMU). As a summary, we present here the results for a longer sample spanning 1985 (1) - 1998 (12). These results are similar to those for the shorter samples, but free of some features which were deemed not robust.

The data is monthly. The endogenous variables: output, prices, interest rates and exchange rates are measured respectively by log of the Index of Industrial Production (IIP), log of the Consumer Price Index (CPI), short term market interest rate (r-mkt) and the log of the exchange rate in national currency units per SDR. The SDR is a standard basked of main currencies, and it provided an intermediate choice between a US dollar exchange rate,

[7]We follow Mojon and Peersman (2001) who also consider them separately for this reason.

which is used often, but may be influenced by some US specific events, and country specific baskets of most relevant currencies, which are less comparable. Most data (as well as those for the exogenous variables: Federal Funds Rate, oil prices, non-fuel commodity prices, German interbank interest rate and German industrial production) are taken from the IMF IFS database, and some from the Eurostat. See the appendix for details.

The interest rate of 0.1 corresponds to 10% (1000 basis points). The variables other than the interest rate are logs of indexes that assume the value 1 in December 1995. The basic specification contains six lags of the endogenous variables and lags zero and one of the exogenous variables. The exception is German industrial production (included as an exogenous variable for countries other than Germany): it is assumed that foreign central banks observe it with a lag, so lags one and two are included. Shorter lag length of the exogenous variables is chosen to conserve the degrees of freedom.

1.5 Results

We approximate the posterior distribution of the estimated coefficients using the Gibbs sampler[8] and compute impulse responses over the horizon of 40 months.

1.5.1 'Mean' impulse responses for the regions

Figure 1.1 presents the 5th, 50th and 95th percentiles of the posterior distribution of impulse responses to a one standard deviation monetary shock, implied by the 'mean' model for each panel ($\bar{\boldsymbol{\beta}}$).

The immediate (period 0) responses of all variables reflect the identifying assumptions: 1) in the month of the shock the output and prices are unaffected and 2) the interest rate raises and the exchange rate falls (appreciates). The uncertainty band for the impact behavior of the interest rate and exchange rate ranges from the 5th to the 95th percentile of all the range where the sign restrictions are satisfied.

The identified MP shock is associated with a median interest rate increase of 40 basis points in the EMU5 and almost 80 bp in the NMS4. The median initial appreciation is respectively 1% and 1.5%. The interest rate increase is reversed after about one year, while the appreciation persists for about 2 years. The economies respond with a transitory output decline and a possibly permanent reduction of the price level.

[8]We generate 2000 draws from the posterior, after discarding the initial 4000 burn-in draws.

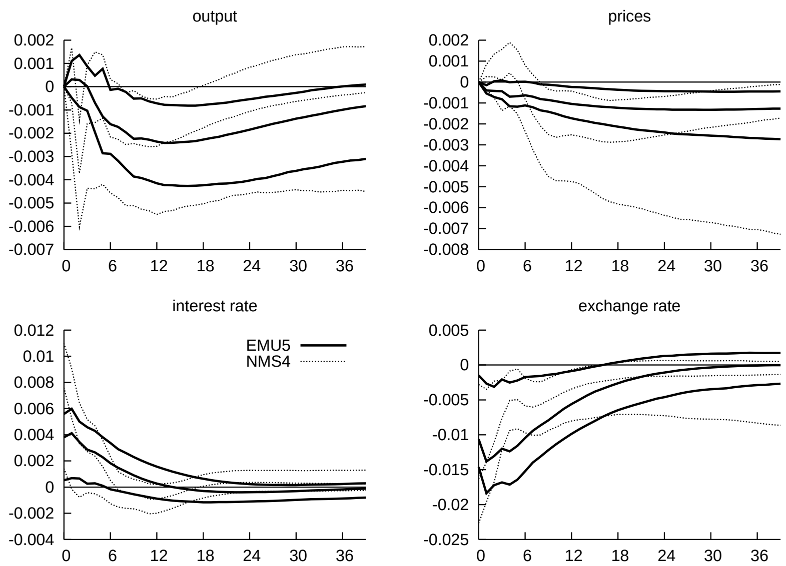

Figure 1.1: Mean impulse responses for the EMU5 and the NMS4: median, 5th and 95th percentiles of the posterior distribution

The finding that interest rate movements in the NMS4 were on average twice stronger than those in the EMU5 is not surprising: First, output growth rates, inflation levels and interest rates in the post-communist countries tended to be higher than in the Western Europe, which was likely to generate higher variance of shocks. Second, the central banks in the region may have believed that, because of the low financial depth in these countries, monetary policy is not very effective there (Ganev et al., 2002), and therefore their policy actions require more vigorous interest rate movements.

A look at the variance decompositions (figure 1.2) suggests that the higher monetary shocks in NMS4 contributed importantly to the variability of output and especially prices in that region. According to the median of the distribution, around 10% of the variability of output is attributed to monetary shocks in both countries. In case of prices, the shares are about 20% in the EMU5 and 30% in NMS4. As usual in the VAR literature, there is a wide uncertainty about the exact figures. The posterior distributions for output overlap significantly, but monetary shocks are more likely to be responsible for a greater share of variance of prices in the NMS4 than in the EMU5.

In order to control for the different size of shocks we standardize the impulse responses, to make them correspond to the same size of the interest rate shock in the first month. We take the EMU5 shock as a benchmark.

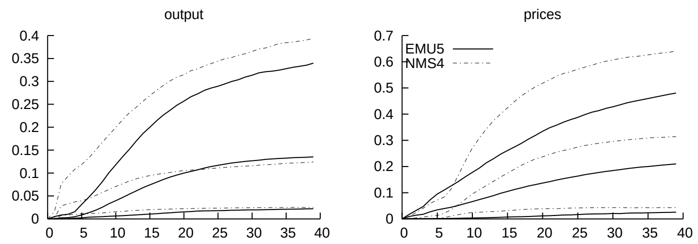

Figure 1.2: Share of monetary shocks in the variance decompositions for the NMS4 and the EMU5 panels: median, 5th and 95th percentiles of the posterior distributions

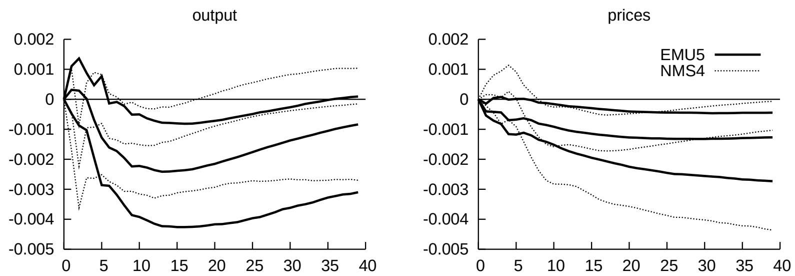

Figure 1.3: Mean impulse responses of output and prices in the EMU5 and the NMS4 corresponding to the same size interest shock in the first month

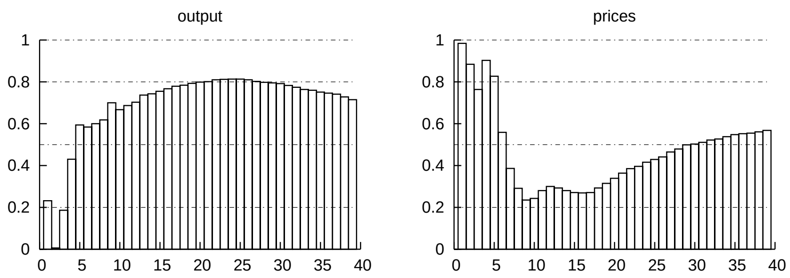

Figure 1.4: Posterior probability that the standardized response of the NMS4 (from figure 1.3) panel is weaker than that of the EMU5 panel

Therefore, we scale down the NMS4 impulse responses, so that the average of the impact and first month response of the interest rate is the same as in the EMU5 panel. (We take the average of impact and first lag effects in order to neutralize partly the different dynamics of interest rates in both panels). Figure 1.3 presents the standardized output and price responses, and figure 1.4 shows the probability that the standardized NMS4 response is weaker than that of the EMU5 panel.

The standardized impulse responses of both regions are quite similar, their 90% probability regions are mostly overlapping. In particular, as in the variance decompositions, we don't find a straightforward confirmation of the belief that the monetary policy is less effective in the NMS4, because of the low financial depth of these countries. It is true that output, in the medium term, responds rather weakly. Also, the price responses are more lagged in the NMS4, and need over 6 months to become significant. However, after this delay, they react more vigorously than in the EMU5.

The observation that output responses tend to be weak, while those of prices quite strong, suggests that the sacrifice ratio facing the NMS4 central bankers could be lower. This possibility is examined closer by comparing the posterior distributions of the sacrifice ratios for both panels, calculated as in Cecchetti and Rich (2001). Their comparison suggests that the sacrifice ratio in the NMS4 might indeed be lower, although the posterior probability of this statement is only 72% with the 36 months horizon.[9]

Central banks in the NMS have shorter track records, and probably enjoy less credibility than their Western European counterparts. If this is the case, their monetary policy has less impact on agent's expectations, which results in longer lags in the response of prices.

The longer lags of price response in the NMS4 panel can be complementarily explained by models of learning in an evolving setup (see e.g. Evans and Honkapohja, 2001). Initially, because of undeveloped financial markets, Central Bank discount rates mattered very little for inflation. As financial markets and credit activity grows, so does the importance of the interest rates, and agents revise estimates of their impact on inflation, and readjust their expectations. However, the nature of learning under uncertainty implies that they adapt their models only partially. As a result, the impact of the Central Bank policies is always initially underestimated, and takes full effect with a delay, after expectations adjust.

However, the most unexpected aspect of the results in figures 1.3 and

[9]This probability exceeds 80% with shorter horizons. The reason why the sacrifice ratios become more similar for longer horizons is that the price decline gets reversed in the long run. The very long run behavior of prices may be, however, less reliably estimated than the medium run behavior.

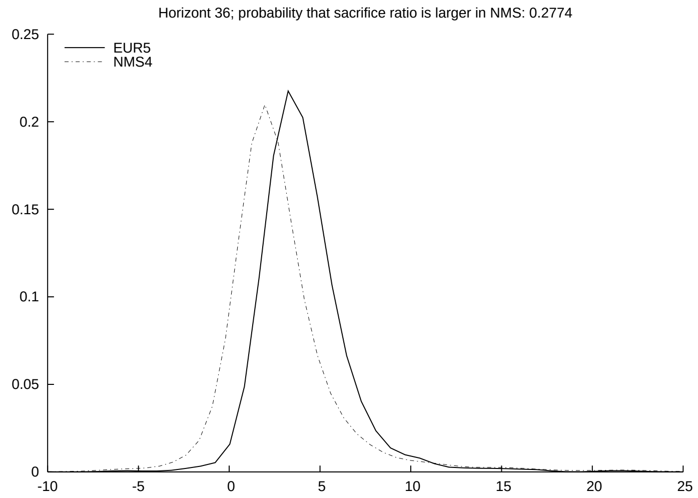

Figure 1.5: Posterior distribution of the sacrifice ratio (36 months horizon)

1.4 is that, after a few months' lag, monetary policy seems to have quite a strong impact on prices in the NMS. This is in spite of the fact, that in the NMS4 indicators of the size of the financial systems, such as the ratio of financial assets/liabilities, or stock market capitalization, to GDP, are lower by a factor of 2 to 5 in comparison with the euro-area average (see Anzuini and Levy, 2004). Apparently, we need to go beyond the simple rule of thumb, that monetary policy is less effective in less financially developed countries, when comparing the Central-Eastern and the Western Europe.

The possibility of lower sacrifice ratios in the NMS suggest other theories, which may be relevant here: The NMS4 have more volatile and, on average, higher inflation rates. Lucas (1973) argues that in an imperfect information model, in countries where aggregate demand fluctuates more, agents adjust their prices more than their outputs. Ball et al. (1988)'s reasoning is, that in higher inflation countries, agents need to adjust their prices more often, and so there is less stickiness. Both models imply, that in such countries aggregate supply curve is steeper, and, consequently, the sacrifice ratio lower, and this is confirmed in their cross-country studies.

There are reasons to believe, that the NMS economies should be responsive to monetary policy: First, the exchange rate channel might be stronger. The NMS4 economies are more open. Moreover, they have less established brands where monopolistic competition is important, so their exports can be more sensitive to exchange rates.

Second, with less developed financial markets, agents may find it more difficult to hedge against the monetary policy changes. Anzuini and Levy (2004) find that the agents in the NMS4 have mostly unhedged foreign debt, and are very exposed to the foreign exchange rate risk. Third, firms in NMS4 find it harder to obtain credit abroad, when domestic credit conditions are tight.

Fourth, even with small institutional financial markets, the prevailing interest rates might still matter for economic decisions and transactions, such as the trade credits, or reinvestment of profits. The NMS4 have a high volume of trade credit (see Anzuini and Levy, 2004) and, since they are catching-up economies, they have more investments compared with GDP than the euro area (Suppel, 2003), which is mostly financed by reinvesting profits.

Summarizing, the story behind figure 1.3 could be following: In the NMS, the policy tightening by the central bank is less credible, and so has little initial effect on expectations, and thus on pricing decisions. The effect of higher interest rates on output is weaker than in the EMU, because of the small financial markets, but not much weaker, in light of the arguments listed above. In addition, this modest aggregate demand contraction is translated more efficiently into prices, because the last are less sticky.

1.5.2 Heterogeneity within EMU and NMS

Heterogeneity of the panels is best reflected by the posterior distribution of the overall tightness parameter λ. As follows from the statistical structure in equations (1.3)-(1.7), probability mass concentrated on low values of λ means that the posterior distributions of country VAR models are close to each other. When, on the contrary, the posterior distributions of parameters for individual countries tend to differ, higher values of the λ are more likely, and get more posterior support.

In the two panels estimated here, the likelihood turns out to be very informative about λ and the support of the posterior distribution is very narrow. This can be assessed by comparing the shape of the impulse responses for e.g. the 5th and the 95th centiles of the simulated distribution of λ: the graphs are almost indistinguishable. Therefore, the data favors a certain intermediate amount of cross-country information pooling, preferring it both to complete homogeneity, and to independence of the individual country models.

The posterior density of the overall tightness parameter λ in each of the panels is presented in figure 1.6. The NMS4 panel turns out to be more heterogeneous.

Figures 1.7 and 1.8 present impulse responses to a monetary shock for each of the analyzed countries. The considered samples are short: for individual countries, they imply only 3.1 observations per estimated parameter in the

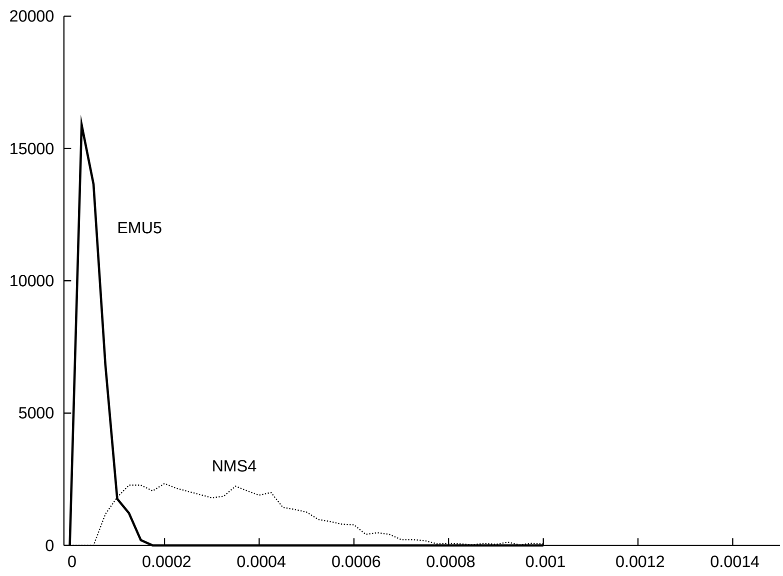

Figure 1.6: Simulated posterior distribution of λ for the euro-area panel and for the NMS4 panel

NMS4 and 2.4 for the Czech Republic, and maximally no more than 5 in the euro area. For the NMS4 countries these are maximum samples under flexible exchange rate regime. In the standard, individual country estimation, results based on such short data would have to be treated with much caution. The results presented here are more reliable, because in the computation of the posterior, the country data is optimally augmented with the information for all other countries in the panel.

Consistently with what the posterior distribution of the tightness parameter λ is suggesting, the impulse responses for EMU5 look more homogeneous than those for NMS4. In the EMU5 panel, Germany is the clear outlier, with the weakest monetary shocks (around 15 bp) and much uncertainty about output and price responses, which renders them insignificant. In Finland, France and Italy the median interest rate shock is around 40 bp, and 50 bp in Spain, which becomes insignificant after about one year. Exchange rate appreciates similarly, by 1-1.5% in all countries. The peak of output response comes after one year, at -0.2% in France, Italy and Spain, and in Finland output reaction is strongest, at -0.3%. Prices fall gradually by about 0.1% in Finland and France, and by about 0.2% in Italy and Spain, which suggests that the latter two may have lower sacrifice ratios, which could be linked with their higher inflation rates

In the NMS4 panel the median interest rate tightening is strongest in Poland (about 90 bp), somewhat weaker in Czech Republic and Slovenia (around 60 bp), and weakest in Hungary, at 40 bp, which is similar as in the EUR5 countries. Exchange rate appreciates by roughly as much as in the EUR5 countries, possibly somewhat stronger (the comparison is blurred by different dynamics: in the NMS the exchange rate appreciates most on impact, while in the EMU the peak response comes after one month), except in Czech Republic, where the appreciation is clearly stronger, at more than 2%. Output falls by about 0.2% in all countries, except Czech Republic where the contraction reaches -0.3%. Prices fall by about 0.2% in Hungary, 0.25% in Slovenia and 0.3% in Poland, and stabilize at the new level after about one year. Czech Republic is an exception, because prices fall by almost 0.4% within the first year, but seem to increase back afterwards. Overall, in Poland, which is an inflation targeter, with largely uncontrolled exchange rate, monetary policy is most volatile, and has most impact. Hungary, where the exchange rate is kept within the narrowest band (among the analyzed countries), sees least impact of monetary policy. This is consistent with the exchange rate channel being crucial for monetary transmission in the region.

1.5.3 Robustness

The results have been checked for robustness to changing the country composition of the panels, sample periods, specification of the VARs, control variables and the identification scheme. The main conclusions go through under those experiments.

Given the short sample size, and the overparametrization of the VAR model, individual countries results are sensitive to changes in the sample period. However, the conclusions for the mean model are generally robust to changing the sample or removing any of the countries from the panel. As mentioned before, market interest rates for Slovakia and Romania hardly follow the central bank interest rates, so these countries were skipped in the basic estimation. When they are included, results for Slovakia are odd since output responds positively, but mean responses are unaffected. Responses for Romania are reasonable, but the average size of the interest rate shock is very large, more than 500 basis points (Romanian market interest rates are very volatile) and this increases the size of the mean interest rate shock to about 250 bp. Other impulse responses are barely affected. Similar situation concerns Portugal, which is an outlier in the euro-area panel, because the unusual swings of its interest rates in the 1992/1993. During the EMS crisis, the Portuguese Central Bank was for some time fending off speculative attacks on the escudo by interest rates increases, which needed to be more

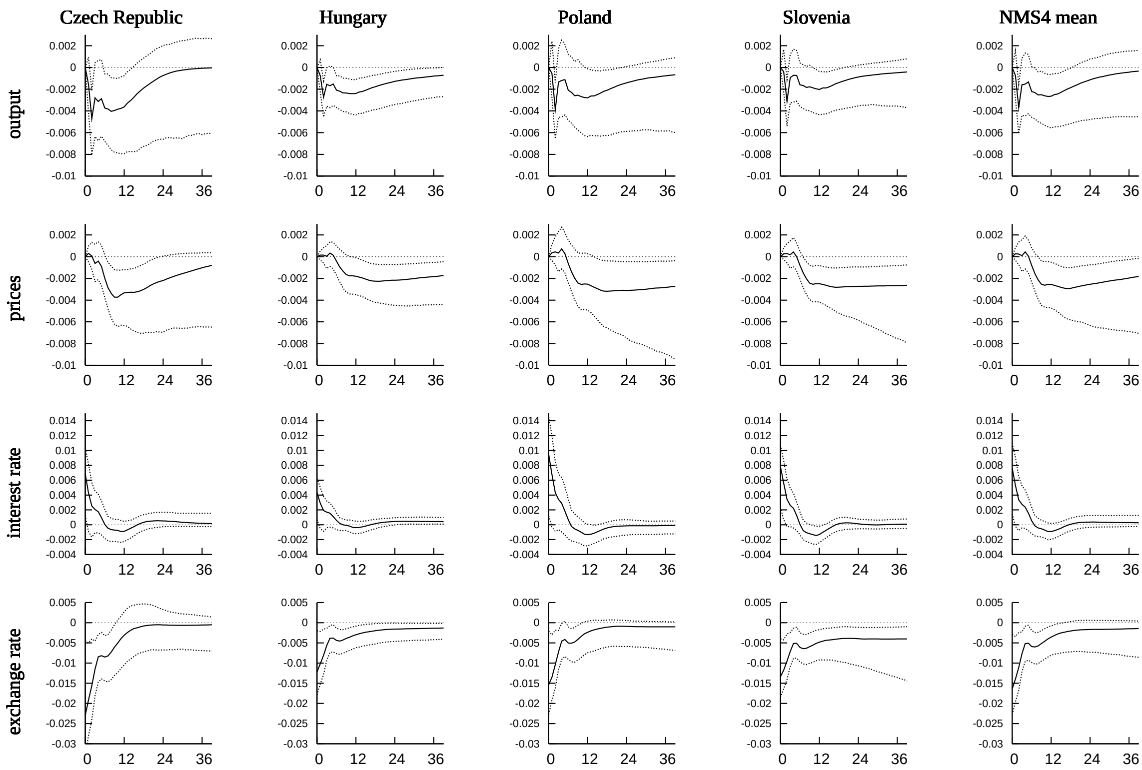

Figure 1.7: Impulse responses in the EMU5 panel

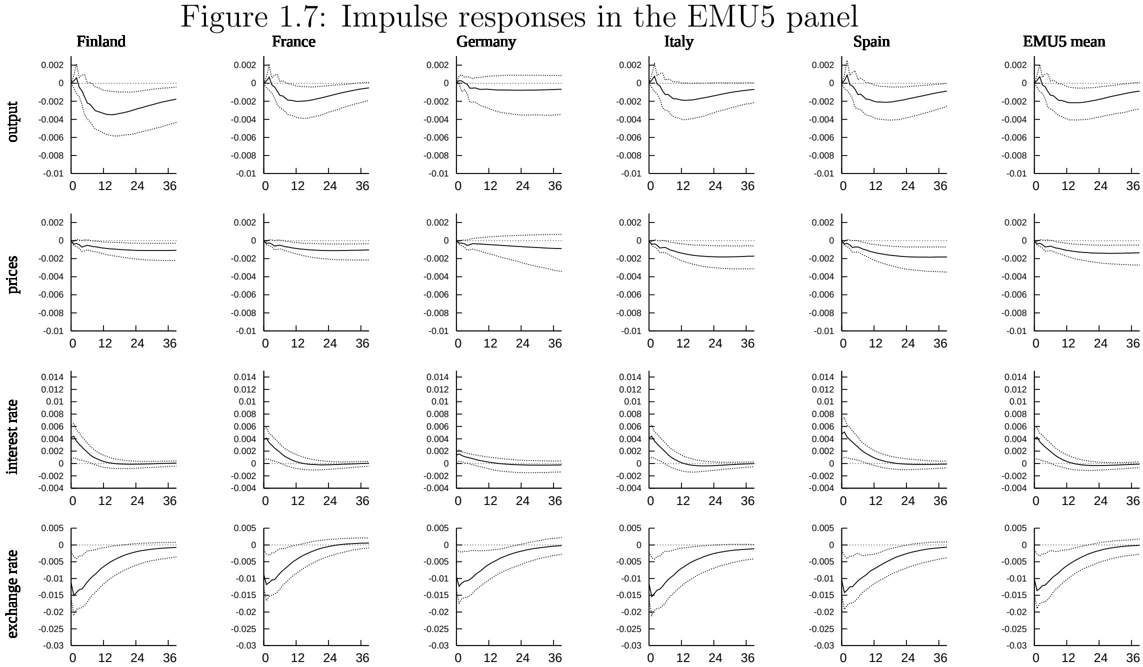

Figure 1.8: Impulse responses in the NMS4 panel

dramatic than in other countries affected by the crisis, partly because of the relatively small volume of the market. The resulting interest rate swings were perceived as temporary anomalies and did not have usual real effects. Therefore, including Portugal in the panel results in a larger size of the interest rate impulses, while barely affecting other variables.

For the NMS4 panel, there is little scope for varying the sample sizes, as they are already very short, but the results are similar for one year shorter samples. For the euro-area panel, we try 9 year samples (roughly equal to the typical sample size for the NMS4 countries) spanning the 1980s until 1998. 9 year samples between 1985 and 1998 produce similar results, which are best characterized by those for the longer 1985-1998 sample. For 9 year samples starting in the first half of the 1980s impulse responses are often insignificant. Apparently, the model lacks features necessary to explain data from early 1980s, but we assume that this does not hamper the comparison performed in this paper.

In another set of experiments, Central Bank interest rates were used instead of the market interest rates. Most of the VAR studies assume, as does this paper, that the Central Bank targets interbank market rates, and the transmission from this intermediate target to the economy is studied. When central bank interest rates are used, comparisons with both the basic NMS4 panel, and the full one including Slovakia and Romania, lead to similar conclusions.

Replacing the SDR exchange rate with the USD exchange rate makes almost no difference for both sets of countries.

If the model estimated without any control variables (only constant terms and lags of endogenous variables), the output and price responses in both panels become more delayed, but deeper in the medium run, compared with the case with all controls included. The comparison of the responses, however, remains unaffected. The contemporaneous German interest rate makes most of the difference, adding further lags does not make a difference. Similarly, adding more lags or changing the set of world control variables often affects individual country responses, but the comparison of the mean responses are unaffected. After some experimentation, it was deemed best to stick with the maximal set of controls and avoid the risk of misspecifying the model. It is hoped that the estimation procedure can 'average out' the noise introduced by possibly excessive number of control variables.

Finally, the analysis was repeated with the recursive identification (i.e. assuming that monetary authorities react with a lag to the exchange rate developments). This identification results in a 'price puzzle': price response is initially positive. For the euro area the 'price puzzle' is not very significant, lasts only about one year, output responds negatively, but the exchange rate seems to depreciate. For the NMS4 panel the positive price response is very strong and output also responds positively.

1.6 Conclusions

This paper makes one of the first systematic comparisons of the responses to monetary shocks in Western Europe and in the New Member States of the EU. The responses of the NMS4 (Czech Republic, Hungary, Poland and Slovenia) turn out to be broadly similar to those in the EMU, but with interesting differences (albeit estimated with significant uncertainty): monetary shocks tend to be stronger, and generate a more delayed, but strong price level response, possibly at lower output cost.

These results suggest that, when considering the differences between the Central-Eastern and the Western Europe, we need to go beyond the simple rule of thumb, that monetary policy is less effective in less financially developed countries. Some of the structural features of the NMS financial systems (less possibilities for hedging, harder access to foreign financial markets), and their export orientation (strong exchange rate channel), may be amplifying the effects of monetary shocks on aggregate demand. Strong effect on prices, with possibly lower sacrifice ratios, that we find in the NMS, are consistent with the findings of Lucas (1973) and Ball et al. (1988), that in economies where the aggregate demand fluctuates more, and inflation is higher, the aggregate supply curve tends to be steeper, and prices less sticky.

Conclusions from the empirical analysis of this paper for the prospects of the EMU accession are not obvious, because of the Lucas critique. VARs are useful for establishing stylized facts about past monetary policy effects, but we don't know to what extent monetary transmission will change in the wake of the EMU accession. Several mechanisms may be at play:

After joining the EMU, the exchange rate channel, which is important for the NMS4 currently, will largely disappear, and domestic firms will gain easier access to wider financial markets. This will considerably weaken the responsiveness of the NMS to monetary policy. The result may be similar to the scenario predicted by Dornbush and observed within the present EMU: that interest rates that are optimal for the European core are too low for the peripheral, poorer and faster growing countries. As a result, the peripheries observe persistently higher inflation rates and may end up developing financial and property markets bubbles.

Overall, however, the results in this paper downplay the importance of the structurally determined weakness of monetary transmission in the NMS. Instead, they bring to the front the issue of longer response lags. If the observed longer lags in price reaction result from lower credibility of the NMS4 central banks, the credibility problem may be solved overnight by joining the monetary union. If learning under uncertainty is the issue, it could be speeded up by a change towards a less uncertain environment.

Finally, the results of this paper provide the following argument to the proponents of the EMU accession: Given the long lags with which the transmission mechanism operates in the NMS4, and the ultimate strength of its effect, it should be harder for the NMS4 central banks to run an effective stabilizing monetary policy. The independent monetary policy is more likely to be a source of additional variability and giving it up could end up being beneficial in the long run.

Appendix 1.A Conditional posteriors for the Gibbs sampler

In the model defined by equations (1.3)-(1.7) the joint posterior is $\propto$

$$\prod_i |\boldsymbol{\Sigma}_i|^{-\frac{T_i}{2}} \exp\left(-\frac{1}{2}\sum_i (\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\beta}_i - \tilde{\boldsymbol{Z}}_i\boldsymbol{\gamma}_i)'(\boldsymbol{\Sigma}_i^{-1}\otimes \boldsymbol{I}_{T_i})(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\beta}_i - \tilde{\boldsymbol{Z}}_i\boldsymbol{\gamma}_i)\right)$$
$$\lambda^{-\frac{IJK}{2}} \exp\left(-\frac{1}{2}\sum_i (\boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}})'\lambda^{-1}\boldsymbol{L}_i^{-1}(\boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}})\right)\prod_i |\boldsymbol{\Sigma}_i|^{-\frac{J+1}{2}}\lambda^{-\frac{v+2}{2}}\exp\left(-\frac{1}{2}\frac{s}{\lambda}\right) \tag{1.14}$$

Define all data, $Y \equiv \{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_I, \boldsymbol{X}_i, \ldots, \boldsymbol{X}_I, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_I\}$ and the set of parameters, $\Theta \equiv \{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_I, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_I, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_I, \bar{\boldsymbol{\beta}}, \lambda\}$.

Conditional posterior for $\boldsymbol{\beta}_i$ is:

$$p(\boldsymbol{\beta}_i|Y, \Theta\backslash\{\boldsymbol{\beta}_i\}) = \mathrm{N}(\boldsymbol{D}_i^{-1}\boldsymbol{d}_i, \boldsymbol{D}_i^{-1}) \tag{1.15}$$

where

$$\boldsymbol{D}_i = \boldsymbol{\Sigma}_i^{-1}\otimes \boldsymbol{X}_i'\boldsymbol{X}_i + \lambda^{-1}\boldsymbol{L}_i^{-1}$$
$$\boldsymbol{d}_i = (\boldsymbol{\Sigma}_i^{-1}\otimes \boldsymbol{X}_i')\,\mathrm{vec}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\boldsymbol{\Gamma}_i) + \lambda^{-1}\boldsymbol{L}_i^{-1}\bar{\boldsymbol{\beta}}$$

Conditional posterior for $\boldsymbol{\gamma}_i$ is:

$$p(\boldsymbol{\gamma}_i|Y, \Theta\backslash\{\boldsymbol{\gamma}_i\}) = \mathrm{N}(\boldsymbol{F}_i^{-1}\boldsymbol{f}_i, \boldsymbol{F}_i^{-1}) \tag{1.16}$$

where

$$\boldsymbol{F}_i = \boldsymbol{\Sigma}_i^{-1}\otimes \boldsymbol{Z}_i'\boldsymbol{Z}_i$$
$$\boldsymbol{f}_i = (\boldsymbol{\Sigma}_i^{-1}\otimes \boldsymbol{Z}_i')\,\mathrm{vec}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{B}_i) = \mathrm{vec}(\boldsymbol{Z}_i'(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{B}_i)\boldsymbol{\Sigma}_i^{-1})$$

Conditional posterior for $\bar{\boldsymbol{\beta}}$ is:

$$p(\bar{\boldsymbol{\beta}}|Y,\Theta\backslash\{\bar{\boldsymbol{\beta}}\}) = \mathrm{N}(\boldsymbol{G}_i^{-1}\boldsymbol{g}_i, \boldsymbol{G}_i^{-1}) \tag{1.17}$$

where

$$\boldsymbol{G}_i = \lambda^{-1}\sum_i \boldsymbol{L}_i^{-1}$$
$$\boldsymbol{g}_i = \lambda^{-1}\sum_i \boldsymbol{L}_i^{-1}\boldsymbol{\beta}_i$$

Conditional posterior for $\boldsymbol{\Sigma}_i$ is:

$$p(\boldsymbol{\Sigma}_i|Y,\Theta\backslash\{\boldsymbol{\Sigma}_i\}) \propto |\boldsymbol{\Sigma}_i|^{-\frac{T_i+J+1}{2}} \exp\left(-\frac{1}{2}\,\mathrm{tr}\,\boldsymbol{\Sigma}_i^{-1}\boldsymbol{U}_i'\boldsymbol{U}_i\right)$$

or

$$p(\boldsymbol{\Sigma}_i|Y,\Theta\backslash\{\boldsymbol{\Sigma}_i\}) = \mathrm{IW}(\boldsymbol{U}_i'\boldsymbol{U}_i, T_i) \tag{1.18}$$

where IW denotes the Inverted Wishart distribution (Bauwens et al., 1999, p.305).

Conditional posterior for λ is:

$$p(\lambda|Y,\Theta\backslash\{\lambda\}) \propto \lambda^{-\frac{IJK+v+2}{2}} \exp\left(-\frac{1}{2}\frac{(s+\sum_i(\boldsymbol{\beta}_i-\bar{\boldsymbol{\beta}})'\boldsymbol{L}_i^{-1}(\boldsymbol{\beta}_i-\bar{\boldsymbol{\beta}}))}{\lambda}\right)$$

or

$$p(\lambda|Y,\Theta\backslash\{\lambda\}) = \mathrm{IG}_2\left(s+\sum_i(\boldsymbol{\beta}_i-\bar{\boldsymbol{\beta}})'\boldsymbol{L}_i^{-1}(\boldsymbol{\beta}_i-\bar{\boldsymbol{\beta}}), IJK+v\right) \tag{1.19}$$

where IG_2 denotes the inverted gamma-2 distribution (Bauwens et al., 1999, p.292).

Simulating the posterior with the Gibbs sampler consists of randomly drawing parameters in Θ from (1.15)-(1.19), always conditioning on all most recently drawn parameters. See e.g. Gelman et al. (1995) for a detailed discussion of the Gibbs sampler.

Appendix 1.B Imposing the sign restrictions

The sign restrictions in (1.13) can be viewed as the priors for the nonzero coefficients of the $\boldsymbol{G}$ which are uniform on the whole real line, on the positive or on the negative part of the real line. The requirement $\boldsymbol{G}\boldsymbol{G}' = \boldsymbol{\Sigma}$ constrains the coefficients to certain intervals, but otherwise all factorizations of the $\boldsymbol{\Sigma}$

are observationally equivalent, i.e. they result in exactly the same value of the likelihood function. Therefore, within this family, the posterior distribution coincides with the prior. The snug is that the coefficients of the $\boldsymbol{G}$ are linked by a nonlinear relationship and they cannot be simultaneously uniform on their admissible intervals.

Technically the sign restrictions are applied in a manner following Kieler and Saarenheimo (1998) and Canova and De Nicoló (2002), obtaining one factorization from another by means of a rotation matrix. The difference is that, thanks to the combination of sign restrictions with some zero restrictions, the resulting search for admissible rotations can be performed analytically, avoiding the computationally intensive numerical search technique of the above papers.

For any factorization $\boldsymbol{G}^*$ such that $\boldsymbol{G}^*\boldsymbol{G}^{*'} = \boldsymbol{\Sigma}$, and any orthogonal matrix $\boldsymbol{D}$, $\boldsymbol{G}^{**} = \boldsymbol{G}^*\boldsymbol{D}$ is also a factorization, i.e. satisfies $\boldsymbol{G}^{**}\boldsymbol{G}^{**'} = \boldsymbol{\Sigma}$. All orthogonal matrices (which correspond to orthogonal linear transformations) are products of sequences of rotations and reflections. Start from the Choleski decomposition of the $\boldsymbol{\Sigma}$. If the zeros in the first two rows are to be preserved, only rotations of the last two columns are allowed. The restriction of diagonal elements to be positive allows us to disregard reflections. Therefore, all matrices satisfying the zero restrictions can be obtained as:

$$\boldsymbol{G}(\theta) = \text{Chol}(\boldsymbol{\Sigma}) \times \text{Rotation}(3, 4, \theta) \tag{1.20}$$

where Chol() denotes the Choleski decomposition and Rotation(x,y,θ) is the matrix that rotates columns x and y by angle θ. Writing the above equation in detail:

$$\boldsymbol{G}(\theta) = \begin{pmatrix} c_{11} & 0 & 0 & 0 \\ c_{21} & c_{22} & 0 & 0 \\ c_{31} & c_{32} & c_{33} & 0 \\ c_{41} & c_{42} & c_{43} & c_{44} \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos(\theta) & -\sin(\theta) \\ 0 & 0 & \sin(\theta) & \cos(\theta) \end{pmatrix}$$

When multiplying out the above matrices, the four sign restrictions on the lower right submatrix of $\boldsymbol{G}$, spelled out in equation (1.13), imply the system of four inequalities:

$$\begin{aligned} c_{33}\cos\theta &> 0 \\ -c_{33}\sin\theta &> 0 \\ c_{43}\cos\theta + c_{44}\sin\theta &< 0 \\ -c_{43}\sin\theta + c_{44}\cos\theta &> 0 \end{aligned}$$

The above system can be solved for the rotation angle θ. The solution depends on the term c_{43}:

$$\theta \in \left(-\frac{\pi}{2}, \arctan\left(-\frac{c_{43}}{c_{44}}\right)\right) \text{ when } c_{43} > 0, \text{ and} \tag{1.21}$$

$$\theta \in \left(\arctan\left(\frac{c_{44}}{c_{43}}\right), 0\right) \text{ when } c_{43} < 0 \tag{1.22}$$

Going through the found range of rotation angles (and postmultiplying the Choleski decomposition of the residual variance by the resulting rotation matrices) we find all the matrices $\boldsymbol{G}$ satisfying the postulated zero and sign restrictions.

In reporting the results, one would like to integrate them over the posterior distribution of θ, which, on the admissible interval, coincides with the prior (since θ doesn't change the value of the likelihood function). The prior can be inferred from the prior for the elements of $\boldsymbol{G}$, but here we stumble on the mentioned problem: $p(\theta) \propto \cos(\theta)$, which corresponds to the uniform distribution of $\boldsymbol{G}_{(3,3)}$, results in highly skewed distributions of $\boldsymbol{G}_{(3,4)}$, $\boldsymbol{G}_{(4,3)}$ and $\boldsymbol{G}_{(4,4)}$, etc. As a compromise, we report results integrated over the *uniform* distribution of θ on its admissible interval, which produces moderately skewed distributions of all parameters. Results obtained with other candidate distributions turned out to be very similar and no conclusions are affected.

Therefore the computation of the posterior is following: for each draw of the residual variance matrix (obtained from the Gibbs sampler) 1) the Choleski decomposition is found, 2) the admissible range for θ is computed from the formula (1.21) or (1.22), 3) a random number is drawn from the uniform distribution on the computed range for θ, 4) matrix $\boldsymbol{G}$ is obtained with formula (1.20).

Bibliography

Anzuini, A. and Levy, A. (2004). Financial structure and the transmission of monetary shocks: Preliminary evidence for the Czech Republic, Hungary and Poland. Bank of Italy, Research Department.

Ball, L., Mankiw, G. N., and Romer, D. (1988). The new Keynesian economics and the output-inflation trade-off. *Brookings Papers on Economic Activity*, 1988(1):1–82.

Bauwens, L., Lubrano, M., and Richard, J.-F. (1999). *Bayesian Inference in*

Dynamic Econometric Models. Advanced Texts in Econometrics. Oxford University Press, Oxford, first edition.

Canova, F. (2005). The transmission of US shocks to Latin America. *Journal of Appplied Econometrics*. forthcoming.

Canova, F. and Ciccarelli, M. (2004). Forecasting and turning point predictions in a Bayesian panel VAR model. *Journal of Econometrics*, 120(2):327–359.

Canova, F. and De Nicoló, G. (2002). Monetary disturbances matter for business fluctuations in the G-7. *Journal of Monetary Economics*, 49:1131–1159.

Canova, F. and Marcet, A. (1995). The poor stay poor: Non-convergence across countries and regions. UPF Working Paper no.137, Revised: March 1998.

Cecchetti, S. G. and Rich, R. W. (2001). Structural estimates of the U.S. sacrifice ratio. *Journal of Business and Economic Statistics*, 19(4):416–27.

Creel, J. and Levasseur, S. (2005). Monetary policy transmission mechanisms in the CEECs: How important are the differences with the euro area? OFCE Document de Travail N 2005-02.

Dornbusch, R., Favero, C., and Giavazzi, F. (1998). Immediate challenges for the European Central Bank. *Economic Policy*, 26:15 – 64.

Ehrmann, M., Gambacorta, L., Martinez-Pags, J., Sevestre, P., and Worms, A. (2003). The effects of monetary policy in the euro area. *Oxford Review of Economic Policy*, 19(Supplement 1):58–72.

Elbourne, A. and de Haan, J. (2005). Financial structure and monetary policy transmission in transition countries. mimeo, University of Groningen and CESifo.

Evans, G. W. and Honkapohja, S. (2001). *Learning and Expectations in Macroeconomics*. Princeton University Press.

Faust, J. (1998). The robustness of identified VAR conclusions about money. *Carnegie-Rochester Conference Series on Public Policy*, 49:207–244.

Ganev, G., Molnar, K., Rybiński, K., and Woźniak, P. (2002). Transmission mechanism of monetary policy in Central and Eastern Europe. CASE Report No. 52.

Gelman, A. B., Carlin, J. S., Stern, H. S., and Rubin, D. B. (1995). *Bayesian Data Analysis*. Chapman and Hall Texts in Statistical Science. Chapman and Hall/CRC, first edition.

Guiso, L., Kashyap, A. K., Panetta, F., and Terlizzese, D. (1999). Will a common european monetary policy have asymmetric effects? *Economic Perspectives, Federal Reserve Bank of Chicago*, pages 56–75.

Hobert, J. P. and Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, 91(436):1461–1473.

Hsiao, C., Pesaran, M. H., and Tahmiscioglu, A. K. (1999). Bayes estimation of short-run coefficients in dynamic panel data models. In C. Hsiao, K. Lahiri, L.-F. L. and Pesaran, M., editors, *Analysis of Panels and Limited Dependent Variables: A Volume in Honour of G S Maddala*, chapter 11, pages 268–296. Cambridge University Press, Cambridge.

Kieler, M. and Saarenheimo, T. (1998). Differences in monetary policy transmission? a case not closed. European Commission Economic Papers No. 132.

Kim, S. (1999). Do monetary policy shocks matter in the G-7 countries? using common identifying assumptions about monetary policy across countries. *Journal of International Economics*, 48:387–412.

Kim, S. and Roubini, N. (2000). Exchange rate anomalies in the industrial countries: A solution with a structural VAR approach. *Journal of Monetary Economics*, 45:561–586.

Leeper, E. M., Sims, C. A., and Zha, T. (1996). What does monetary policy do? *Brookings Papers on Economic Activity*, (2).

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions - five years of experience. *Journal of Business and Economic Statistics*, (4):25–38.

Lucas, R. E. (1973). Some international evidence on output-inflation tradeoffs. *The American Economic Review*, 63(3):326–334.

Mihov, I. (2001). Monetary policy implementation and transmission in the European Monetary Union. *Economic Policy*, 33:371–406. with discussion.

Mojon, B. and Peersman, G. (2001). A VAR description of the effects of monetary policy in the individual countries of the Euro area. ECB Working Paper no.92.

Pesaran, H. M. and Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, 68:79–113.

Ramaswamy, R. and Slok, T. (1998). The real effects of monetary policy in the European Union: What are the differences? *IMF Staff Papers*, 45(2):374–391.

Sims, C. A. and Zha, T. (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, 39(4):949–968.

Suppel, R. (2003). Comparing economic dynamics in the EU and CEE accession countries. ECB working paper no. 267.

Uhlig, H. (2001). What are the effects of monetary policy on output? results from an agnostic identification procedure. Tilburg University. Center for Economic Research Discussion paper 1999-28.

Zellner, A. and Hong, C. (1989). Forecasting international growth rates using Bayesian shrinkage and other procedures. *Journal of Econometrics*, 40:183–202.

Chapter 2

Prior for Growth Rates, Small Sample Bias and the Effects of Monetary Policy

Joint with Albert Marcet[1]

Abstract: A Bayesian with a flat prior would say that OLS is the optimal estimator in a VAR. However, from a classical point of view, the well-known bias of OLS should be corrected. We show that OLS appears optimal to a Bayesian because the flat prior puts a very large weight on parameters that imply huge growth rates in the first few periods of the sample. This leads us to propose a prior that excludes huge growth rates. We argue that this prior, which is readily acceptable to most economists, breaks the classical/Bayesian dichotomy, as it can be interpreted as delivering a bias correction, and it has several advantages over proposed corrections for the bias. We show how to compute the posterior exactly and with a practical shortcut. We illustrate its effect in a VAR of Christiano et al. (1999), and show that the small sample bias causes OLS to understate the impact of monetary shocks on output.

2.1 Introduction

The good finite-sample properties of OLS (unbiasedness and minimum MSE) only hold under the assumption of exogenous regressors. In particular, in a time series context, it is well known that OLS is biased and that it underestimates the highest root of the process for the relevant case of a root close to

[1]Institut d'Anàlisi Econòmica CSIC, UPF and CEPR

1.[2] Related to the bias is the well known fact that the short sample distribution of OLS is asymmetric, so that the standard methods for constructing confidence intervals based on symmetry are likely to be inaccurate. Most applied work on time series econometrics relies on asymptotics in order to justify using OLS and traditional (symmetric) confidence intervals, but this is just a way of ignoring the finite sample distribution.

The bias is substantial in VAR applications. The impulse-response coefficients are proportional to the root raised to a power, so that a small deviation in the VAR coefficients is magnified in the IRF coefficients at medium lags. The short-sample bias is likely to underestimate the size of the impulse responses and to diminish the actual significance of the results obtained.

This suggests that the OLS estimated root should be adjusted upwards to compensate for this bias. Some procedures have been designed for this purpose from a classical point of view.[3] These procedures have been shown to reduce the MSE (in short samples) in some cases.

From a Bayesian point of view, however, using OLS and symmetric confidence intervals is easy to justify. The posterior distribution of the VAR coefficients under a *flat prior* is symmetric, centered at the OLS estimate. This implies that OLS is the optimal point estimate with a mean square error loss function. Many VAR applications nowadays are proceeding in this way.

Sims and Uhlig (1991) show that it is possible for the posterior distribution of parameters to be symmetric around OLS in spite of the skewness of the short sample distribution of the OLS estimator, and they conclude that using OLS is justified despite the short sample bias. Furthermore, these authors argue that the use of Bayesian estimators is often much easier than the use of classical finite sample corrections, so that they propose analyzing VAR from a Bayesian point of view even if the researcher does not want to use any a-priori knowledge and, therefore, wishes to use a flat prior.

But we find this conclusion unsatisfactory. It appears there is a dichotomy between the Bayesian and classical approaches in time series. Embracing Bayesianism (and a flat prior) justifies using OLS, but most researchers would like to know why is it that OLS suddenly becomes a good estimator when you change your point of view. The short sample bias still is there and, therefore, it seems as if the (highest root of) OLS should be adjusted upwards.

In this paper we attempt to resolve these issues. First, we argue that

[2]The bias of OLS has been known and studied from a classical point of view since at least Quenouille (1949), Hurwicz (1950), Marriott and Pope (1954) and Kendall (1954).

[3]Among others: Quenouille (1949) proposed a jackknife estimator, Orcutt and Winokur (1969) and Roy and Fuller (2001) use analytical expressions of the bias, Andrews (1993) numerically obtains a median-unbiased estimator and MacKinnon and Smith (1998) correct the bias with bootstrap.

the Bayesian and classical approaches are easily reconciled. We show that imposing a prior distribution that growth rates are unlikely to be very high (we call this a *delta* prior) implies that the Bayesian posterior is asymmetric and it calls for an upward adjustment of OLS. We argue that the delta prior is not controversial to most researchers for many series of interest where the short sample bias would be an issue, so that it should be widely acceptable. We show how to compute the posterior exactly and with a very practical shortcut. As a result, using the delta prior can be understood as a way of correcting for the short sample bias in a practical way.[4]

To make these points we first argue that the optimal properties of OLS hinge on the fact that the flat prior gives a very large weight to parameter values that imply huge growth rates at the beginning of the sample. Most economists will agree that the US GNP is unlikely to grow by *more* than, say, 100% in a quarter, but it is precisely the flat prior gives a very high probability to this event that OLS will appear to be a good estimator. We make this point in detail by performing a 'helicopter tour' similar to that of Sims and Uhlig (1991), and showing that the posterior distribution becomes asymmetric as soon as huge growth rates are given low probability. Then, OLS is no longer the estimator that minimizes mean square error to a Bayesian. A Bayesian would adjust OLS in the same direction as a classical econometrician, and the posterior probability intervals are asymmetric as in the distribution of OLS in short samples.

This leads us to propose a prior distribution that gives low probability to crazy initial growth rates. This prior is, we believe, widely acceptable: most economists will agree that US GNP is very unlikely to grow by more than 100% in a quarter. The prior excludes precisely those parameter values that influence heavily the optimality of OLS under the flat prior and it links classical and Bayesian literature.

We show that, in fact, the delta estimator (i.e., the mean of the posterior under the delta prior) has very desirable properties even from a classical point of view, since it has a lower MSE than classical estimators, even those designed to improve short sample properties by correcting for the short sample bias.

We show how to implement this prior on growth rates with a shortcut that gives an approximation to posterior probabilities. Under this shortcut computational costs similar to those involved in computing the posterior in the standard case.

[4]The relation is, actually, very close: classical results consider initial values for the series that are not too far from its mean (or some 'counterpart' for the explosive processes), and such assumption has the effect of restricting the initial growth rate to be "reasonable".

In addition, we also show a more time-consuming strategy to compute posterior probabilities exactly. This is of independent interest, as it shows how to compute posterior probabilites when the prior is over the behavior of the process itself and not on the parameters of the model. We show that the posterior has to be adjusted by 1) enlarging the actual sample with realizations of the model, 2) finding the maximum likelihood estimator for each enlarged sample, 3) averaging out over estimators obtained, giving small posterior probability to enlarged samples with low probability realizations and giving large posterior probability to samples that deliver an estimator close to the one with the original sample.

As an application, we show the impact of bias, classical bias corrections, and the delta prior on the estimates of the macroeconomic effects of monetary shocks. We concentrate on the well known VAR study of the US data, by Christiano et al. (1999). We first show that the effect of the small sample bias is readily visible in their reported results. We show that the estimation reported in CEE understates the effect of monetary shocks on output by almost a factor of two. It also somewhat underestimates the uncertainty about the obtained impulse responses, so that the price puzzle, (which CEE claim to have ruled out) is not completely rejected in our results.

The rest of the paper is organized as follows. Section 2 sets ground for our subsequent arguments by illustrating how initial conditions affect the precision of the OLS estimator in time series. Section 3 extends the Sims and Uhlig (1991) 'helicopter tour' including a constant term and it shows that adding the delta prior both the posterior and the short-sample distribution of OLS are asymmetric. Section 4 presents the delta prior and the computation of the posterior first under a convenient shortcut and second exactly. Section 5 explains the link between the delta prior and the classical time series bias literature, and it relates our prior to other priors for VARs proposed in the literature, including the Minnesota prior and the no-change prior of Sims (1996). Section 6 shows the implications of the prior on growth rates that we propose in Christiano et al. (1999)'s study of the effects of a monetary policy shock on macroeconomic variables in the US. We conclude in section 8.

2.2 Why OLS appear to be so good

We argue that a Bayesian is likely to exaggerate the virtues of OLS if parameter values that imply huge initial growth rates are given sufficiently high prior probability.

Precision of the OLS estimate depends on the variance of explanatory variables in the analyzed sample. In terms of undergraduate econometrics, the larger the dispersion of the explanatory variable, the smaller the term $(X'X)^{-1}$, and the lower the variance of OLS.

The dispersion, in the time series context, depends on the initial conditions. To illustrate this point, consider the AR(1) model with the intercept:

$$y_t = \alpha + \rho y_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathrm{N}(0, \sigma_\epsilon^2), \quad y_0 \text{ given} \tag{2.1}$$

ϵ i.i.d. For the present example, we take: $y_0 = 0$, $\rho = 0.95$, $\sigma = 1$ and the number of observations, $T = 50$. The process has a stationary distribution with mean $\mu = \alpha/(1-\rho)$. Figure 2.1 displays two realizations of the process y. The first column of graphs is a scatterplot of the above regression, each dot represents the independent variable (y_{t-1}) against the dependent variable (y_t). The next column plots y_t, and the third column plots Δy_t, against time. The two panels only differ in the constant term α. Panel A presents a process with $\alpha = 0$, while panel B presents a process with $\alpha = 2$, which implies that it starts 40 standard errors away from the mean ($\mu = 40$). The random errors ϵ are the same in both panels. The solid line in the scatter-plots is the actual regression line implied by the true parameters in (2.1), the dashed line is the fitted regression estimated by OLS for this realization.

In the upper panel, the series fluctuates about its long run mean and the fitted regression line is significantly flatter than the truth. In this panel the realization contains small growth rates. Note that the slope of the dashed line in the scatter-plot is the value of the estimated OLS coefficient, and that it is lower than the actual regression line, reflecting the fact that the OLS bias is quite large in that realization. In panel B, the transition from the remote starting value to the steady state, which dominates the first half of the sample, results in much higher dispersion of the values of the explanatory variable y_{t-1}. With this high dispersion of the explanatory variable OLS estimates the parameters much more precisely. This is why the fitted regression in panel B is much closer to the truth. Notice from the plots of the first difference of the series that y grows unusualy fast in the first periods, compared to the long-run growth rate.

It is well known that OLS is the best estimator (with a mean square error loss function) under a flat prior. But a flat prior on the constant term (given y_0) gives much more weight to parameter values that imply a mean far away from the starting point, as in panel B, than to parameter values that imply a mean close to the initial condition, as in panel A. OLS is very precise for parameter values as panel B, and no wonder that the Bayesian reasoning suggests it is optimal on average, since a flat prior gives a very large probability to realizations with very high initial growth rates.

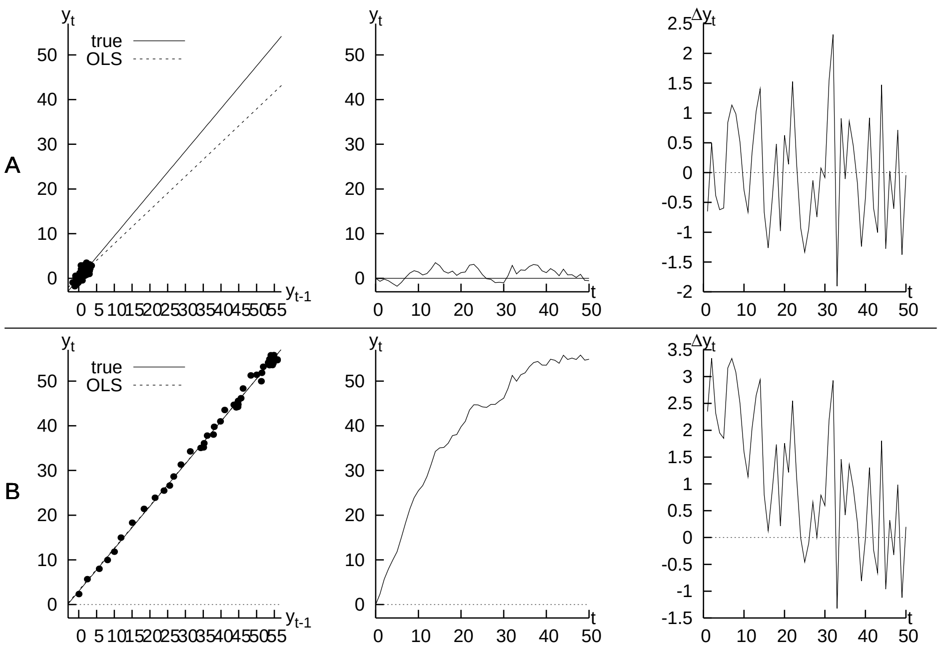

Figure 2.1: Two cases of the AR(1) process and the performance of the OLS estimator of the coefficients. The first column shows scatter plots of y_t against y_{t-1} and the true and fitted regression lines. The second column plots the same y_t against time. The third column plots first difference of y_t against time.

We will argue that if we give low prior probability to cases like B, then the small sample bias of OLS becomes apparent also to a Bayesian, and the Bayesian posterior delivers a bias correction similar to the one in the classical literature.

2.3 The helicopter tour with a constant term

2.3.1 Model without constant term, with flat prior

For completeness we review here the results in Sims and Uhlig (1991) (SU). In the AR(1) model *without constant term*:

$$y_t = \rho y_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \mathrm{N}(0, \sigma^2) \tag{2.2}$$

SU study the joint (bivariate) distribution of ρ and $\hat{\rho}$. They point out that the result will be symmetric around a peak $\hat{\rho}$.[5] They find numerically this joint distribution and display different cuts of it in three-dimensional graphs.[6]

The cross-sections along the fixed-ρ lines, as in Figure 2.2, represent the small sample distribution of $\hat{\rho}$ given ρ. The cross-sections along the fixed-$\hat{\rho}$ lines as in Figure 2.3 represent the distribution of ρ given a value for the estimate $\hat{\rho}$; these cross sections are a summary of the posterior distribution of ρ given $\hat{\rho}$. These cuts reflect two well known facts: *i)* the short sample distribution of OLS under the classical approach (Figure 2.2) shows a downward bias of OLS as well as an asymmetric distribution, *ii)* the posterior under a flat prior represented by Figure 2.3 is symmetric and centered at the OLS estimator.

The difference in the two cuts justifies the dichotomy between the classical and Bayesian approach in time series models that SU advocated. If the researcher does not want to claim knowledge of the parameters, to the extent that the Bayesian estimator and posterior probabilities are optimally derived from Bayes' rule, the relevant cut is the one in Figure 2.3, and this justifies, among other things, using plain OLS as a point estimate. A classical econometrician, however, would concentrate on Figure 2.2 and try to find a bias correction.

2.3.2 Model with constant term and flat prior

Key to our analysis of initial growth rates is the fact that the constant term is not known. Therefore, we first show that the results of SU's helicopter tour extend to the case when the constant term in equation (2.1) is unknown and it needs to be estimated as well. In fact, the dichotomy between classical and Bayesian is deepened.

[5]They point out that under the flat prior for ρ, its posterior (assuming known σ), is gaussian:

$$p(\rho|y) = N\left(\hat{\rho}, \frac{\sigma^2}{\sum_{t=1}^{T} y_{t-1}^2}\right) \tag{2.3}$$

where $\hat{\rho}$ is the OLS estimate. The variance of this distribution depends on the particular realized vector y. To obtain the joint p.d.f. of ρ and $\hat{\rho}$ one has to integrate on each instance of y. But given that the posterior is symmetric, integrating y out will not destroy this symmetry (Sims and Uhlig, 1991, p.1593). As explained in Sims (1988), the downward bias of the OLS estimator is exactly compensated by the increasing precision of OLS for higher values of ρ and, as a result, the posterior is symmetric around the OLS estimator.

[6]We have reproduced the numerical calculations of SU. We take a grid of values for ρ and, for each ρ, generate 10000 data vectors of length T=100 from model (2.2), starting from $y_0 = 0$. For each data vector compute $\hat{\rho}$. Lining up the histograms of $\hat{\rho}'$s obtain a surface which is the joint p.d.f. of ρ and $\hat{\rho}$ under a flat prior for ρ.

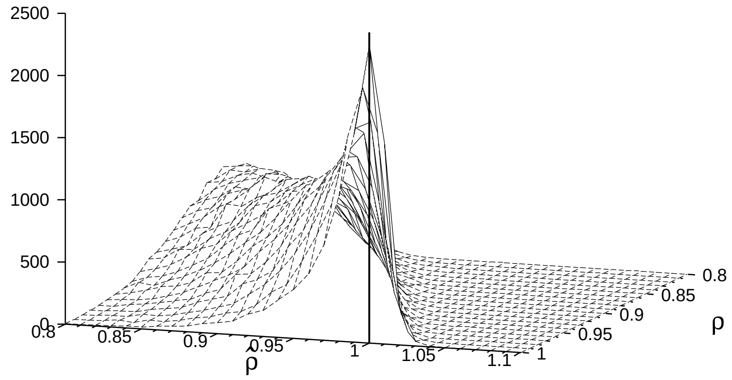

Figure 2.2: Figure 3 from Sims and Uhlig (1991): Joint frequency distribution of ρ and $\hat{\rho}$ sliced along $\rho = 1$

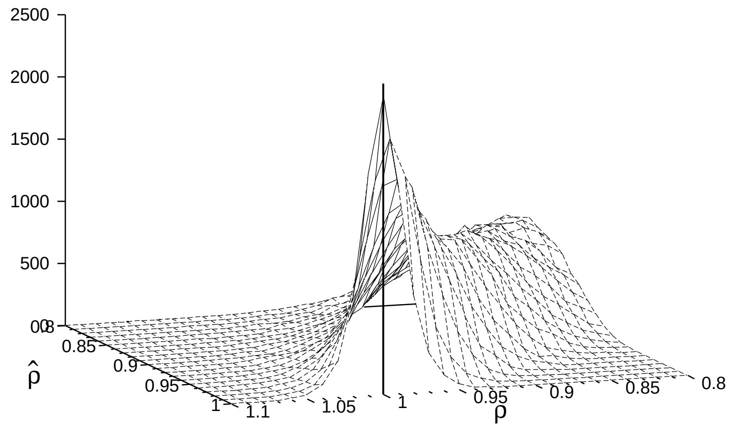

Figure 2.3: Figure 4 from Sims and Uhlig (1991): Joint frequency distribution of ρ and $\hat{\rho}$ sliced along $\hat{\rho} = 1$

It has been pointed out from a classical point of view (for example, Andrews (1993, p.158)) that the small sample bias of OLS is much stronger when α has to be estimated. Therefore, it seems that from a classical point of view a correction for the bias is needed even more urgently in this case.

But a Bayesian is unimpressed by these classical results: it turns out that, under a flat prior, the posterior distribution in the model with a constant is also symmetric around OLS and, therefore, OLS remains the best estimator even if with the larger bias.[7]

The joint bivariate distribution of $(\rho, \hat{\rho})$ is displayed in the next two figures.[8] The cut along a given estimate is depicted in Figure 2.4 (which is the analog to figure 2.3). The distribution of $\rho \mid \hat{\rho} = .95$ is indeed symmetric, the main change is that it is now very tightly concentrated along the $\rho = \hat{\rho}$ line.

We do not show the cut for a fixed ρ to save space. The result is very similar to Figure 2.4.

This figure illustrates our claim that the dichotomy emphasized by SU is even stronger once the constant is estimated. Classical econometricians have emphasised that the small sample bias in a model with the constant is much stronger (Andrews, 1993). Therefore, upon observing $\hat{\rho}$ close to 1, a classical econometrician would recognize that this estimate is a product of a large downward bias and believe that the true ρ is likely to be considerably greater. If this econometrician is concerned about short sample issues he would try to use a bias correction estimator. But, as follows from Figure 2.4, a Bayesian with a flat prior is even more convinced that the value of OLS is a good estimator of the true parameter when a constant is introduced. What is going on?

The simulation underlying Figure 2.4 makes it clear that the reason that the Bayesian views OLS as such a good estimator is that for stationary values

[7]The reasoning about symmetry of the posterior $\rho|\hat{\rho}$ is similar as before. Now we have to consider the bivariate normal posterior for (ρ, α). We need to integrate out the α, but the resulting marginal posterior of ρ is:

$$p(\rho|y) = N\left(\hat{\rho}, \frac{\sigma^2}{\sum_{t=1}^{T}(y_{t-1} - \bar{y})^2}\right) \tag{2.4}$$

which is symmetric, and after integration w.r.t. y, the resulting posterior $\rho|\hat{\rho}$ distribution is also symmetric around $\hat{\rho}$.

[8]We just adapt the procedure of Sims and Uhlig (1991). We take the values of ρ on the grid; for convenience we assume a normal prior for $\alpha \sim \mathrm{N}\left(0, \sigma_\alpha^2\right)$ and draw constant terms α from this distribution and take $\sigma_\alpha \to \infty$. For each draw of (ρ, α) we simulate the process given by (2.1) starting from $y_0 = 0$ and construct the bivariate distribution of $(\rho, \hat{\rho})$ following Sims and Uhlig (1991). $\sigma_\alpha = 10$ turns out to be sufficiently big to make the cross-section along $\hat{\rho} = 0.95$ close to a flat prior.

of ρ ($|\rho| < 1$), large draws of the constant term correspond to large means of the process in the long run (given by $\alpha/(1-\rho)$). Since the process always starts at $y_0 = 0$, high probability of large α's means that realizations as in panel B of Figure 1 are very likely, in such cases OLS can be arbitrarily precise.[9]

2.3.3 Introducing a prior on the initial growth rate

The discussion in the previous section leads us to introduce a prior on the growth rates. This gives low probability to observing realizations of the process which start very far away from the mean, like in panel B of figure 1. We will see how the posterior becomes similar to the short-sample distribution.

The growth in period 1 is:

$$\Delta y_1 = \alpha + (\rho - 1)y_0 + \epsilon_1 \tag{2.5}$$

Our prior states that the growth attributable to the dynamics of the model (excluding the variability of ϵ_1) is normally distributed with mean g and variance σ_g:

$$\alpha + (\rho - 1)y_0 \sim \mathrm{N}(g, \sigma_g^2) \tag{2.6}$$

We call this the 'delta' prior.

Most economists will have an easy time specifying 'reasonable' prior mean g and variance σ_g^2 for a given series. It is much easier to answer the question 'what is a reasonable growth rate for output' than to answer 'what is a reasonable value for the roots of a process in a multivariate autoregression'.

A note of caution is needed about translating prior ideas on the growth rate into the delta prior. Notice that the total prior variance of the growth rate is given by

$$\mathrm{Var}_{\mathrm{Pr}}(\Delta y_1) = \sigma_g^2 + \sigma_\epsilon^2 \tag{2.7}$$

In words, the total a-priori uncertainty is given by the uncertainty about true parameters plus the underlying uncertainty of the shocks ϵ. Hence, if an economist thinks that a reasonable prior standard deviation for the growth rate is 30%, this only gives the left side of (2.7), and the variance on the delta prior is then $\sigma_g^2 = (0.3)^2 - \sigma_\epsilon^2$.

Let us now call up a third helicopter and have a tour of the p.d.f. of $\rho, \widehat{\rho}$ under this prior. In the present case, with $y_0 = 0$, the prior on growth rate just means $\alpha \sim \mathrm{N}(g, \sigma_g^2)$, and the prior on growth rate simply translates as

[9]Large draws of α unambiguously cause high growth rates in the beginning of the sample. ρ close to one also implies large mean of the process, but also a slower convergence to it, so its effect on the initial growth rate is ambiguous.

a prior on α. We take $g = 0$ and $\sigma_g = 0.1$, which, if the data were in logs, would rule out with 95% probability initial 'systematic' (not caused by ε_1) growth rates outside the (-20%,20%) range, which is a very loose requirement for most growing macroeconomic time series.

The resulting bivariate distribution of $(\rho, \hat{\rho})$ is depicted in figure 2.5, with a cut to display the distribution of $\rho \mid \hat{\rho} = .95$. Under the delta prior, the posterior is no longer symmetric around OLS: it is skewed towards larger values.[10] There is no longer a dichotomy between Bayesian and classical views: the mean of Bayesian posterior would now adjust upwards the estimate, correcting for the OLS bias in a similar way as a classical econometrician would.

Therefore, if the huge initial growth rates as in panel B of Figure 2.1 are given small prior probability, a Bayesian 'adjusts' OLS in a similar way as a classical econometrician. This is not a coincidence. In most classical studies of the short-sample distribution of OLS the initial value of the variable is forced to be not too far from the mean of the process, so that the classical results have been implicitly using a prior similar to the one we propose. We study this in detail in section 5.

2.4 Estimation with the delta prior

In many practical applications, especially multivariate, introducing a prior on just the growth rate of the first period is not enough to influence the results. The large number of right hand side variables and parameters implies that just restricting the first growth rate does not have much of an impact on the posterior. For this reason, and because it is still (we believe) an acceptable prior to most observers of economic activity, we propose a prior that imposes a distribution on the growth rates in the first T_0 periods, where T_0 is fixed and does not grow with the sample size.

One may be tempted to impose the prior on all the observations in the sample and to set $T_0 = T$. But this is not reasonable, because it implies that the weight of the prior increases as the sample size grows, therefore the tightness of the prior grows with the sample size. Furthermore, imposing the delta prior on a large number of observations basically forces the parameter values to be consistent with a unit root which is, to our taste, too restrictive,

[10] The reader should take notice that, following SU, the values on the ρ axis values are increasing from right to left in the graph. This may be confusing at first, since it is natural to anticipate the *upward* correction to be illustrated by a shift of the probability mass to the *right* on the picture. We adopt this plotting convention for consistency with SU. They use it because it allows to switch smoothly between the two cross-sections of the plot in the 'helicopter tour' fashion, without having to flip one dimension.

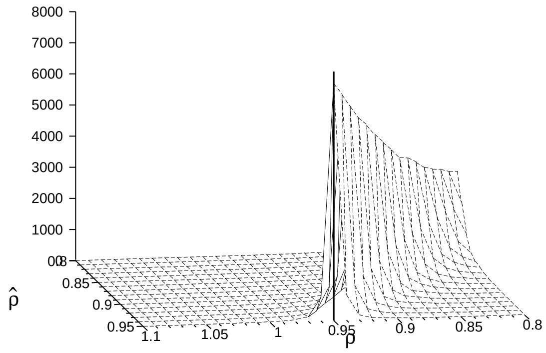

Figure 2.4: Joint frequency distribution of $(\rho, \hat{\rho})$ sliced along the $\hat{\rho} = 0.95$ line (as in Figure 5 from Sims and Uhlig (1991)). Model with a constant term, and an approximately flat prior, starting from $y_0 = 0$

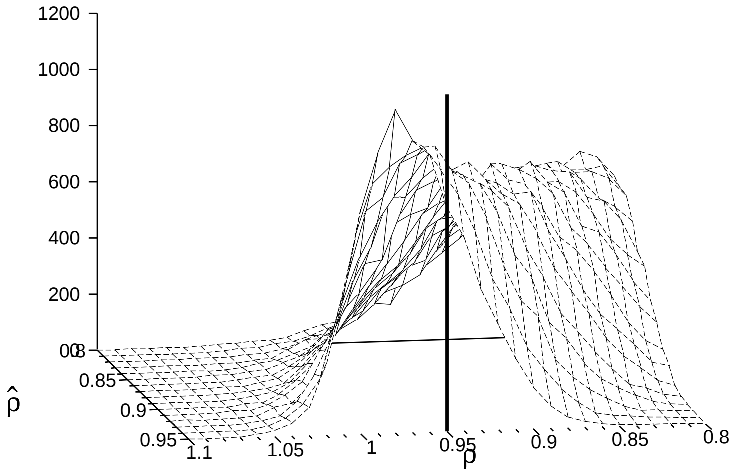

Figure 2.5: Joint frequency distribution of $(\rho, \hat{\rho})$ sliced along the $\hat{\rho} = 0.95$ line (as in Figure 5 from Sims and Uhlig (1991)). Model with a constant term, drawn with $\sigma_g = 0.1$, starting from $y_0 = 0$

and it goes away from the purpose of imposing minimal prior knowledge on the parameters and let the data speak as much as possible. Furthermore, the researcher may be willing to leave open the possibility that the initial growth rates are not similar to the long run growth rates, for example, in order to allow for a transition period due to an initial condition away from the steady state.

As usual, we model the process $\{y_t\}$ as a $J-$dimensional VAR(P) process:

$$y_t = \sum_{i=1}^{P} \Phi_i \; y_{t-i} + \Gamma' w_t + \epsilon_t \quad t \geq 0$$

$\epsilon_t \sim \mathrm{N}(0, \Sigma_\epsilon)$ i.i.d. w_t is a vector of N exogenous variables - typically it would consist only of a 1, and then Γ would be a vector of constant terms. With $T+P$ observations, we define $Y \equiv [y_1, ..., y_T]'$ as a $T \times J$ matrix gathering T observations determined by the model, and the VAR(P) process can be also written as:

$$Y = XB + E \tag{2.8}$$

Here, $X \equiv [Y_{-1}, Y_{-2}, \ldots, Y_{-P}, W]$ collects the lagged values of Y, and exogenous variables W (in the model with just the constant term, W contains only a column of 1's), $B \equiv [\Phi_1, \ldots, \Phi_P, \Gamma']'$ and $E \equiv [\epsilon_1, ..., \epsilon_T]'$. The initial conditions $Y_0 \equiv [y_{-P+1}, ..., y_0]'$ are fixed.

Now, the delta prior states that the parameter values have to be consistent with reasonable growth rates in the first few periods. Say that a researcher is willing to express his/her prior uncertainty about growth rates by stating that

$$\Delta y_t \sim N(g, \Sigma_\Delta) \quad \text{for } t = 1, ..., T_0 \tag{2.9}$$

Here, g is the vector of prior means and Σ_Δ is the vector of *total* prior variance (that is, in terms of our discussion of equation (2.7), this is equal to $\sigma_g^2 + \sigma_\epsilon^2$). Note that, given the initial conditions Y_0, this gives a distribution for the first T_0 observations of the level of the series y.

The presumption is that most economists can easily find reasonable values for g, and even if they choose large values for the variance Σ_Δ, the prior will be sufficient to perform a bias-adjustment that, as we will argue in section 5, often performs even better than classical estimators.

2.4.1 The delta posterior

To characterize the posterior, note as usual that the likelihood conditional on initial P observations is proportional to:

$$L(Y; B, Y_0) \propto |\Sigma_\epsilon|^{-\frac{T}{2}} \exp\left(-\frac{1}{2} \operatorname{tr} \Sigma_\epsilon^{-1} E'E\right)$$

It turns out that the delta posterior is obtained by averaging out normal distributions obtained by stacking the data and realizations of y drawn from the delta prior. Formally, let φ denote the density of $[y_1, ..., y_{T_0}]'$ consistent with the delta prior and the observed initial conditions Y_0; i.e., φ is obtained from $y_t \mid y_{t-1}, ..., y_1 \sim N(g + y_{t-1}, \Sigma_\Delta)$ for $t = 1, ..., T_0$. Let $\overline{Y} \equiv [\overline{y}_1, ..., \overline{y}_{T_0}]'$ denote a particular realization of the first T_0 observations drawn from this density. Let $\widehat{B}(Y, \overline{Y})$ be the OLS estimator of B obtained with the sample of size $T + T_0$ resulting from stacking the actual sample Y and a given realization $\overline{Y}$. Let $\widehat{\Sigma}(Y, \overline{Y})$ be the variance-covariance matrix of the OLS estimator of B in the stacked sample. Assume the variance of the shocks Σ_ϵ is known. In this proposition, assume for simplicity that T_0 is larger than the number of parameters to estimate, formally, $T_0 > K \equiv PJ + N$.

Proposition 1 *The posterior implied by the delta prior satisfies*

$$Post(B; Y) = \frac{1}{\widetilde{\mathcal{K}}(Y)} \int_{R^{T \times J}} \frac{\mathcal{K}(Y, \overline{Y})}{\mathcal{K}(\overline{Y})} \Phi\left(B; \widehat{B}(Y, \overline{Y}), \widehat{\Sigma}(Y, \overline{Y})\right) \varphi\left(\overline{Y}\right) d\overline{Y} \tag{2.10}$$

where

- $\Phi(\cdot; \mu, \Sigma)$ *is the normal density with mean* μ *and variance* Σ
- $\mathcal{K}(\cdot)$ *is the integrating constant that converts the likelihood of the corresponding sample into a pdf that integrates to one. More precisely*

$$\mathcal{K}(Y) \equiv \int L(Y; \overline{B}) d\overline{B} =$$
$$= (2\pi)^{-\frac{(T-K)J}{2}} |\Sigma_\epsilon|^{-\frac{T-K}{2}} |X'X|^{-\frac{J}{2}} \exp\left(-\frac{1}{2} \operatorname{tr}\left(\hat{E}'\hat{E}\Sigma_\epsilon^{-1}\right)\right) \tag{2.11}$$

  (see Appendix C) where $K = PJ + N$ *is the number of parameters in each equation*

- $\widetilde{\mathcal{K}}$ *is an integrating constant, to be computed separately, insuring that (2.10) integrates to one.*

Proof

Throughout the proof and the paper, L is the likelihood function under normal distribution of the errors.

Let f_B be the prior distribution of the parameters implied by the delta prior. To find this prior, first note that the joint prior density of $(B, \overline{y}_1, ..., \overline{y}_{T_0})$ given by the delta prior is given simply by

$$\begin{aligned} p(B, \overline{y}_1, ..., \overline{y}_{T_0}) &= p(B \mid \overline{y}_1, ..., \overline{y}_{T_0})\; p(\overline{y}_1, ..., \overline{y}_{T_0}) \\ &= \frac{L(\overline{Y}; B, Y_0)}{\mathcal{K}(\overline{Y})}\; \varphi(\overline{Y}) \end{aligned}$$

(that is, the 'posterior' of B conditional on $\overline{Y}$, times the marginal of $\overline{Y}$). The first equality is simple probability rules, the second equality uses familiar formulae for posterior of B when no additional knowledge of the parameters is available, and $\mathcal{K}(\overline{Y})$ is the integrating constant

$$\mathcal{K}(\overline{Y}) \equiv \int L(\overline{Y}; \overline{B}, Y_0)\; d\overline{B} \quad .$$

Therefore, the prior satisfies

$$f_B(B) = \int_{R^{T_0 \times J}} \frac{L\left(\overline{Y}; B, Y_0\right)}{\mathcal{K}(\overline{Y})}\; \varphi\left(\overline{Y}\right)\; d\overline{Y} \tag{2.12}$$

The usual formula for the posterior now gives

$$Post(B; Y, Y_0) = \frac{L\left(Y; B, Y_0\right)}{\widetilde{\mathcal{K}}(Y)} \int_{R^{T_0 \times J}} \frac{L\left(\overline{Y}; B, Y_0\right)}{\mathcal{K}(\overline{Y})}\; \varphi\left(\overline{Y}\right)\; d\overline{Y}$$

We have abused notation slightly, since the first function L in this formula has a higher dimensional argument Y than the second L so these are really two different functions and they can only be distinguished by the argument. $\widetilde{\mathcal{K}}$ is the integrating constant of the whole posterior, which we will determine later.

Now, moving the first L inside the integral, multiplying out the exponential terms in the likelihoods and including the appropriate constant of integration we have

$$\begin{aligned} Post(B; Y, Y_0) &= \frac{1}{\widetilde{\mathcal{K}}(Y)} \int_{R^{T_0 \times J}} \frac{L\left(Y; B, Y_0\right) \cdot L\left(\overline{Y}; B, Y_0\right)}{\mathcal{K}(\overline{Y})} \varphi\left(\overline{Y}\right) d\overline{Y} \\ &= \frac{1}{\widetilde{\mathcal{K}}(Y)} \int_{R^{T_0 \times J}} \frac{\mathcal{K}(Y, \overline{Y})}{\mathcal{K}(\overline{Y})} \frac{L\left(Y, \overline{Y}; B, Y_0\right)}{\mathcal{K}(Y, \overline{Y})} \varphi\left(\overline{Y}\right) d\overline{Y} \end{aligned}$$

where, for each realization $\overline{Y}$, the term $L\left(Y, \overline{Y}; \overline{B}, Y_0\right)$ is the likelihood obtained by stacking the actual sample and each realization $\overline{Y}$. The usual argument says that the normalized likelihood $\frac{L\left(Y, \overline{Y}; \overline{B}, Y_0\right)}{K(Y, \overline{Y})}$ is equal to a normal distribution centered at OLS of the sample $\left(Y, \overline{Y}\right)$ and with the OLS estimated variance-covariance matrix . Therefore, $\frac{L(Y,\overline{Y};B,Y_0)}{\mathcal{K}(Y,\overline{Y})} = \Phi\left[B;\widehat{B}(Y,\overline{Y}),\widehat{\Sigma}(Y,\overline{Y})\right]$, and this gives the result.

Clearly, the formula for $\widetilde{\mathcal{K}}$ integrates this posterior:

$$\widetilde{\mathcal{K}}(Y) = \int_{R^{T_0 J+(1+JP)J}} \frac{\mathcal{K}(Y,\overline{Y})}{\mathcal{K}(\overline{Y})}\Phi\left[\overline{B};\widehat{B}(Y,\overline{Y}),\widehat{\Sigma}(Y,\overline{Y})\right]\varphi\left(\overline{Y}\right)d(\overline{Y},\overline{B})$$

□

The formula (2.10) says that the posterior is a weighted average of normal distributions, each distribution found by adding a hypothetical realization $\overline{Y}$ to the actual data. The weight given to each hypothetical realization is $\frac{\mathcal{K}(Y,\overline{Y})}{\mathcal{K}(\overline{Y})}\varphi\left(\overline{Y}\right)$.

This weight is readily interpreted. Obviously, the term $\varphi\left(\overline{Y}\right)$ just gives more weight to a priori more likely realizations. The ratio of integrating constants is related to the comparison of a model which accounts for the whole stacked sample $(Y,\overline{Y})$ with one B, against a model which used separate coefficients for Y and for $\overline{Y}$. To justify this more precisely, note

$$\begin{aligned}\frac{\mathcal{K}(Y,\overline{Y})}{\mathcal{K}(\overline{Y})} &= \mathcal{K}(Y)\frac{\mathcal{K}(Y,\overline{Y})}{\mathcal{K}(Y)\mathcal{K}(\overline{Y})} = \\ &= \frac{\mathcal{K}(Y)}{(2\pi)^{\frac{KJ}{2}}\left|\Sigma_\epsilon\right|^{\frac{K}{2}}}\left(\frac{|X'X|\,|\overline{X}'\overline{X}|}{|X'X+\overline{X}'\overline{X}|}\right)^{\frac{J}{2}} \\ &\times\exp\left(-\frac{1}{2}\operatorname{tr}\left[\left(RSS(Y,\overline{Y})-RSS(Y)-RSS(\overline{Y})\right)\Sigma_\epsilon^{-1}\right]\right)\end{aligned} \tag{2.13}$$

where the first equality is trivial, the second equality uses the formula (2.11) and simple algebra. Here, X are the regressors in the observed sample, $\overline{X}$ the regressors in the hypothetical realization, and RSS is the residual sum of squares for each sample where residuals are computed by OLS applied to each sample. The ratio after the first equality is a Bayes factor comparing a model for the whole stacked sample with one using separate coefficients for Y and for $\overline{Y}$. This ratio re-weighs each realization of $\overline{Y}$ and it gives less weight to realizations such that

1. $\overline{Y}$ behaves very different from Y in the sense that a Chow test of constancy of coefficients would be rejected

2. the regressors $\overline{X}$ are not very informative about individual coefficients

To argue 1), notice that the term in the exponent in (2.13) is proportional to a Chow test, comparing the sum of square residuals of the restricted and unrestricted regressions when the unrestricted allows a whole different set of coefficients for some of the observations.

To argue 2) notice that the drawn regressors $\overline{X}$ are not informative if they are highly collinear, or have little variability. Then the determinant $\left|\overline{X}'\overline{X}\right|$ is close to zero while the other determinants involved in (2.13) need not be zero.

To implement this posterior it is easy to first draw a $\overline{Y}$ from the distribution φ, then draw a parameter from the corresponding normal distribution and to weigh this by the ratio of integrating constants inside the integral. This should allow computation, first of all, of the constant $\widetilde{\mathcal{K}}$, or, when using (2.13), the Bayes factor $\frac{\mathcal{K}(Y)}{\widetilde{\mathcal{K}(Y)}}$ which will be more conveniently scaled. Then, all quantities of interest can be obtained from the weighted sample.

2.4.2 The conditional posterior

We now suggest an approximate way of finding the posterior. The idea is to impose the prior on the first T_0 observations but conditioning each observation t on the *observed* past y's. We now extend the discussion for the one-variable one-lag case, from section 3.

Since the growth rate is given by

$$\Delta y_t = B'(y_{t-1}, ..., y_{t-P}, 1) - y_{t-1} + \epsilon_t$$

(terms in the parenthesis are stacked vertically) we have that this growth rate is influenced by the shock ϵ and the model determined growth, given by $B'(y_{t-1}, ..., y_{t-P}, 1) - y_{t-1}$.

The prior determines the mean and variance for the model determined growth. Note that

$$E_{\Pr}\Delta y_t = E_{\Pr}\left[B'(y_{t-1}, ..., y_{t-P}, 1) - y_{t-1}\right] = g \tag{2.14}$$

$$\begin{aligned} E_{\Pr}[\Delta y_t \Delta y_t'] = \Sigma_\Delta = &\text{var-cov}_{\Pr}[B'(y_{t-1}, ..., y_{t-P}, 1) - y_{t-1}] \\ &+ \Sigma_\epsilon \qquad\qquad\qquad t = 1, ..., T_0 \end{aligned}$$

therefore, the 'model determined' growth rates in the first T_0 observations are normally distributed with means g and variance

$$\text{var-cov}[B'(y_{t-1}, ..., y_{t-P}, 1) - y_{t-1}] \equiv \Sigma_g = \Sigma_\Delta - \Sigma_\epsilon \tag{2.15}$$

and with the autocorrelation matrix S which is the same for all variables.

Normally we would have to integrate past $y's$ out of the above mean and variance, because the delta prior has implications for all y's within dates 1 to T_0, but the shortcut we propose is to use the above formula by fixing past y's that appear in (2.14) and (2.15) and setting them equal to the observed ones in the data. This gives

$$\mathrm{vec}\,(X_0 B - Y_{-1}) \sim \mathrm{N}(g \otimes \iota_{T_0}, \Sigma_g \otimes S) \tag{2.16}$$

(where ι_{T_0} denotes a column vector 1's, of length T_0) or

$$p(\mathrm{vec}\,B|X_0) \propto \\ \exp\left(-\frac{1}{2}\left(\mathrm{vec}\,B'(\Sigma_g^{-1} \otimes X_0' S^{-1} X_0)\,\mathrm{vec}\,B - 2\,\mathrm{vec}\,B'\,\mathrm{vec}\left(X_0' S^{-1}(Y_{-1} + \iota_{T_0} g')\Sigma_g^{-1}\right)\right)\right) \tag{2.17}$$

which is a normal prior, conjugate with the likelihood. If we use the standard RATS prior for the error variance:

$$p(\Sigma_\epsilon) \propto |\Sigma_\epsilon|^{-\frac{v+J+1}{2}} \tag{2.18}$$

where $v = PJ + K$, i.e. the number of variables in each equation, the posterior for B and Σ can be easily simulated with the Gibbs sampler. We will refer to this prior as the conditional delta-prior.

This prior is only 100% consistent with the delta prior in the case that $T_0 = 1$, because in that case the values of y in (2.14) and (2.15) are given in the initial conditions Y_0. But if the prior is imposed on more observations (i.e., if $T_0 > 1$) past values that appear in the mean and variance are unknown a priori, they should be agreeable with the distribution implied by the prior and not set equal to the observed data. But this delta prior conditional on actual data is easy to use, it involves similar computations as the RATS prior and it is the one we use in this version of the working paper.

2.5 Delta prior and classical bias corrections

We argue in this section that the delta-prior, although inspired by Bayesian principles, can be also justified from a classical perspective. In the related paper (Jarociński and Marcet, 2005) we discuss at length the relationship between classical bias correction and Bayesian estimators. We argue there, that focusing on the bias is unjustified and that a classical estimator could be constructed that takes care of mean square error. We relate carefully

Bayesian estimation and classical bias corrections by showing a series of estimators designed to behave well in short samples; each of these estimators is related to its predecessor, and the first estimator is close to bias correction while the last is close to Bayesian estimator. In this way, Bayesian and classical estimation can be regarded as being quite close.

In this section we focus on the properties of the delta prior relative to bias correcting estimators. The section is quite involved, so here we summarize the results. First we argue that studies of classical bias correction are related to the delta prior in two ways: these studies have specified certain assumptions for the distribution of the initial variable (initial conditions). Even though this was done for technical reasons (in order to suppress dependence on nuisance parameters), the assumptions for initial conditions that have been used turn out to avoid huge growth rates in the first few periods. In this sense, the literature on classical bias correction has imposed a requirement on initial conditions that works similarly to the delta prior. The second point of contact is that the posterior under the delta prior adjusts the mean of the estimator in the same qualitative direction as bias correction estimators and, in this sense, the delta prior could be described as providing a 'Bayesian bias correction'.

But the bias corrections have two problems: first they actually give biased estimators, second, there is no reason that removing the bias is a good objective per se. It is well known that an unbiased estimator may have a very large mean square deviation. We show that the delta estimator is substantially better in terms of MSE than classical even when, in a classical spirit, we consider many fixed parameter values. The lower mean square error of the delta estimator obtains for a wide range of parameter values that include the parameters of interest in a model that grows, even when the parameter is fixed. We conclude that the delta-prior can be motivated on classical grounds.

2.5.1 Modelling initial conditions

This subsection discusses various ways to model the initial condition that have been used in the literature. Although it is a fairly long digression, we need it in order to set common ground for the discussion of classical bias corrections and alternative Bayesian priors proposed in the literature.

The first step is to reparameterize the model as:

$$y_t - \mu = \rho \left(y_{t-1} - \mu \right) + \epsilon_t \qquad \text{for } t = 1 \ldots T \tag{2.19}$$

This parametrization is a special case of (2.1) for:[11]

$$\alpha = \mu(1 - \rho) \tag{2.20}$$

The specified distribution for y_0 will be related to μ. Under (2.19) the process reverts to μ if $|\rho| < 1$, goes away from μ if $|\rho| > 1$, and it drops out when $\rho = 1$.[12]

Classical estimators need to specify fully the distribution of the process in order to derive the short sample properties of the estimator. One could assume, for example, $y_0 = \mu$, but this would be very restrictive since, in practice, there is no reason that the first period in the sample is exactly at the centering parameter μ. Usually a distribution of the initial variable y_0 is specified. Since the focus is on estimating ρ, it is convenient to specify a distribution that causes the OLS estimator to be independent from the 'nuisance' parameters of the model: μ and σ. In that way, the results on estimators of ρ are valid for a wide range of nuisance parameters. A general condition for this independence is

Result 2 Assume the initial condition in model (2.19) is given by:

$$y_0 = \mu + \sigma\psi \tag{2.21}$$

where ψ is a random variable. Then, if ψ independent of the shocks ϵ and its distribution is independent of μ and σ, the distribution of the OLS estimator of ρ in (2.1) is independent of μ and σ.[13]

[11] One can not go the other way around: if $\rho = 1$ in (2.1) there is no μ that satisfies the above equation and generates the same y's. This will have implications for the practical application of this parametrization when the observed variable has a positive growth rate and is close to a unit root, because a trend will have to be introduced in this case to have a growing variable.

[12] Zivot (1994) explains that the motivation for suppressing the constant term α when $\rho = 1$ is that it changes interpretation in the unit root case, from determining the level of the process to determining its drift. When no trend is included for the general ρ, it is not reasonable to allow for a drift in just one point of the parameter space: $\rho = 1$.

[13] Similar results have been used in the literature. As can be seen, the proof is very simple, but we could not find a formal proof, so we offer it here for completeness. The proposition is very similar to the property of $\hat{\rho}$ discussed in Andrews (1993, Appendix A), which contains a verbal proof for $|\rho| \leq 1$ and a particular distribution for ψ, but allowing for a trend.

Proof: Define normalized errors: $u \equiv \epsilon/\sigma$. (2.21) allows to write:

$$y_t = \mu + \sigma\left(\sum_{i=1}^{t} \rho^{t-i} u_i + \rho^t \psi\right) = \mu + \sigma\tilde{y}_t$$

All classical papers that we know of use a version of (2.21). Of course, each choice of ψ will result in a different small sample distribution of $\hat{\rho}$ conditional on ρ. Let us now review various assumptions on ψ.

The case $y_0 = \mu$ discussed above corresponds to $\psi = 0$.

Bhargava (1986) constructs invariant (to μ and σ) tests for unit roots taking

$$\begin{aligned} \psi &\sim \mathrm{N}\left(0, \frac{1}{1-\rho^2}\right) && \text{when } |\rho| < 1 \\ \psi &\sim \mathrm{N}(0,1) && \text{otherwise} \end{aligned} \tag{2.22}$$

so that the independence of the distribution of ψ on the nuisance parameters required in the proposition holds. This yields $y_0 \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{1-\rho^2}\right)$ in the stationary case, so it is assumed that the process is in the stationary distribution (as if it had been running forever). It yields $y_0 \sim \mathrm{N}(\mu, \sigma^2)$ in the non-stationary case, that is, it says that y_0 was generated from the model and the initial condition was brought back one period to set $y_{-1} = \mu$.

Andrews (1993) considers median unbiased estimators for ρ by taking a similar starting point for the stationary case and an arbitrary starting point (i.e. ψ is an arbitrary constant) for the unit root case (he does not consider explosive values of ρ).

MacKinnon and Smith (1998) take:

$$y_0 \sim \mathrm{N}\left(\mu, \sigma^2\right) \tag{2.23}$$

so that they take $\psi \sim \mathrm{N}(0,1)$ for all ρ.

Uhlig (1994) proposes a formulation which encompasses many cases: he assumes $y_{-S} = \mu$, where S is a given constant that the researcher has to specify. This provides an alternative ψ.

To provide a common nomenclature for all these cases, we call Uhlig's proposal the 'S' model. Then the case $y_0 = 0$ becomes the '$S = 0$' model, the assumption of MacKinnon and Smith (1998) as in (2.23) becomes the '$S = 1$' model and we call the assumption of Bhargava (1986) in (2.22) the '$S = \infty$' model.

---

where $\tilde{y}$ is the process with the zero mean, which would obtain from the same realization of errors, but rescaled to have a unit variance. Then it is a matter of simple algebra to show that:

$$\hat{\rho} \equiv \frac{T \sum y_t y_{t-1} - \sum y_{t-1} \sum y_t}{T \sum y_{t-1}^2 - \left(\sum y_{t-1}\right)^2} = \frac{T \sum \tilde{y}_t \tilde{y}_{t-1} - \sum \tilde{y}_{t-1} \sum \tilde{y}_t}{T \sum \tilde{y}_{t-1}^2 - \left(\sum \tilde{y}_{t-1}\right)^2}$$

□

We point out here that there are many ways of choosing the distribution of initial variables. It would seem that these initial conditions could have very different implications, and so that MacKinnon and Smith (1998) and Andrews (1993) may be correcting completely different biases.

2.5.2 Bayesian Bias Correction

We can now compare the delta-prior to classical bias correction. A Bayesian posterior mean is guaranteed to have the lowest MSE on average over the parameter space given by the prior. But from a classical point of view this is of little interest, the question is how an estimator performs for a given value of the parameters. Let us dub the mean of the delta posterior as the *delta estimator*. We show the Bias and MSE properties of delta estimator compared to alternative estimators from a classical point of view, for given parameter values.

The bias of various estimators is examined in a Monte Carlo study in which, in order to highlight the small sample problems, we assume $T = 25$. We consider the AR(1) model for various ρ's. We model the initial condition with $S = 1$.[14] We consider delta prior with mean 0 and two standard deviations: 0.2 (denoted delta1) and 0.05 (denoted delta2).

The performance of the delta estimator is compared with that of OLS and a bootstrap-bias-corrected (CBC) estimator of MacKinnon and Smith (1998) (that is, the bias correction is derived with the actual model $S = 1$) in a Monte Carlo experiment. Kilian (1998) essentially uses CBC to improve the small sample properties of inference with VARs. Figure 2.6 shows the bias for each estimator and each possible true value of ρ.[15] Obviously, the largest bias is with OLS. The CBC estimator is only approximately unbiased (see, Jarocinski and Marcet (2005) for a discussion) so the picture reflects that the bias is not zero but, as is well known, CBC reduces the bias considerably relative to OLS. We can see that for high values of ρ the bias of the delta estimator is in between that of OLS and CBC.[16]

Therefore, the Bayesian estimator can also achieve a bias correction. The correction is not as 'good' as the CBC estimator. But it is clear that achiev-

[14] To compute the values in this and next figure we took 100,000 realizations of the process and values $\rho = 0, 0.05, \ldots, 1.2$.

[15] We only report positive values of ρ. Since we are concerned with possibly non-stationary series we can ignore negative values of ρ. These would imply oscillations in the series, so they imply large changes from one period to the next and, clearly, the delta prior does not work very well when this is the truth.

[16] We have done similar computations when the CBC adjustment is computed with $S = 0$ and $S = \infty$. The general picture does not change.

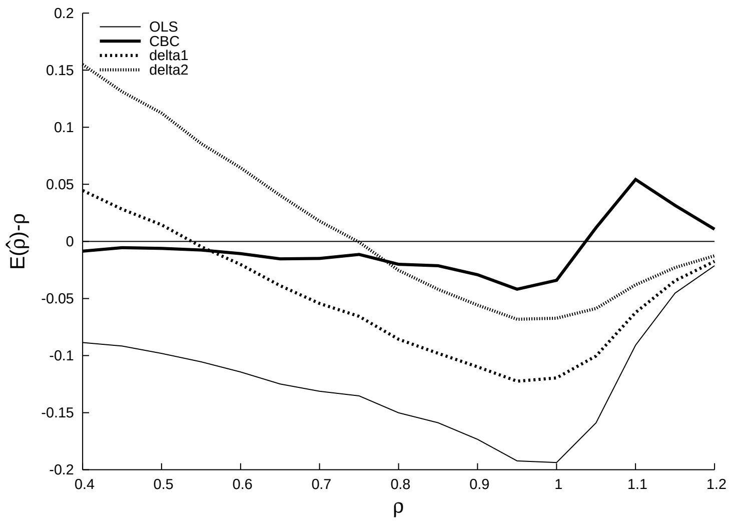

Figure 2.6: Bias of the OLS, CBC and two delta estimators in a Monte Carlo experiment, sample size: T=25.

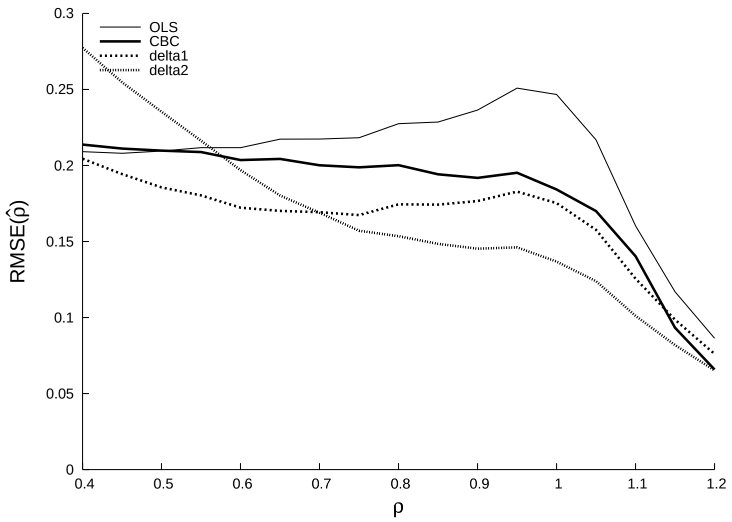

Figure 2.7: RMSE of the OLS, CBC and two delta estimators in a Monte Carlo experiment, sample size: T=25.

ing (or getting close to) unbiasedness is not an important goal per se. A researcher cares about obtaining estimates that are on average close to the true value, so what is relevant is the total mean square deviation from the true parameter and, as is well known, unbiased estimators can have large MSE. We now turn to studying the MSE but still remain in the classical setup, and study the MSE for fixed values of ρ.

Figure 2.7 is a more relevant measure of how big deviations are, as it reports the root MSE for the three estimators under consideration at various values of ρ. The figure shows that Bayesian estimators beats OLS and the CBC for sufficiently high values of ρ. This surely contains the relevant roots in practice for possibly non-stationary series. Actually, the delta estimator can be substantially better, for example, at $\rho = 1$ the RMSE under CBC is around 30% larger than with the delta2 estimator. This increase is very large, close to the one obtained if half of the sample is thrown away.

Notice that, in this case, all the cards seem to be stacked in favor of the CBC: the bias correction is designed assuming $S = 1$ and this is, in fact, the truth. In the real world this (or any other) assumption on the initial condition is quite unwarranted, so we would expect CBC to be at an even larger disadvantage in practice. We have computed all possible combinations of $S = 0, 1$ and ∞ for both the actual model and the bias correction and we obtain similar results.

Our conclusion is that, as long as a researcher is willing to say that the true ρ stays between, say, .7 and 1.1, the delta prior is a better alternative than available bias corrections, even from a classical point of view.

2.6 Other Bayesian priors

Our paper contributes to the literature on priors for the VARs. We have discussed at length the advantages of the delta prior over the flat prior. Several alternatives to the flat prior in time series have been proposed and they usually also compensate the small sample bias. We start by commenting differences with the Minnesota, Jeffrey's and shrinkage priors. More related to our work will be the discussion of Sims (1996) and Sims and Zha (1998) prior as well as the S-prior of Uhlig (1994).

2.6.1 Minnesota, Jeffrey's and shrinkage priors

Applied work often uses the so-called Minnesota prior, with a Normal-Wishart prior centered at the unit root (Doan et al., 1984, see), which pushes the AR estimates towards a unit root. This estimator can be interpreted as delivering a bias correction, since in practice it usually increases the root of the process relative to OLS. But this correction is arbitrary, completely determined by the precision of the prior. It is difficult for an applied researcher to determine what are is his/her own beliefs about the parameters of the AR process and relate these to knowledge about how the economy works. This prior has been shown to be effective for forecasting purposes, but economists using a VAR to summarize the properties of the data generally avoid using this prior: if the empirical results are crucially determined by the precision of the prior, criticizing the prior becomes an obvious way to challenge the empirical results.

Phillips (1991) criticized the flat prior on the grounds that it was not truly uninformative. He suggested to use Jeffrey's prior which, he argued, is truly uninformative and it appropriately gives larger weight to higher roots. By favoring higher roots, the prior implies some bias correction compared with OLS. This sparked an intensive debate, to which we have nothing to add, but we will just note that Jeffrey's prior is not commonly found in practical applications. In part, this is because of computational difficulties that make VAR applications hard. We just feel that most economists will feel more comfortable stating their prior about the growth rate than about parameters of the process. Also, our results show that criticism of the flat prior by Phillips (1991) on the grounds that the flat prior gave same weight to roots close to .5 or to 1 was missing the main point, since the problem are the huge initial growth rates. Our prior still gives the same weight to these alternatives but it does deliver a bias correction, and it reconciles the classical and Bayesian viewpoints.

Another alternative to the flat prior is the 'shrinkage' prior introduced for VARs by Ni and Sun (2003), which intends to curb overfitting by shrinking the parameters towards zero. As a result, conditionally on starting close to zero, it also constrains the initial growth rates implied by the parameters, so it ends up having a similar effect to our prior. Monte Carlo simulations of the authors suggest that on balance it tends to shrink mostly the constant terms, while somewhat pushing upwards the estimated slope coefficients, and therefore it corrects the small sample bias.

2.6.2 Priors similar to the delta prior

We have found two priors that happen to be close to the delta prior.

Sims (1996) mentioned that, even though a flat prior justified using OLS from a Bayesian prospective, the results were unsatisfactory and OLS gave a systematic overestimation of the error in the first few observations. This is, in a way, observing the small sample bias and saying it should be corrected.

Sims proposed to introduce a 'dummy initial observation' with no change in the variable. This is discussed in more detail in Sims and Zha (1998). This prior reflects the 'belief that no-change forecasts should be 'good' at the beginning of the sample' (Sims, 1996, p.5). It is, similarly as the Minnesota prior, an experience-based prior, introduced by adding a dummy observation with a variance. This prior is a special case of the delta prior when the prior mean is $g = 0$ and, somehow, the first P observations are kept constant (perhaps by assuming that $\sigma_g = 0$ for $T_0 - 1$ periods and setting $\sigma_g = \sigma_\varepsilon$ in the period T_0). Looking at the expression in our Result 1, it is clear that the delta posterior is consistent with a φ that places probability 1 on such an evolution of y. In this sense, the prior of Sims (1996) is a special case of the delta prior. We feel that the delta prior is a more explicit way to express the uncertainty about the data, that it has a clearer interpretation and it is easier for an applied researcher to relate to it.

Also close to the delta prior is the work of Uhlig (1994). It has been stated often that conditioning on the first few observations is incorrect, because, especially in a small sample, these initial observations carry valuable information about the parameters. Uhlig (1994) was an attempt at using the exact likelihood, putting the distribution of initial conditions in the likelihood.[17] But, as should be clear from our discussion of modelling initial conditions above, there is not a clear way how to do this. Uhlig proposed to link parameters and initial conditions by saying how many periods the model has been operating (Uhlig's S) *and* by stating what was the initial condition at the beginning of these periods. One alternative is to pick $S = \infty$ but this is as arbitrary as any other assumption: why not 10 periods?, why not since the end of WWII? If S is fixed, to what number? why assume the variable was at the centering parameter μ at the beginning? Furthermore, for roots equal or larger than 1 we have to assume a finite number of periods (usually one) and an arbitrary initial condition at that time (usually $y_{-1} = \mu$). It is probably because of this reason that the S-prior has not been used much in applied work.

In the appendix we report our Monte Carlo experiments with the S-prior in the univariate case, and we propose how to extend it to a VAR. It turns out, that the Uhlig's S-prior has some of the nice properties of the delta-prior. In the appendix we show that introducing the S-prior delivers a similar asymmetry of the posterior, that $S = 0$ has very good properties in a classical sense, and that in an applied case with $S = 1$ it provides similar empirical results. We feel that these results reinforce the theme of the paper, namely, that if huge growth rates are discarded in the prior then a Bayesian

[17] We call his approach a prior, but he stated it as writing the exact likelihood.

observes an asymmetric posterior and much more precision can be gained from imposing such priors. It is because the S-prior also reduces the growth in the first few observations, and it makes transitions as in panel B of figure 1 very unlikely, that it has some features in common with the delta prior. However, we feel that the delta prior is easier to use for applied work.

2.7 Christiano et al. (1999) estimation of the effect of monetary shocks in the US

To study the effect of bias and bias-correcting priors on macroeconomic VARs we first replicate the estimation of the effects of the monetary shock in the US from Christiano et al. (1999)[18], referred to further as CEE99. The authors estimate a VAR with output (Y, measured by the log of real GDP), prices (P, the log of the implicit GDP deflator), commodity prices (PCOM, the smoothed change in an index of sensitive commodity prices), federal funds rate (FF), total reserves (TR, in logs), nonborrowed reserves (NBR, in logs) and money (M1 or M2, in logs). All data are quarterly and the sample is 1965:3 - 1995:2. The residuals are orthogonalized with the Choleski decomposition of the variance (with this variable ordering) and the monetary shock is the one corresponding to the federal funds rate.

2.7.1 Impulse responses to monetary shocks

Figure 2.8 displays the responses of output to a monetary shock estimated with three procedures. The solid line is the responses from OLS estimator, equal across columns. The extreme dashed lines contain the 95% confidence bands and the middle line signals the median. The first column reproduces the results in CEE99 using a bootstrap procedure that is an attempt to provide an idea about uncertainty about the point estimates of impulse responses, while disregarding the small sample bias ('other-percentile' bootstrap). The second column is arguably the most popular way of constructing confidence bands, using the RATS procedure (Doan, 1992) for generating the Bayesian posterior bands for impulse responses assuming a flat prior. It is, therefore, centered at OLS and symmetric. The third column is the delta prior, conditional on the initial observations.

The bootstrap confidence bands for some series are displaced towards zero, compared with the point estimate, reflecting the small sample bias in

[18]We choose this example because the data is on the Internet, and because the authors use the bootstrap error bounds, which highlight the small sample bias.

time series. (The sample is quite large - 120 observations - so the bias may not be striking at first glance, but in smaller samples the bootstrap confidence bounds often exclude point estimates. Also, it is possible that this application has so many parameters that the sampling error is very large and the bias is, by comparison, small potatoes. We are implementing these comparisons for some other applications where the bias may be larger.)

But in any case it is clear that the bias causes an underestimation of the effects of a monetary shock on output. The bias seems to be particularly large in the response of output and it is there also for the response of PCOM.

The next step is to impose the prior on initial growth rates of the variables. The parameters of the basic specification of the prior are given in table 2.1. We have chosen these values mechanically by setting the means equal simply to the average growth rates observed in the whole sample, and standard deviations are half of those observed. A more opinionated reader can plug in his/her own preferred values, but the results are not going to change much. Recall that in order to obtain the total variance one has to add the variance of the innovations, so the prior is quite loose. We set $T_0 = 10$.

Table 2.1: Basic specification of the Δ-prior

variable	prior mean annual growth rate	prior standard deviation
Y	2.7	1.8
P	5.0	1.2
PCOM	3.2	100
FF	0.0	2.4
TR	5.4	4.5
NBR	5.2	3.3
M1	6.5	2.0

In contrast to the non-informative prior, the delta-prior (see the plots in column 3 of figure 2.8) corrects the bias and typically shifts the posterior probability mass of the impulse responses away from zero. Notice that the 'other-percentile' bootstrap bands adjust in the wrong direction: in seeing the bootstrap in the upper left graph, the researcher should acknowledge that the bias reduces the absolute value of the response and, therefore, the researcher should report even more negative responses. But just displaying the bootstrap bands and median goes in the wrong direction. By contrast, the delta prior does the job automatically, and it gives more negative responses than OLS. The effect is strongest exactly where the bias, indicated by the deviation of the median from OLS, was strongest. Consequently, the

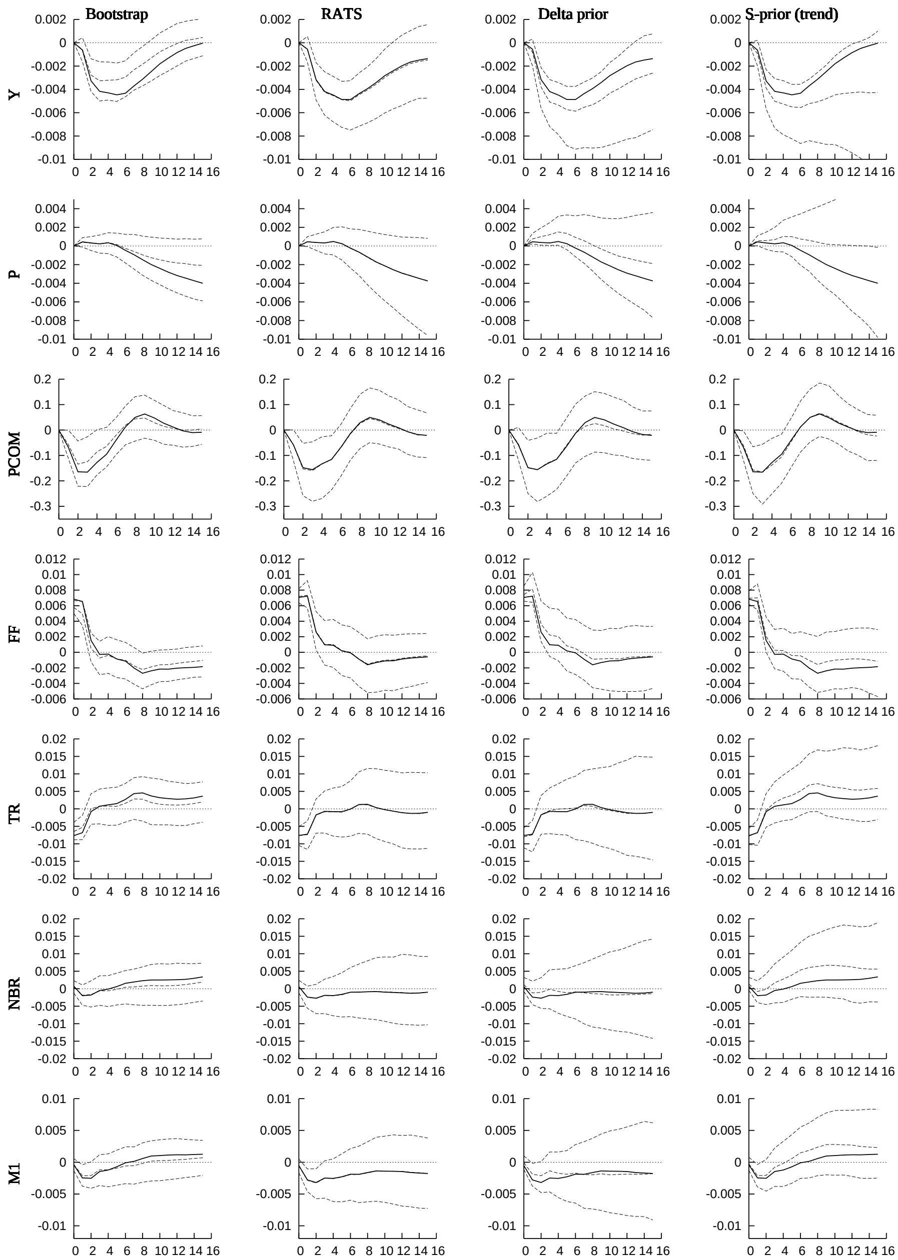

Figure 2.8: Impulse responses to monetary shocks: OLS point estimate, median and the 95% uncertainty bounds generated by 'other-percentile' bootstrap, RATS routine (flat prior), Δ-prior (with parameters specified in the text) and $S = 1$-prior (the latter with trend, see the appendix)

responses of output estimated with the delta-prior are more negative and persistent, and the size of the interest rate shock is much more positive, giving more significance to the results of a VAR.

Another effect of the Δ-prior is that, tilting the estimates towards non-stationarity, it produces wider uncertainty bands then the non-informative prior, especially for longer lags. We believe that this reflects the fact that (despite the efforts of many researchers) it is virtually impossible, with a sample of 120, to ascertain the long run effect of a shock.[19]

Finally, we report what one would find with the S=1 prior for VAR's, according to our interpretation of Uhlig's prior in the Appendix. In this case, for the estimation to make sense, a trend must be included. The results are very close to the delta-prior. We feel this is reassuring since, as we said in section 6, the S-prior is another way to put very small probability on huge growth rates.

2.7.2 Comparison with Kilian (1998) bootstrap-after-bootstrap

Kilian (1998) proposed a classical procedure to construct error bands for impulse responses, called bootstrap-after-bootstrap. It is intended to improve the coverage of bootstrap intervals by approximately removing small-sample bias at each step of the simulation. Bootstrap is used to approximately remove the bias, and the underlying assumption is that the bias is constant in the neighborhood of the OLS estimate.

As described in Kilian (1998), the first step consists of estimating and correcting the bias of the OLS point estimate. In the second step, the error bands are generated: series are repeatedly simulated from the data generating process implied by the corrected OLS estimate, a VAR is estimated by OLS at each simulated data set, the OLS estimate is corrected for bias analogously as in the first step, and finally impulse responses are computed and stored. In a simplified, but computationally much cheaper version of this algorithm, the bias estimate obtained in the first step is reused in the second step, for correction of all OLS estimates on generated data (instead of performing a separate bootstrap for each of them). We use this simplified bootstrap-after-bootstrap here.

[19] There is no paradox in the fact that the informative prior widens uncertainty bands, since their width depends both on the width *and* the location of the uncertainty bands of the autoregressive coefficients. Consider the response to a unit shock after 16 lags in the AR(1) model: $\rho \in (0.5, 0.7)$ corresponds to the response in the range (0.000015,0.0033), while a twice shorter range $\rho \in (1, 1.1)$ corresponds to responses in the range (1,4.6).

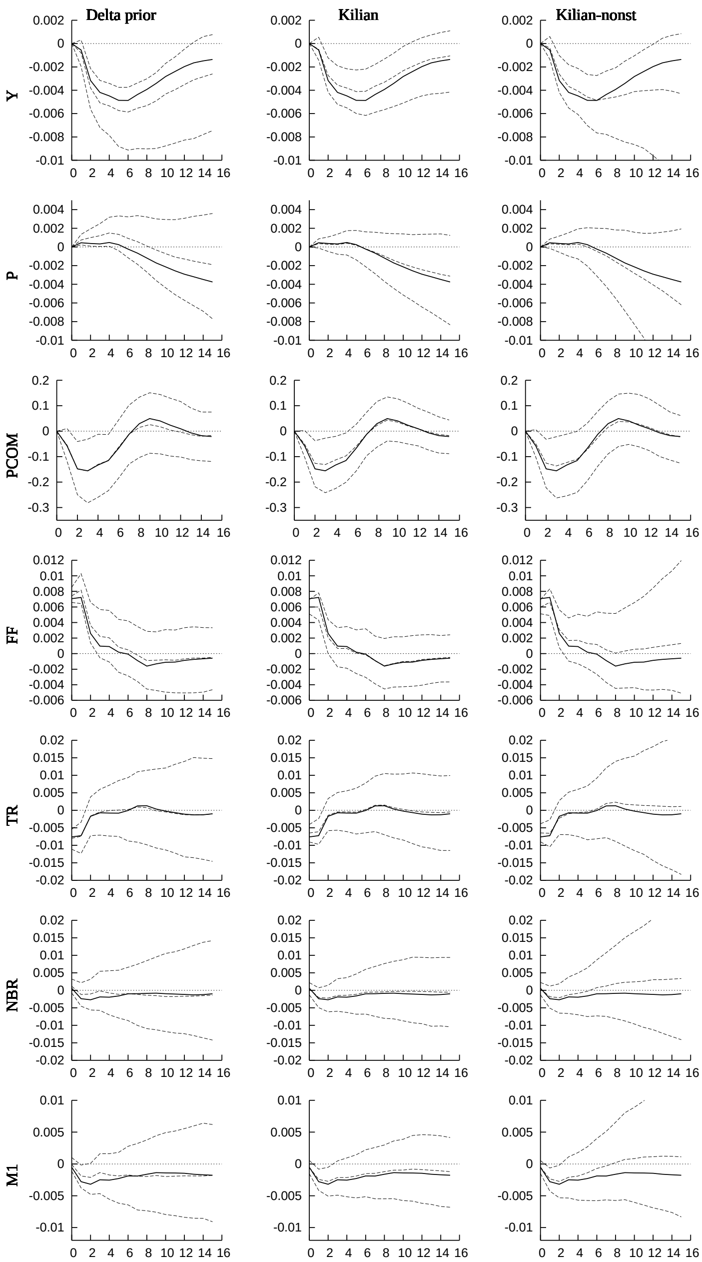

Figure 2.9: CEE99 responses to monetary shocks: OLS estimates and percentiles 2.5, 50 and 97.5 of the distribution of impulse responses. Delta prior, 'Kilian' - bootstrap-after-bootstrap method and 'Kilian-nonst' - bootstrap-after-bootstrap method without correction of nonstationarity.

Bias correction pushes the roots of the process towards the nonstationary range and explosive responses may often result. This may, however, be a spurious effect caused by the assumption of constant bias. Kilian suggests to shrink bias estimates in all cases, when the bias corrected estimate would become explosive, and thus guarantee stationarity of all the simulated distribution of impulse responses. Sims and Zha (1998), when applying Kilian's method, deviate here and allow for explosiveness. We report results from both approaches.

The first column in 2.9, labeled 'delta prior' shows again, for easier comparison, results obtained with delta prior. The second column, labeled 'Kilian', shows results obtained with Kilian's simplified bootstrap-after-bootstrap method, constraining roots of bias-corrected OLS estimates to be stationary. The results are quite similar to the 'naive bootstrap' bands from figure 2.8: the error bands are narrow and in some cases (output and federal funds rate) displaced towards zero, displaying the small sample bias towards stationarity. The third column, labeled 'Kilian-nonst' (for 'nonstationary'), shows the error bands computed with simplified bootstrap-after-bootstrap, but without imposing stationarity. The results are more similar to the results of the delta-prior estimation, where stationarity is not imposed either. The resulting bands reflect it, becoming very wide for higher lags. Overall, the figure illustrates the dilemma involved in applying bootstrap-after-bootstrap, when the root of the system is close to unity: imposing stationarity discards much of the bias correction. Allowing nonstationarity, on the other hand, exposes the results to the inaccuracy of the assumption of constant bias. We guess that it is because of the failure of this assumption in practice, that the bands contain too much nonstationarity, and are so wide for farther lags. It is possible that applying the full, not simplified, version of Kilian's algorithm gives better results.

2.8 Conclusions

One theme of the paper has been that as long as huge growth rates are given small weight there is no dichotomy between Bayesian and classical estimation in time series. Both with the S-prior and our delta prior the posterior looks asymmetric and the same issues that had been discussed in short sample classical analysis arise in Bayesian econometrics. A Bayesian adjusts upwards (towards nonstationarity) the OLS estimator in a similar way as corrections proposed in the classical literature.

Our preferred alternative is the delta prior. It should be uncontroversial that most economic variables can not have huge growth in any given period;

introducing this knowledge about the economy reconciles a Bayesian posterior with the certainty that classical econometricians (at least those that correctly worry about short sample issues) have about the fact that OLS should be adjusted upwards close to a unit root.

But correcting the bias turns out not to be a very good way to estimate parameters and, at the same time, take care of short sample issues. It seems that a Bayesian estimator based on a delta prior is also the best from the classical point of view.

We have illustrated the effect of the delta-prior in a practical case: the estimation of the macroeconomic effects of a monetary policy shock, following Christiano et al. (1999). The posterior corrects the small sample bias towards stationarity, and implies stronger responses, in particular of output and interest rates, than those found by the authors.

Further work is planned on the issues raised by this chapter. We have to implement the exact posterior calculations and explore further implications of the delta prior. In particular, it will be interesting to examine the forecasting performance of the delta-prior compared with the Minnesota prior and other standard forecasting tools. Second, it will be interesting to see the effect of delta-prior in other structural VAR applications, where short sample issues may be important.

Appendix 2.A S-priors for the AR(1)

In this section we explore the implications of the initial conditions, formulated following Uhlig (1994). Uhlig considers processes which had started at the mean S periods before the first observation in the sample.[20]

Let us first consider the case $S = 0$ and call a fourth helicopter to tour a joint p.d.f.. Now, the initial condition become

$$y_0 = \mu \tag{2.24}$$

with probability one, which is analogous to the condition $y_0 = 0$ of Sims and Uhlig (1991).

Two views of the resulting joint distribution of $\rho, \widehat{\rho}$ are displayed in figures 2.10 and 2.11. Figure 2.10 shows the fixed-ρ cross-section, which is the small sample distribution of $\hat{\rho}$. It is skewed towards lower values of $\hat{\rho}$, because of the small sample bias. Comparing with Figure 2.2 we can see that this bias is stronger when the constant term is estimated. We can also see that, as ρ increases, the distribution becomes tighter, but not as fast as that in Figure 2.2.

Figure 2.11 shows the fixed-$\hat{\rho}$ cross-section which is the Bayesian posterior of ρ given $\hat{\rho}$. Consistently with the observation of Andrews (1993), the balance between the downward bias of OLS, and its increasing precision, is tilted towards the former. The posterior accounts for this and is skewed towards values of ρ higher then the observed $\hat{\rho}$, so that, in this figure, when we condition on an observed value $\widehat{\rho} = 1$, the mean of the posterior is larger than one. Therefore, in this case we also have that there is no strong dichotomy between Bayesian and classical analysis: both recommend correcting for the bias of OLS.

The general 'S-prior' (Uhlig (1994) and (Zivot, 1994, p.569-570)) is:

$$p(\mu|\sigma, y_0, S) = \sigma^{-1}\nu(\rho, S)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\nu(\rho, S)^{-1}(y_0 - \mu)^2\right\} \tag{2.25}$$

with

$$\begin{aligned} \nu(\rho, S) &= (1-\rho^{2S})/(1-\rho) \qquad && |\rho| \neq 1 \\ &= S && |\rho| = 1 \end{aligned} \tag{2.26}$$

[20]Strictly speaking, Uhlig (94) does not derive a 'prior' from the above specification of the initial condition, but he uses it to write the 'exact likelihood'. This is, of course, a semantic issue, since the resulting posterior is the same as if the above specification for the initial condition is considered a prior given the initial observations, which is our preferred interpretation.

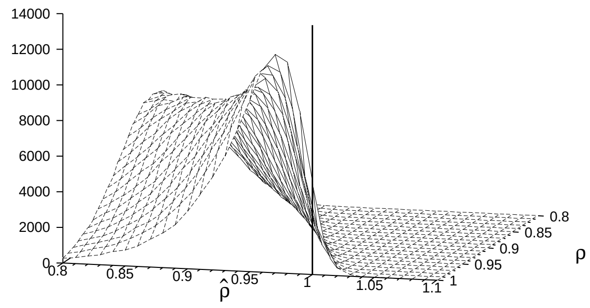

Figure 2.10: Joint frequency distribution of $(\rho, \hat{\rho})$ sliced along the $\rho = 1$ line (as in Figure 3 from Sims and Uhlig (1991)). Model with a constant term, starting from $y_0 = \mu$

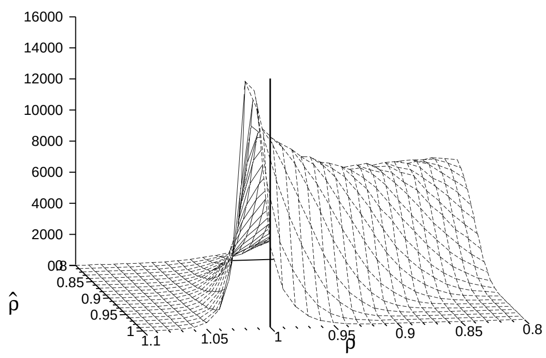

Figure 2.11: Joint frequency distribution of $(\rho, \hat{\rho})$ sliced along the $\hat{\rho} = 1$ line (as in Figure 4 from Sims and Uhlig (1991)). Model with a constant term, starting from $y_0 = \mu$

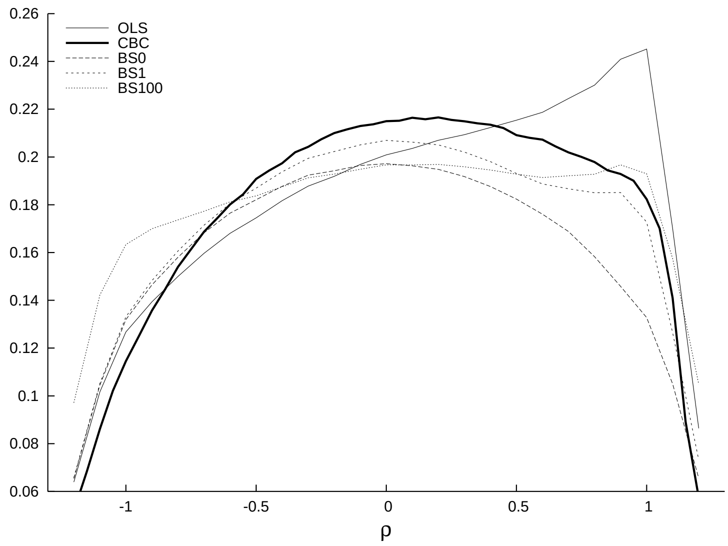

Figure 2.12: RMSE of the OLS, BS0, BS1 and BS100 estimators in a Monte Carlo experiment with $S = 0$, sample size: $T = 25$.

These priors are then complemented with the non-informative prior for ρ and σ:

$$p(\rho, \sigma) \propto \sigma^{-1} \tag{2.27}$$

The $S = 1$ prior is conditionally conjugate, and the full posterior distribution can be conveniently simulated with the Gibbs sampler. For the general S the conjugacy is lost, but the marginal posterior for ρ can be obtained analytically by a straightforward integration of the posterior with respect to μ and σ. The resulting formula is given in Zivot (1994, equation (43)).

The bias and mean squared error of the Bayesian 'S-prior' estimators are examined in a Monte Carlo experiment similar to the one in section 5. We denote the Bayesian posterior means obtained with $S = 0, 1$ and 100 priors respectively as BS0, BS1 and BS100.

We simulate the AR(1) processes with initial conditions corresponding to $S = 0, 1$ and $\infty/1$ and in each case try all 3 Bayesian estimators, in addition to OLS and the CBC. The results are presented in figures 2.12, 2.13 and 2.14.

The BS1 and BS100 estimators perform somewhat better, but are broadly similar to the delta prior: in RMSE terms they beat OLS for the positive ρ and the CBC on an even larger interval (except that the BS100 has slightly worse RMSE than the CBC for ρ around 1). In the range where they perform well, their bias (unreported here) is stronger then that of the CBC, although

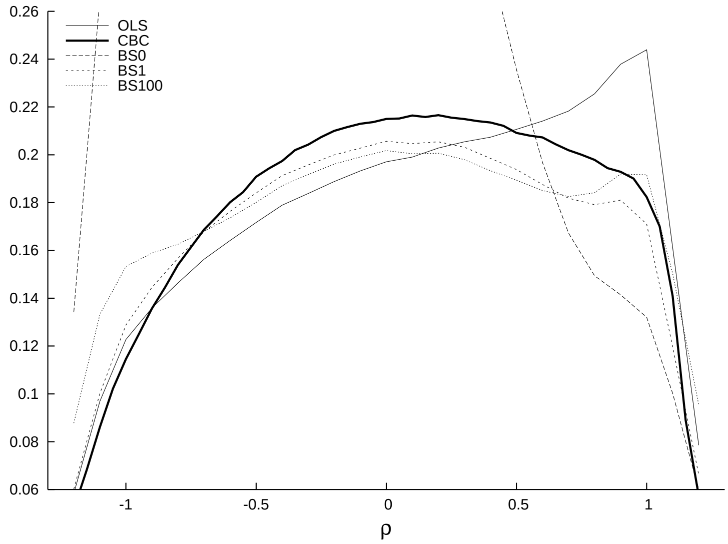

Figure 2.13: RMSE of the OLS, BS0, BS1 and BS100 estimators in a Monte Carlo experiment with $S = 1$, sample size: $T = 25$.

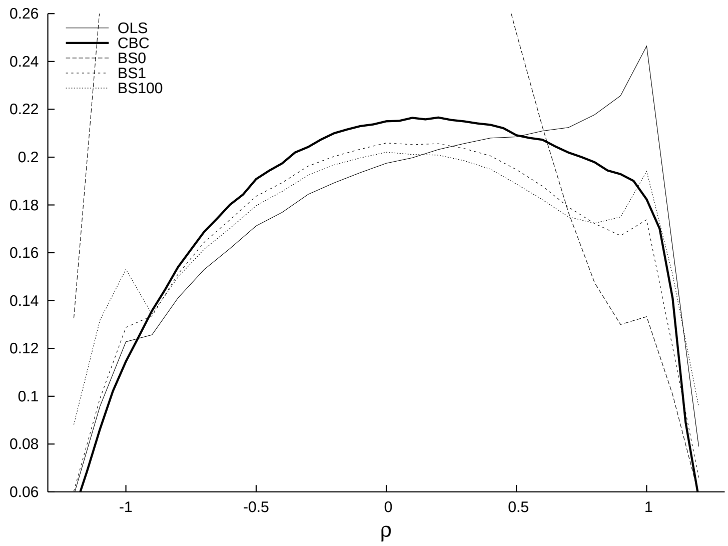

Figure 2.14: RMSE of the OLS, BS0, BS1 and BS100 estimators in a Monte Carlo experiment with $S = \infty$, sample size: $T = 25$.

reduced compared with OLS. They are not very sensitive to the accuracy of the prior, e.g. the BS1 estimator performs similarly also when the prior assumption is false: for the processes starting at the mean (Figure 2.12), and in the stationary distribution (Figure 2.14).

That last observation is not true for the BS0 estimator, which gives a particularly strong advantage over the alternatives when the prior is correct (for the processes starting exactly at the mean, Figure 2.12), but for some parameter values it is very misleading when the prior is wrong. The graphs are truncated, so it cannot be seen that the RMSE of BS0 has a peak of 0.65 (more then 3 times that of the alternatives) around $\rho = -0.6$ for $S = 1$ (Figure 2.13) and of 0.9 (more then 6 times that of the alternatives) around $\rho = -0.9$ for $S = \infty/1$ (Figure 2.14). Nevertheless, in the range which is most relevant in practice, i.e. for high positive values of ρ, BS0 is always more precise then the alternatives.

Appendix 2.B The S-prior for VAR's

In this section we generalize the 'S=1' prior to the VAR(P) model, evolving around an arbitrary exogenous process (for example a constant mean, or a linear trend).

Let Y be a $T \times J$ matrix gathering T observations on J jointly endogenous variables. Let L denote a lag operator and define a $P \times 1$ vector $l \equiv [\mathrm{L}, \mathrm{L}^2, \ldots, \mathrm{L}^P]'$. Define $X \equiv l' \otimes Y = [Y_{-1}, Y_{-2}, \ldots, Y_{-P}]$, a $T \times JP$ matrix with P lagged matrices Y. W is a $T \times K$ matrix with K exogenous variables (in case of a time invariant mean, $W = \iota_T$, a $T \times 1$ vector of ones, while in case of a linear trend $W = [\iota_T, \tau]$, where $\tau = [1, 2, \ldots, T]'$). $Z \equiv l' \otimes W = [W_{-1}, W_{-2}, \ldots, W_{-P}]$ is a $T \times KP$ matrix with P lagged matrices W. Finally. U is a $T \times J$ matrix with T independent normal vectors of length J, each with a variance Σ.

Deviations of Y from the exogenous component evolve according to a VAR(P) process, perturbed by shocks U:

$$Y - W\Gamma = (X - Z\,(I_P \otimes \Gamma))\,\Pi + U \tag{2.28}$$

where Π is a $JP \times J$ matrix of VAR coefficients and Γ is a $K \times J$ matrix with means and slopes of the time trend for each endogenous variable. Reduced form of the above model is:

$$Y = W\tilde{\Gamma} + X\Pi + U \tag{2.29}$$

with

$$\tilde{\Gamma} = \Gamma - (l' \otimes \Gamma)\Pi = \Gamma(I - \Phi_1\mathrm{L} - ... - \Phi_P\mathrm{L}^P)' \tag{2.30}$$

where Φ_i is a matrix of VAR coefficients of lag i, so that $\Pi = (\Phi_1, \Phi_2, \ldots, \Phi_P)'$. Equation (2.30) is a generalization of the restriction $\alpha = \mu(1-\rho)$ (2.20) for the AR(1) case, and analogously to that case it has implications when the lag polynomial in Φs contains some unit roots. When Γ consists only of constant terms, (2.30) guarantees that those corresponding to series that have unit roots are suppressed, and the random walks have no drifts. When Γ contains constant terms and time trends, (2.30) ensures that the time trends corresponding to unit root series are suppressed, consistently with the postulate of Uhlig (1994).

Likelihood conditional on initial P observations is proportional to:

$$L(Y; \Pi, \Gamma) \propto |\Sigma|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2} \operatorname{tr} \Sigma^{-1} U'U\right\} \tag{2.31}$$

A common noninformative prior for the error variance and the autoregressive coefficients is:

$$p(\Sigma, \Pi) \propto |\Sigma|^{-\frac{1}{2}(J+1)} \tag{2.32}$$

The prior for Γ must be proper, to compensate the indeterminacy at unit root implied by restriction (2.30). A 'bias-correcting' prior assumption, in the spirit of the previous sections, relates the coefficients of the deterministic component of the process to the initial observations:

$$Y_0 - W_0\Gamma = U_0 \tag{2.33}$$

where Y_0 is a $T_0 \times J$ matrix with T_0 initial observations on the endogenous variables, W_0 is a $T_0 \times K$ matrix with the corresponding values of the exogenous variables, and U_0 a $T_0 \times J$ matrix with T_0 independent normal vectors length J with a variance Σ_0. We take Σ_0 to be the OLS estimate of the VAR errors.

The implied prior for Γ is normal, centered on the OLS estimate of Γ on the T_0 initial observations:

$$p(\text{vec}\Gamma | Y_0, W_0, \Sigma_0) = \text{N}(\text{vec}\hat{\Gamma}_0, \Sigma_0 \otimes W_0'W_0^{-1}) \tag{2.34}$$

where

$$\hat{\Gamma}_0 \equiv (W_0'W_0)^{-1} W_0'Y_0$$

The simplifying assumption in (2.33), which delivers conjugacy of the prior, is that the first T_0 deviations from the deterministic component are independent, and only starting with time period 0 the time dependence implied by Π kicks in. An exact multivariate and multilag counterpart to the Uhlig's S-prior would assume, that the process had been equal to its deterministic component up to period $-S$ and then started evolving according to

(2.28). But then the variance of U_0 would depend in a complicated way on Π and the conjugacy of the prior would be lost. The prior in (2.34) implies that the initial T_0 observations are used only to infer on the coefficients of the deterministic process Γ, but not the Π. This may be a simplification, but it is still better then the case of the flat prior, where the information in the initial observations is ignored altogether.

The joint posterior implied by (2.31),(2.32) and (2.34) is:

$$p(\Pi, \Gamma, \Sigma | Y, Y_0) \propto |\Sigma|^{-\frac{T+J+1}{2}} \exp\left\{ -\frac{1}{2} \operatorname{tr}(\Sigma^{-1} U'U + \Sigma_0^{-1} U_0' U_0) \right\} \tag{2.35}$$

The conditional posteriors are:

$$p(\Sigma | \Pi, \Gamma, Y, Y_0) = \mathrm{IW}(U'U, T) \tag{2.36}$$

$$p(\mathrm{vec}\Pi | \Gamma, \Sigma, Y, Y_0) = \mathrm{N}\left(\mathrm{vec}\tilde{\Pi}, \Sigma \otimes (\tilde{X}'\tilde{X})^{-1} \right) \tag{2.37}$$

$$p(\mathrm{vec}\Gamma | \Pi, \Sigma, Y, Y_0) = \mathrm{N}\left(G^{-1} g, G^{-1} \right) \tag{2.38}$$

where IW is an Inverted Wishart distribution,

$$\begin{aligned}
\tilde{X} &\equiv X - Z(I_P \otimes \Gamma) \\
\tilde{\Pi} &= (\tilde{X}'\tilde{X})^{-1} \tilde{X}'(Y - W\Gamma) \\
G &= A'(\Sigma^{-1} \otimes I_T) A + \Sigma_0^{-1} \otimes W_0' W_0 \\
g &= A'(\Sigma^{-1} \otimes I_T) \mathrm{vec}(Y - X\Pi) + \left(\Sigma_0^{-1} \otimes W_0' W_0 \right) \mathrm{vec}\hat{\Gamma}_0 \\
A &= I_J \otimes W - (\Pi' \otimes Z) \left((I_P \otimes K_{JP})(\mathrm{vec} I_P \otimes I_J) \otimes I_K \right)
\end{aligned}$$

and K_{JP} is the commutation matrix (Magnus and Neudecker, 1988, p.46). A sample from the posterior (2.35) can be generated by Gibbs sampler, i.e. by drawing in turn from (2.36), (2.37) and (2.38).

Appendix 2.C Integrating constant for a multivariate normal likelihood

For completeness we put here the derivation of the integrating constant given in (2.11). Throughout we condition on a known error variance Σ.

Likelihood conditional on initial P observations is equal to:

$$L(Y; B) = (2\pi)^{-\frac{TJ}{2}} |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2} \mathrm{vec}(Y - XB)'(\Sigma^{-1} \otimes I_T) \mathrm{vec}(Y - XB) \right) =$$

$$(2\pi)^{-\frac{TJ}{2}} |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2} \left(\mathrm{vec}\, B'(\Sigma^{-1} \otimes X'X) \mathrm{vec}\, B - 2 \,\mathrm{vec}\, B' \mathrm{vec}(X'Y\Sigma^{-1}) + \operatorname{tr} Y'Y\Sigma^{-1} \right) \right) \tag{2.39}$$

It is easy to see that this expression is proportional to the multivariate normal density of B: $N\big(\mathrm{vec}(X'X)^{-1}X'Y, \Sigma \otimes (X'X)^{-1}\big)$ or

$$p(B|Y) = (2\pi)^{-KJ/2}|\Sigma|^{-K/2}|X'X|^{J/2}$$
$$\exp\left(-\frac{1}{2}\,\mathrm{vec}\big(B-(X'X)^{-1}X'Y\big)'\big(\Sigma^{-1}\otimes X'X\big)\,\mathrm{vec}\big(B-(X'X)^{-1}X'Y\big)\right)$$
$$= (2\pi)^{-KJ/2}|\Sigma|^{-K/2}|X'X|^{J/2}$$
$$\exp\left(-\frac{1}{2}\left(\mathrm{vec}\,B'(\Sigma^{-1}\otimes X'X)\,\mathrm{vec}\,B - 2\,\mathrm{vec}\,B'\,\mathrm{vec}(X'Y\Sigma^{-1}) + \mathrm{tr}\,Y'X(X'X)^{-1}X'Y\Sigma^{-1}\right)\right) \tag{2.40}$$

By definition, the normalizing constant is such that:

$$p(B|Y) = \frac{L(Y;B)}{\mathcal{K}_f(Y)}$$

and therefore the formula for the normalizing constant is:

$$\mathcal{K}_f(Y) = \int L(Y;B)dB = \frac{L(Y;B)}{p(B|Y)} =$$
$$(2\pi)^{-\frac{(T-K)J}{2}}|\Sigma|^{-\frac{T-K}{2}}|X'X|^{-\frac{J}{2}}\exp\left(-\frac{1}{2}\,\mathrm{tr}\,Y'(I-X(X'X)^{-1}X')Y\Sigma^{-1}\right)$$
$$= (2\pi)^{-\frac{(T-K)J}{2}}|\Sigma|^{-\frac{T-K}{2}}|X'X|^{-\frac{J}{2}}\exp\left(-\frac{1}{2}\,\mathrm{tr}\big(\hat{E}'\hat{E}\Sigma^{-1}\big)\right) \tag{2.41}$$

Bibliography

Andrews, D. W. K. (1993). Exactly median-unbiased estimation of first order autoregressive / unit root models. *Econometrica*, 61(1):139–165.

Bhargava, A. (1986). On the theory of testing for unit roots in observed time-series. *Review Of Economic Studies*, 53(3):369–384.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (1999). Monetary policy shocks: What have we learned and to what end? In Taylor, J. B. and Woodford, M., editors, *Handbook of Macroeconomics*, number 1A, chapter 2, pages 65–148. Amsterdam: North-Holland.

Doan, T., Litterman, R., and Sims, C. (1984). Forecasting and conditional projections using realistic prior distributions. *Econometric Reviews*, 3(1):1–100.

Doan, T. A. (1992). *RATS User's Manual version 4*. Estima, Suite 612, 1800 Sherman Ave., Evanston, IL 60201, second edition.

Hurwicz, L. (1950). Least-squares bias in time series. In Koopmans, T. C., editor, *Statistical Inference in Dynamic Economic Models*. Wiley, New York.

Jarociński, M. and Marcet, A. (2005). Bootstrap mean squared error optimization. work in progress.

Kendall, M. G. (1954). Note on the bias in the estimation of autocorrelation. *Biometrika*, XLI:403–404.

Kilian, L. (1998). Small-sample confidence intervals for impulse response functions. *The Review of Economics and Statistics*, 80(2):218–230.

MacKinnon, J. G. and Smith, A. A. (1998). Approximate bias correction in econometrics. *Journal of Econometrics*, 85:205–230.

Magnus, J. R. and Neudecker, H. (1988). *Matrix Differential Calculus with applications in statistics and econometrics*. John Wiley and Sons, New York.

Marriott, F. H. C. and Pope, J. A. (1954). Bias in the estimation of autocorrelations. *Biometrika*, XLI:393–403.

Ni, S. and Sun, D. (2003). Noninformative priors and frequentist risks of bayesian estimators of vectpr-auoregressive models. *Journal of Econometrics*, 115:159–197.

Orcutt, G. H. and Winokur, H. S. (1969). First order autoregression: Inference, estimation, and prediction. *Econometrica*, 37(1):1–14.

Phillips, P. C. (1991). To criticize the critics: An objective bayesian analysis of stochastic trends. *Journal Of Applied Econometrics*, 6(4):333–364.

Quenouille, M. H. (1949). Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society Series B*, 11:68–84.

Roy, A. and Fuller, W. A. (2001). Estimation for autoregressive time series with a root near 1. *Journal Business and Economic Statistics*, 19(4):482–493.

Sims, C. and Uhlig, H. (1991). Understanding unit rooters: A helicopter tour. *Econometrica*, 59(6):1591–1599.

Sims, C. A. (1988). Bayesian skepticism on unit root econometrics. *Journal of Economic Dynamics and Control*, 12:463–474.

Sims, C. A. (1996). Inference for multivariate time series models with trend.

Sims, C. A. and Zha, T. (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, 39(4):949–68.

Uhlig, H. (1994). On Jeffreys prior when using the exact likelihood function. *Econometric Theory*, 10(3-4):633–644.

Zivot, E. (1994). A bayesian analysis of the unit-root hypothesis within an unobserved components model. *Econometric Theory*, 10(3-4):552–578.

Chapter 3

Determinants of Economic Growth: Will Data Tell? Joint with Antonio Ciccone[1]

Abstract: By now, many factors inhibiting and facilitating the economic growth of countries have been proposed. Will (imperfect) international aggregate income data allow us to tell which matter when all factors are treated symmetrically a priori? We find that growth determinants emerging from agnostic cross-country regression approaches are sensitive to arguably small variations in the international income data. For example, although minor by past standards, the revision of the Penn World Table 6.0 income data for 1960-96 leads to substantive changes in growth determinants. Agnostic empirical approaches also yield only limited coincidence regarding growth determinants when using international income estimates obtained with alternative methodologies.

3.1 Introduction

Why does income grow faster in some countries than others? Following Kormendi and Meguire (1985), Grier and Tullock (1989), and Barro (1991), most empirical work deals with limited international data by focusing on cross-country regressions with few explanatory variables (for recent reviews see Barro and Sala-i-Martin, 2004, and Durlauf, Johnson, and Temple, 2005). Selected variables are usually motivated by their importance in the theoretical debate. But as growth theory is not unequivocal regarding key

[1] ICREA-UPF

explanatory variables there is usually only partial overlap among the explanatory variables considered. As a result, it often remains an open question whether a proposed growth factor would remain significant if combined with another explanatory variable in the literature.

It is therefore natural to try and determine the factors inhibiting and facilitating economic growth when all of the variables suggested in the literature are treated symmetrically a priori. The basic idea is to see in which direction the data guides an agnostic. Following the initial contributions of Levine and Renelt (1992) and Sala-i-Martin (1997a,b), this is the task tackled by the recent contributions of Fernandez, Ley, and Steel (2001), Sala-i-Martin, Doppelhofer, and Miller (2004), and Hendry and Krolzig (2005). As the suggested explanatory variables are many, agnostic empirical approaches inevitably end up with a large number of variables relative to the number of countries. We show that as a result, the growth determinants emerging from such approaches are sensitive to arguably small changes in the international income data. For example, the minor corrections of the 1960-96 Penn World Table 6.0 income data incorporated into the latest version of the PWT lead to substantive changes in the determinants of growth. Overall, the two versions of the PWT agree on fewer than half the determinants of 1960-96 growth. And even when the two datasets agree on a growth factor's statistical significance, there is substantial disaccord on the magnitude of the effect. Most of the disagreement regards growth factors that are prominent in the literature and key for economic policy.

Given the difficulties faced by those who estimate international incomes, data imperfections are not surprising. A major challenge is the limited coverage and quality of the underlying price benchmark and national income data. Many income estimates are therefore obtained by extrapolation and margin of errors for income levels and growth rates are often large (e.g. Summers and Heston, 1991; Heston, 1994) Another reason for imperfect data are errors that arise when building a complex database. The PWT income data has therefore been subject to periodic revisions to eliminate errors, incorporate improved national income data, or account for new price benchmarks. For example, the 1960-96 income data in the PWT 6.1 (the latest version) corrects the now-retired PWT 6.0 estimates.[2] Relative to previous revisions, changes seem minor. For example, the correlation between 1960-96 income per capita growth rates in the two databases is 0.98, which is high compared to the 0.88 correlation of PWT 6.0 income per

---

[2] For an explanation see the PWT website page http://pwt.econ.upenn.edu/whatsnew.htm.

capita growth for 1960-1985 with the corresponding data from PWT 5.6 (the previous version, which was based on older national income and less price benchmark data). But despite the strong positive correlation between the income data in PWT 6.0 and 6.1, the two datasets yield rather different 1960-96 growth determinants using agnostic empirical approaches. Consider for example the Bayesian Averaging of Classical Estimates approach of Sala-i-Martin, Doppelhofer, and Miller (2004). SDM use this approach to obtain the 1960-96 growth determinants with PWT 6.0 data. When we update their results with the corrected data in the PWT 6.1 it turns out that the two versions of the PWT disagree on 16 of 29 determinants of 1960-96 growth.

Such disaccord affects many growth factors that have featured prominently in the literature. For instance, a key policy question is whether countries more open to international trade have experienced faster economic growth (e.g. Sachs and Warner, 1995; Rodrik and Rodriguez, 2000; Warner, 2003). It is therefore not surprising that the cross-country evidence on the relationship between trade openness and economic growth is referred to in reports or speeches by officials of the United Nations, International Monetary Fund, and the Organization of American States as well as briefings to government officials.[3] But while trade openness continues to be a statistically significant 1960-96 growth factor with PWT 6.0 data according to the Bayesian criterion of Sala-i-Martin, Doppelhofer, and Miller (2004), we find it to be insignificant with PWT 6.1 data. Bayesian Averaging of Classical Estimates with PWT 6.0 income data also yields that malaria prevalence in the 1960s is a statistically significant negative determinant of 1960-96 growth. As pointed out by Sachs (2005), such a finding has important consequences for spending on malaria prevention. But with PWT 6.1 data for the 1960-96 period, malaria prevalence in the 1960s is an insignificant growth factor. Another much debated barrier to economic development is ethnolinguistic fractionalization, a possible explanation for poor growth in Africa (e.g. Easterly, 2001; Alesina et al., 2003; Alesina and La Ferrara, 2005). But while this factor is a statistically significant negative 1960-96 growth determinant with PWT 6.0 income data, it is insignificant with PWT 6.1 data. Government consumption and real exchange rate

---

[3] For interesting examples see the 2000 and 2004 International Monetary Fund's World Economic Outlook, the 2001 report of the United Nations Economic Commission for Africa, the New Zealand Treasury's 1999 Briefing to the Incoming Government, and Frankel (1999) as well as Salazar-Xirinachs (2001).

distortions, usually seen as a proxy of market distortions, are two further examples of prominent growth factors that turn insignificant using PWT 6.1 income data (e.g. Barro, 1991; Caselli, Esquivel, and Lefort, 1996; Sala-i-Martin, 1997a,b; Acemoglu, Johnson, and Robinson, 2002).

Growth factors that appear statistically insignificant using the Bayesian Averaging of Classical Estimates approach with PWT 6.0 income data but become significant with PWT 6.1 data play an equally important role in the literature. Consider for example the effect of more democratic regimes on economic growth. Existing cross-country studies do not find a statistically significant positive link between democracy and growth (e.g. Limongi and Przeworski, 1993; Barro, 1996). Sala-i-Martin, Doppelhofer, and Miller's (2004) analysis using Bayesian Averaging of Classical Estimates with PWT 6.0 income data confirms this finding for the 1960-96 period. But PWT 6.1 data for the same period yields a statistically significant positive link between more political rights and economic growth, which is line with the recent evidence from democratic transitions in Papaioannou and Siourounis (2005). Fertility is an insignificant 1960-96 growth determinant with PWT 6.0 data. With PWT 6.1 data, high fertility has a statistically significant negative effect on economic growth, confirming previous findings (e.g. Barro, 1991, 1996; Barro and Lee, 1994). Natural resource abundance proxied by hydrocarbon deposits also emerge as a statistically significant growth determinant using PWT 6.1 data, although the effect is positive and therefore in contradiction with previous findings (e.g. Sachs and Warner, 1995; Gylfason, 2001; Kronenberg, 2004). And population density also becomes a statistically significant positive growth factor, which contrasts with mixed results in the cross-country literature (e.g. Sachs and Warner, 1995; Hall and Jones, 1999) but is consistent with the evidence on regional productivity differentials (e.g. Ciccone and Hall, 1996).

Old and corrected PWT income estimates for 1960-96 also yield sizable differences regarding the magnitudes of growth effects for many factors. Consider for example conditional convergence, maybe the most emphasized issue in the empirical cross-country literature since its beginnings (e.g. Kormendi and Meguire, 1985, Grier and Tullock, 1989, Barro, 1991; Barro and Sala-i-Martin, 1992, Mankiw, Romer, and Weil, 1992). According to the PWT 6.0 estimates of Sala-i-Martin, Doppelhofer, and Miller (2004), an agnostic would conclude that the half-life of income per capita differences between countries is 115 years. But the corrected income data for 1960-96 yield a much shorter half-life of 45 years. Cross-country regressions have also been used to assess the impact of climate change on economic growth (e.g. Congressional Budget Office, 2003). It is therefore interesting to take the effect of the population share in the tropics on 1960-96 economic growth as a second example. Both old and corrected PWT income data yield a negative effect. But according to the PWT 6.1 data the unconditional posterior mean effect is 60% smaller in absolute value. The differences in magnitudes in these two examples are large from the point of view of applied theory or economic policy but actually smaller than the average gap across all 67 growth factors considered.

Our finding that the minor differences between the PWT 6.1 and 6.0 data result in large discrepancies regarding statistical significance and impact for many growth factors is not specific to the Bayesian Averaging of Classical Estimates approach. We find similar results using the Bayesian Model Averaging approach of Fernandez, Ley, and Steel (2001) and the General-to-Specific approach of Hendry and Krolzig (2005).

These results using the PWT panel suggest that growth determinants emerging from these agnostic empirical approaches may be sensitive to minor variations in the income data. To examine this issue, we apply the three approaches considered to perturbed PWT 6.1 income data. For example, in one case we work with a series of 30 growth perturbations that are generated by adding a small amount of noise to PWT 6.1 income per capita growth. Each of the 30 growth perturbations is selected to have a correlation between 0.975 and 0.985 with PWT 6.1 growth data for 1960-96 (the interval is centered on the 0.98 correlation between the PWT 6.0 and 6.1 growth data). We then determine the 1960-96 growth determinants for each of the 30 growth perturbations taking the list of 67 potential explanatory variables of Sala-i-Martin, Doppelhofer, and Miller (2004) as a starting point. Using the Bayesian Averaging of Classical Estimates approach we find that 42 of the 67 variables are robust for at least one growth perturbation but only 4 variables are always robust. More than half of the candidate variables therefore emerge as growth determinants for some growth perturbation but not for another. This is a reason for concern given the imperfection of available international income estimates. The results are almost identical using Bayesian Model Averaging of Fernandez, Ley and Steel (2001). The General-to-Specific approach yields that none of the candidate explanatory variables always form part of the final parsimonious model.

We also examine the performance of the three agnostic approaches using a second series of growth perturbations obtained by averaging the first series and PWT 6.1 growth for 1960-96. This results in 30 growth perturbations with a correlation between 0.99 and 0.995 with PWT 6.1 growth. But despite this almost perfect correlation with PWT 6.1 data, we still find sizable differences in indicators of statistical significance across growth perturbations using Bayesian Averaging of Classical Estimates. For example, consider the distribution of the Bayesian posterior inclusion probability of individual variables across the 30 growth perturbations. As a summary of the spread of this distribution, we take the ratio of its 90-th to the 10-th percentile. For the variable capturing more democratic political regimes, this ratio is 7.07. For the population share in the tropics it is 3.8 and for ethnolinguistic fractionalization it is 2.69. The average 90-th/10-th percentile ratio across all 67 growth factors we consider is 2.57. Differences in the magnitudes of growth effects across the 30 growth perturbations are also large for many factors. For example, while the effect of natural resource abundance proxied by hydrocarbon deposits on growth is positive for all growth perturbations, the posterior unconditional mean effect at the 90-th percentile is 554% greater than that at the 10-th percentile. The effect of a larger government consumption share on growth, on the other hand, is always negative. But the posterior mean effect at the 10-th percentile is 516% greater than that at the 90-th percentile in absolute value. The gap between mean effects in these two examples is approximately equal to the average gap across all 67 growth factors. When we examine why the results of agnostic approaches are sensitive to small variations in the income data, the number of candidate explanatory variables considered at the outset turns out to be key. The sensitivity diminishes when we start out with relatively small sets of candidate variables.

Differences between PWT income estimates are mostly due to later versions incorporating additional primary data. But international income estimates may also differ for methodological reasons. For example, the PWT and World Bank PPP international income comparisons are based on the same national income and price benchmark data but alternative aggregation methods. The World Bank uses the so-called Elteto-Koves-Szulc (EKS) system while the PWT relies on the Geary-Khamis (GK) system. A key property of the GK system is that it is additive and therefore yields internationally comparable results for GDP as well as its components. The main advantage of the EKS system used by the World Bank, as well as the Statistical Agency of the European Union and of the OECD, is that it is grounded in consumer theory (e.g. Diewert, 1978). Methodological differences between the World Bank and the PWT data would be a lesser concern if the two datasets resulted in similar growth determinants, but they are often in disaccord. For example, using the Bayesian Model Averaging of Fernandez, Ley and Steel yields disagreement on 8 of 9 growth factors for the 1975-1996 period (WB PPP income estimates are only available since 1975). The General-to-Specific approach yields disaccord on 13 of 19 variables and Bayesian Averaging of Classical Estimates on 17 of 28 variables. It is important to be aware that the methodological choices underlying international income estimates have such strong effects on growth determinants.

The well-known international income estimates of Maddison (e.g. Maddison, 1989) focus on the production side of national income accounts, not the expenditure side as the PWT and the World Bank estimates. The two approaches are usually seen as complementary (Maddison and van Ark, 1994; van Ark and Timmer, 2001). International income data constructed from the production side of national income accounts is available from the Groningen Growth and Development Centre and The Conference Board. When we compare 1960-96 growth determinants obtained with GGDC-TCB and PWT income data, we again find only limited coincidence. For example, the two datasets disagree on around 80% of the growth determinants using Bayesian Model Averaging of Fernandez, Ley and Steel or the General-to-Specific approach of Hendry and Krolzig and around 60% using Bayesian Averaging of Classical Estimates. This limited coincidence regarding growth determinants implies that any definite assessment of the factors inhibiting and facilitating economic growth will require settling the difficult methodological issues that arise when estimating international income differences (e.g. United Nations, 1992).

The remainder of the paper is structured as follows. Section II contains a brief discussion of the three agnostic approached considered. In Section III, we obtain the 1960-96 growth determinants using the three approaches with the corrected income data from the PWT 6.1. We also analyze differences with PWT 6.0 growth determinants and the sensitivity of growth determinants to small variations in the income data. Section IV compares the growth determinants that emerge with using aggregate income estimates based on alternative methodologies. Section V concludes.

3.2 Three Agnostic approaches to Robust Growth Determinants

Consider the problem of identifying determinants of economic growth over a certain period, using cross-country data. We start with K candidate explanatory variables, gathered in vector $\boldsymbol{x}$. We refer to subsets of $\boldsymbol{x}$, denoted by $\mathbf{x}^j$ ($j=1,...,2^K$), as Models. Consider cross-country growth regressions of the form:

$$y_i = x_i^j b^i + e_i^j \qquad (1)$$

where subscript i denotes countries (i=1…T), j – models, y_i is the growth rate of per capita GDP in country i, $x^j{}_i$ is a vector of k^j explanatory variables, included in model j, b^j is a vector of coefficients reflecting the effects of explanatory variables on growth and $e^j{}_i$ is an error term.

What makes the problem tricky is the small sample size. With large number of observations (large T) it would be sufficient to run the regression with the full set of candidate variables, and the coefficients of the irrelevant ones would all be insignificantly different from zero. However, with T close to K such approach is too inefficient, and as a result too noisy, to be of practical interest (the inefficiency results from estimating coefficients of many irrelevant variables), and with T<K it is even infeasible. Smaller models are more tightly estimated, but they give very disparate results.

The problem of identifying determinants of economic growth among a list of candidate variables, using the available cross-country data, has been approached in the literature in two ways. The Bayesian way is to attach probabilities to possible alternative models, and consider the effect of variables on average, across models. Variants of this approach have been applied in Sala-i-Martin, Doppelhofer and Miller (2004) and Fernandez, Ley and Steel (2001). The Classical way, followed in Hendry and Krolzig (2004), is to use a series of econometric tests to single out the true model. Below, we provide a short description of methodologies used in the three mentioned papers. More complete discussions can be found in the papers themselves and in references quoted therein.

3.2.1 Bayesian methods: FLS-BMA and BACE

The Bayesian approach to model uncertainty, known as Bayesian Model Averaging (BMA), is to attach prior probabilities to competing models, and update them with the information contained in the data, to obtain posterior probabilities of models. Any results of interest are then computed on average across all models, with weights equal to posterior model probabilities. Bayesian Model Averaging is discussed, among others, in Leamer (1978) and in Hoeting et al. (1999). In the context of empirical growth research, its use is advocated among others in Brock and Durlauf (2001). In this paper we use two specifications: from Fernandez, Ley and Steel (2001) (which we refer to as FLS-BMA) and Sala-i-Martin, Doppelhofer and Miller (2004) (which, following the authors, we refer to it as BACE – Bayesian Averaging of Classical Estimates).

The ingredients of a BMA exercise are: prior for the models *p(Mj)*, prior for the regression coefficients in the models b^j and the likelihood function of the data given models. The choice of the likelihood function is natural: that of the linear regression with Gaussian errors (1). In specification of the priors, both Fernandez, Ley and Steel (2001) and Sala-i-Martin, Doppelhofer and Miller (2004) limit the subjectivity of their Bayesian analyses in two ways. First, their priors treat all variables symmetrically. Second, they strive to diminish the effect of the priors on the coefficients in each model.

Prior probability of a model in Fernandez, Ley and Steel (2001) is equal for each possible model, irrespective of its size and composition. Sala-i-Martin, Doppelhofer and Miller (2004) use a more general prior, which still treats all variables symmetrically, but favors models of a certain specified size $\tilde{k}$, and downweighs smaller or larger models. In their baseline specification, Sala-i-Martin, Doppelhofer and Miller (2004) favor smaller models then Fernandez, Ley and Steel (2001).

Prior for coefficients in the models, assumed in Sala-i-Martin, Doppelhofer and Miller (2004), is noninformative, and it results in posterior distributions which coincide with the classical sampling distributions of OLS coefficients (thus the name: Bayesian Averaging of Classical Estimates). This prior is not a proper distribution, but it emerges as a limit of a certain sequence of increasingly loose priors.[4] Fernandez, Ley and Steel (2001) consider a similar class of loose priors, but instead of taking a limit, they choose such finite parameter values which result, in practice, in a negligible effect on the posterior.

The posterior probabilities of models are derived to be proportional to:

$$p(M_j \mid y) \propto p(M_j) T^{-\frac{k_j}{2}} SSE_j^{-\frac{T}{2}} \qquad (2)$$

in Sala-i-Martin, Doppelhofer and Miller (2004) (see their equation 6, or equation 4.16 in Leamer (1978)), and

$$p(M_j \mid y) \propto p(M_j) \left(\frac{g}{g+1} \right)^{\frac{k_j}{2}} \left(\frac{1}{g+1} SSE_j + \frac{g}{g+1} (y-\bar{y})'(y-\bar{y}) \right)^{-\frac{T-1}{2}} \qquad (3)$$

[4] See Sala-i-Martin, Doppelhofer and Miller (2004) and Leamer (1978) for details.

in Fernandez, Ley and Steel (2001) (see their equation 8). In the above formulas, SSE_j is the sum of squared OLS residuals in model j, g = $1/\max\{T,K^2\}$, y is a vector of growth rates for all countries and $\bar{y}$ is the average growth rate in the sample.

In both approaches, the terms multiplying the prior model probability contain a premium for the fit and a punishment for including many parameters. It is easy to see, that as T becomes large, expression in (3) converges to that in (2). In small samples, however, they differ: it can be seen that in FLS-BMA the punishment for overparametrization is tougher, and the premium for the fit is lower, and consequently this approach favors smaller models then the BACE.

Overall, the strategy of both approaches is similar: both treat all variables symmetrically a priori, and give a high posterior weight to parsimonious models with a good fit. The particular modeling choices in the two approaches have offsetting effects: prior for models favors smaller models more strongly in Sala-i-Martin, Doppelhofer and Miller (2004), but the prior for coefficients results in a higher premium for small models in Fernandez, Ley and Steel (2001). Ultimately, the average posterior model size tends to be similar, but the weighting of individual models is different in the two approaches.

The output of BMA exercises, which is reported in both papers, and on which we concentrate in this paper, consists of *posterior probabilities of inclusion* of individual variables, *unconditional posterior means* of their coefficients and *posterior means* of coefficients *conditional on inclusion* of the variable. Posterior inclusion probability of a variable is computed as a sum of posterior probabilities of all models which include a given variable. Unconditional posterior mean of a coefficient is computed as a weighted average of posterior means of coefficients in individual models, with weights proportional to the posterior model probabilities. Whenever a model includes a given variable, its posterior mean coincides exactly, in the case of BACE, and approximately, in FLS-BMA, with the OLS point estimate. When a variable is not included, its coefficient is taken to be zero. Posterior means conditional on inclusion are computed similarly, but averaging only over models which include a given variable.

As regards the robustness of the two approaches to the measurement error in the data, notice that their verdicts depend crucially on the fit of alternative models (sums of squared residuals are raised to very high exponents, which amplifies small differences in fit). With many, often correlated, candidate variables, we will have many models with a similar fit. In a small sample, even if the measurement error were in principle unrelated to all variables, it might, by pure chance, improve the fit of some models and worsen it in others. As a result, the weighting of the models will change, producing different posterior inclusion probabilities and posterior means of coefficients.

3.2.2 General-to-Specific approach - PcGets

Hendry and Krolzig (2004) identify determinants of economic growth using a General-to-Specific strategy, [5] as implemented in their PcGets computer package, which performs automatic model selection. The algorithm sifts relevant from irrelevant variables by performing a series of econometric tests. It is described among others in Hendry and Krolzig (2001) and Hendry and Krolzig (2005). Throughout this paper by General-to-Specific we will mean their particular implementation of this strategy.

The algorithm starts with the General Unrestricted Model (GUM), which includes all the candidate variables. Then, exclusion of individual variables is being tested, until arriving at a final parsimonious model which dominates all alternatives. Common objections to such procedures are addressed: First, the problem of spurious significance from repeated testing is limited by the design of testing strategies (combination of individual and joint tests), and played down by the authors as unimportant in practice. Second, the obtained result does not depend on which insignificant variable is removed first, i.e. path independence is ensured.

Path independence is guaranteed by following multiple reduction paths, starting with each feasible initial deletion. Final outcomes of consecutive reductions are called terminal selections. The union of all obtained terminal selections becomes the new GUM, and the process is repeated, until either a unique model, or a set of irreducible undominated models emerges. In the latter case, the final model is chosen based on an information criterion.

The output of the PcGets algorithm is the final, specific model, which includes only the variables which have a statistically significant effect on the dependent variable. The coefficients are estimated with OLS, and standard t-ratios and R^2 are reported.

[5] See Campos, Ericsson and Hendry (2005) for a comprehensive discussion of General-to-Specific methodology and related literature.

Measurement error in the variables distorts the outcomes of all steps and thus can affect final results. In fact, because models with poorer performance are rejected at some stage and never taken into account anymore, the final results can be more affected by a measurement error in the data, then in Bayesian Model Averaging algorithms, where all models are always taken into account, even if with lower weights.

3.3 1960-96 Growth Determinants Using PWT 6.1

We now use the three agnostic approaches to obtain the 1960-96 growth determinants with the corrected PWT 6.1 income data and compare the results to those obtained with PWT 6.0 data. As the comparisons suggest that growth determinants identified by the three agnostic approaches may be sensitive to minor changes in the data, we also examine the sensitivity of the three approaches to small, random perturbations of PWT 6.1 income data.

3.3.1 Corrected 1960-96 Growth Determinants

We start by implementing Bayesian Averaging of Classical Estimates approach. Then we turn to the Bayesian Model Averaging approach of Fernandez, Ley and Steel and conclude with the General-to-Specific approach.

3.3.1.1 BACE approach

In Table II, we list all 67 candidate explanatory variables and their posterior inclusion probabilities (InclProb) as well as conditional (CondMean) and unconditional mean effects (UncondMean) according to PWT 6.1 income data. We also include comparisons of these PWT 6.1 BACE statistics with their PWT 6.0 counterparts. The MAX/MIN InclProb column gives the larger of the PWT 6.1 and PWT 6.0 inclusion probability divided by the smaller inclusion probability. For example, the inclusion probability of Political Rights is 22% with PWT 6.1 and only 3.8% with PWT 6.0. Hence, the MAX/MIN ratio of the two inclusion probabilities is 6.17. The [MAX-MIN]/MIN[ABS] CondMean statistic is the larger minus the smaller conditional mean, relative to the smaller absolute mean value. For example, the conditional mean of Population Density in 1960 is 2.048E-05 with PWT 6.1 and only 1.30E-05 with PWT 6.0. Hence, the [MAX-MIN]/MIN[ABS] statistic is 0.58. The last column contains exactly the same statistic for the unconditional mean.

Sala-i-Martin, Doppelhofer, and Miller (2004) classify variables with a posterior inclusion above 0.1 as robust and those with a posterior inclusion probability above 0.08 but below 0.1 as marginally robust. We simply refer to all variables with a posterior inclusion probability above 0.08 as robust. Using this criterion, the variables numbered 1 through 20 are robust with PWT 6.1 data. Among these variables those in bold type are not robust with PWT 6.0. For example, Air Distance to Big Cities has a posterior inclusion probability of 42% with PWT 6.1 and is therefore classified as robust with PWT 6.1. But with PWT 6.0 the inclusion probability is only 3.3%, which makes this variable a non-robust growth factor. The variables numbered 21 through 67 are not robust with PWT 6.1. Among these variables those in bold type are robust with PWT 6.1. Hence, variables in bold type are those that are robust using one dataset but not the other. For example, Years Open 1950-94 has a posterior inclusion probability of 5.29% with PWT 6.1 but an inclusion probability of 12.5% with PWT 6.0. Hence it becomes non-robust when switching to the corrected PWT data.

Overall, there is considerable disagreement regarding robust long-run growth factors between the PWT 6.0 and 6.1 income data. The two datasets disagree on 16 of the 29 variables that are robust using one of them (55%). The MAX/MIN inclusion probability statistic also signals disaccord. For example, the average MAX/MIN ratio of the variables that are classified as robust with PWT 6.1 is 3.4. Hence, the larger of the two PWT 6.1 and 6.0 inclusion probabilities exceeds the smaller inclusion probability on average by a factor above 3 for these variables. When we consider all 67 variables, the average MAX/MIN ratio is 2.36.

The growth factors with the largest increase in the posterior inclusion probability when switching to the corrected PWT income data are Air Distance to Big Cities, Political Rights, and Timing of Independence. Those with the largest decrease are Malaria Prevalence in 1960s, Spanish Colony, and Ethnolinguistic Fractionalization. Broadly speaking, the growth factors that become more robust using the PWT 6.1 income data appear related to institutions (e.g. Political Rights, Timing of Independence) and economic geography (e.g. Air Distance to Big Cities, Population Density in 1960, Population Density Costal in 1960s). Growth factors losing importance relate to trade openness (e.g. Openness Measure 1965-74, Years Open 1950-94, Real Exchange Rate Distortions) and the weight of the government in the economy (e.g. Gov. Consumption Share in 1960s, Public Investment Share).

The two datasets also yield different results regarding conditional and unconditional mean effects of the 67 variables. Consider Malaria Prevalence in 1960s for example. The gap between the larger and the smaller mean conditional on inclusion for this variable is more than 23 times the smaller absolute value. When averaged across all 67 variables, the [MAX-MIN]/MIN[ABS] difference between the PWT 6.1 and 6.0 conditional mean is 4.33. The [MAX-MIN]/MIN[ABS] difference for the unconditional mean, the more important statistics from an agnostic point of view, shows even stronger disagreement. Population Density 1960, for example, enters positively using both versions of the PWT, but the size of the effect with PWT 6.1 is 12.63 times that with PWT 6.0. Hence, the [MAX-MIN]/MIN[ABS] statistic is 11.63. The [MAX-MIN]/MIN[ABS] statistic for the unconditional mean averages to 15.08 for the variables that are robust with PWT 6.1 data and to close to 12.5 for all 67 variables. The median gap between larger and smaller unconditional mean relative to the smaller absolute value is 1.93 when we consider only variables that are robust with PWT 6.1 data and 2.69 when we consider all 67 variables.

3.3.1.2 FLS-BMA approach

In Table III, we summarize the results of implementing FLS-BMA with the corrected PWT income data and compare these results to those obtained with PWT 6.0 data. Using FLS-BMA, the two datasets disagree on 15 of 30 growth determinants (50%), which is close to the 55% share we obtained when implementing BACE. The extent of disaccord between the two datasets indicated by the average MAX/MIN inclusion probability ratios is also similar. Using FLS-BMA, the average MAX/MIN ratio of inclusion probabilities across all variables is 2.22, somewhat smaller than the corresponding BACE figure. Among the variables that are robust with PWT 6.1, the average MAX/MIN statistic is 2.85, again somewhat smaller than using BACE. Differences in the conditional and unconditional means between the two datasets are also similar using FLS-BMA than BACE.

Overall, the corrected PWT 6.1 income data leads to similar changes using both the BACE and the FLS-BMA approach. For example, the 3 growth determinants with the largest increase in the posterior inclusion probability with PWT 6.1 coincide (Air Distance to Big Cities, Political Rights, Timing of Independence) and so do the 3 growth determinants with the larges decrease in inclusion probability (Malaria Prevalence in 1960s, Spanish Colony, Ethnolinguistic Fractionalization). Moreover, according to the FLS-BMA approach, the corrected PWT income data also lead to greater posterior inclusion probabilities for variables related to institutions and economic geography and smaller inclusion probabilities for variables related to trade openness and the weight of the government in the economy. Interestingly, the two Bayesian approaches also coincide in that, according to the PWT 6.1 data there is faster conditional convergence. The conditional mean upon inclusion of initial log income per capita is -0.0086 with PWT 6.0 data using both the BACE and the FLS-BMA approach. With PWT 6.1 data, both approaches yield a mean effect of -0.0153. This difference implies that the half-life of income per capita difference shrinks from 80 years to 45 years, approximately one generation. The gap is even wider when we use the unconditional mean effect to calculate implies half-lives of income per capita differences.

3.3.1.3 General-to-Specific approach

In Table IV, we summarize the results using the PcGets approach. This approach yields only few important growth determinants. For example, with PWT 6.1 the only variables entering the specific model are Fraction Confucius, Fraction Population in Tropics, Fraction Buddhist, and Investment Price. While these variables are a subset of those emerging as robust from the two Bayesian approaches, they are generally not among those with highest inclusion probabilities (the exception is Investment Price). For example, the specific model does not contain the log GDP in 1960, the growth determinant with highest Bayesian posterior inclusion probability whether we use BACE or FLS-BMA, which implies that the model does not generate conditional convergence. The PcGets approach also yields considerable disagreement regarding the robust determinants of long-term growth between the PWT 6.0 and 6.1 income data. The two datasets disagree on 5 of the 7 variables that are robust using one of them (72%).

3.3.2 Sensitivity Analysis

Our results using the corrected PWT suggest that robust growth determinants identified by the three agnostic approaches considered could be sensitive to minor perturbations of the data. To examine this issue further, we now apply the three agnostic approaches to growth and income series that are generated by perturbating PWT 6.1 data for 1960-96. The distributions from which we draw such perturbations are calibrated using the differences between PWT 6.1 and 6.0 data.

3.3.2.1 Generating Perturbed Data

The sensitivity analysis we focus on uses data that only perturbs 1960-96 income per capita growth of the PWT 6.1, which we denote by the vector $\hat{y}_{1960-96}^{PWT\,6.1}$ where the *i*-th element refers to country *i*. We generate a sequence *k*=1,2,… of perturbed growth vectors (growth perturbations) $\hat{y}_{1960-96}^{k}$ as $\hat{y}_{1960-96}^{k} = \hat{y}_{1960-96}^{PWT\,6.1} + e^{k}$ where e^{k} is drawn from a normal distribution with zero mean. The variance-covariance matrix of this distribution is assumed to be diagonal with the *i*-th element on the diagonal equal to $a + b \log y_{i,1960}^{PWT\,6.1}$ where $y_{i,1960}^{PWT\,6.1}$ is PWT 6.1 income per capita in 1960 of country *i*. When *b* is strictly negative, this formulation implies smaller errors in richer countries. The parameters *a* and *b* are obtained by regressing the squared differences between PWT 6.1 and 6.0 growth rates on a constant and PWT 6.1 log income per capita in 1960; *a* is set equal to the intercept of this regression and *b* equal to the slope (which is significantly negative, indicating as expected that revisions for higher-income countries were smaller on average). We draw from the distribution until we have generated 30 growth perturbations whose correlation with PWT 6.1 growth is between 0.975 and 0.985 (the interval in centered on 0.98, the correlation between PWT 6.1 and 6.0 growth rates).

We also want to examine the growth determinants that emerge using the three agnostic approaches when the correlation between perturbed growth rates and PWT 6.1 growth is above 0.99. To do so, we generate a second set of 30 growth perturbations by averaging each element of the first set and PWT 6.1 growth. This is equivalent to generating new growth perturbations $\hat{y}_{1960-96}^{k,average}$ as PWT 6.1 growth plus half the disturbance e^{k} used to generate the first set of perturbations, i.e. $\hat{y}_{1960-96}^{k,average} = \hat{y}_{1960-96}^{PWT\,6.1} + 0.5e^{k}$. The resulting 30 growth perturbations have a correlation with PWT 6.1 growth data between 0.99 and 0.995.

In addition we perform a sensitivity analysis based on data that perturbs both 1960-96 growth and log 1960 per capita income of the PWT 6.1. Perturbed log income data is generated using an approach that is analogous to that for growth rates. The main new element is that our calibration of the distribution used to draw growth and log income perturbations takes into account the correlation between PWT 6.1 and 6.0 growth differences on the one hand and log 1960 income differences on the other. The calibrated distribution is used to simultaneously draw growth and level perturbations until we have generated 30 growth and level perturbations such that: (i) the growth perturbations have a correlation with PWT 6.1 growth between 0.975 and 0.985; (ii) the log income perturbations have a correlation with PWT 6.1 log 1960 income per capita between 0.9475 and 0.9525 (the interval in centered on 0.95, the correlation between PWT 6.1 and 6.0 log 1960 income per capita). A second set of perturbed growth and level data is generated by simply halving the perturbations underlying the first set. The resulting 30 growth perturbations have a correlation with PWT 6.1 growth data between 0.99 and 0.995 and the 30 level perturbations have a correlation with PWT 6.1 log 1960 income per capita between 0.985 and 0.995. A detailed description of how we generate perturbed data can be found in the Appendix.

3.3.2.2 BACE approach

Applying BACE to each of the 30 growth perturbations yields a distribution of BACE statistics (posterior inclusion probabilities, posterior means upon inclusion, posterior unconditional means) for each of the 67 candidate variables. In Table V, we characterize these distributions for each candidate variable using 4 summary measures:

(i) The median BACE posterior inclusion probability. By comparing this statistic for different candidate variables, we can assess which appears to be more robustly related to growth.

(ii) The BACE posterior inclusion probability at the 90-th percentile of the distribution relative to that at the 10-th percentile (90PCT/10PCT). This statistic allows us to assess the dispersion of posterior inclusion probabilities.

(iii) The posterior mean conditional on inclusion at the 90-th percentile minus that at the 10-th percentile, relative to the smaller absolute value ([90PCT-10PCT]/MIN[ABS]). This allows us to assess the dispersion of conditional means across growth perturbations.

(iv) The [90PCT-10PCT]/MIN[ABS] statistic for the unconditional mean. This statistic could be greater or smaller than that of the conditional mean, depending on the correlation between posterior inclusion probabilities and conditional means.

The variables in the small box are those robust across all 30 growth perturbations (variables where all 30 posterior inclusion probabilities are above 8%) and the variables in the large box are those robust for at least one growth perturbations. The variables are ordered according to the smallest posterior inclusion probability across growth perturbations.

Table V shows that 42 variables are robust for at least one growth perturbations. Only 4 variables are always robust (log GDP in 1960, Investment Price, Primary School in 1960, Population Density Coastal 1960). As there are a total of 67 candidate variables, this implies that almost 2/3 of the candidates turn out to be robust at least once and less than 1/10 of the variables robust at least once are always robust. To put it differently, more than half of the candidate variables (38=42-4) emerge as growth determinants for some growth perturbation but not for another. Because posterior inclusion probabilities are sensitive to growth perturbations, even a variable with a median inclusion probability of 60% (East Asian Dummy) is non-robust for some growth perturbations. Moreover, of the variables that turn out to be sometimes robust, several have median posterior inclusion probabilities of only 2% (Religion Measure, British Colonial Dummy, Nominal Government GDP Share in the 1960s).

Another way to see that posterior inclusion probabilities are sensitive to growth perturbations is by examining the 90PCT/10PCT statistic. For example, Population Density Coastal in 1960s, although robust for all 30 growth perturbations, has an inclusion probability at the 90-th percentile that is almost 6 times that at the 10-th percentile. Averaging across all variables yields a 90PCT/10PCT inclusion probability of 4.57.

Posterior means conditional on inclusion are also sensitive to growth perturbations. In the case of the British Colonial Dummy, for example, the difference between the conditional mean at the 90-th percentile and that at the 10-th percentile is more than 19 times the absolute mean at the 10-th percentile. The average across all variables of the [90PCT-10PCT]/MIN[ABS] statistic for the conditional means is 14.63. The same average calculated for the unconditional mean is 33.32.

The dispersion of BACE statistics remains large when we focus on the 75-th and the 25-th percentiles of the distribution (instead of the 90-th and the 10-th). For example, for posterior inclusion probabilities, the 75PCT/25PCT statistic averages to 2.12 across all variables, while the average of the [75PCT-25PCT]/MIN[ABS] statistic is 2.82 for conditional and 5.49 for unconditional means.

Another way of assessing the sensitivity of BACE statistics with respect to growth perturbations is to compare the results of each growth perturbation with those using PWT 6.1 growth using the measures employed when comparing PWT 6.1 with PWT 6.0 results. Using this approach, we obtain an average MAX/MIN statistic for the posterior inclusion probability of just below 2, while the [MAX-MIN]/MIN[ABS] statistic for the unconditional posterior mean averages to 5.6.

Table VI summarizes the results of applying BACE to the 30 growth perturbations whose correlation with PWT 6.1 growth is between 0.99 and 0.995. The number of growth determinants that are robust at least once (in the large box) is now 35, still more than half of the 67 candidate variables. Of the 35 variables robust at least once, 9 are always robust (26%). Because posterior inclusion probabilities continue to be sensitive to growth perturbations, even variables with a median inclusion probability of 24% (Life Expectancy in 1960, Air Distance to Big Cities) are non-robust in some cases. And among the sometimes robust variables there are 2 with a median inclusion probability of only 3% (Socialist Dummy, Ethnolinguistic Fractionalization). The 90PCT/10PCT statistic also continues to show that posterior inclusion probabilities are sensitive to growth perturbations, even for those variables that are always robust. For example, the inclusion probability at the 90-th percentile is more than twice that at the 10-th percentile for 5 of the 9 variables in this category. Averaging across all variables yields a 90-th/10-th percentile inclusion probability ratio of 2.57. This number more than doubles to 5.43 when we average the maximum to minimum inclusion probability ratio. The indicator of the magnitude of growth effects relevant to an agnostic, the unconditional posterior mean, also continues to be sensitive to growth perturbations. The average across all variables of the [90PCT-10PCT]/MIN[ABS] statistic is 5.19. Focusing on the 75-th and the 25-th percentiles of the distribution (instead of the 90-th and the 10-th) continues to indicate sizable dispersion in the BACE statistics across growth perturbations. For example, for posterior inclusion probabilities, the 75PCT/25PCT statistic averages to 1.62 across all variables. And the average of the [75PCT-25PCT]/MIN[ABS] statistic is 2.8 for unconditional means. Comparing the BACE results of each growth perturbation with those using PWT 6.1 and averaging across growth perturbations yields a MAX/MIN posterior inclusion probability statistic of 1.4 and a [MAX-MIN]/MIN[ABS] statistic for the unconditional posterior mean of 2.4.

Application of BACE to the perturbed growth and log 1960 income data yields results that are in line with those in Tables V and VI. For example, applying BACE to the first set of 30 growth and level perturbations yields that 47 of the 67 candidate variables are robust at least once (70%) and only 3 variables are always robust (log GDP in 1960, Investment Price, Primary

School in 1960). Halving the perturbations yields that 37 variables are robust at least once and 9 are always robust. The correlation of median inclusion probabilities with those obtained by perturbing growth rate only is 0.99.

3.3.2.3 FLS-BMA approach

Applying FLS-BMA to the first set of 30 growth perturbations yields similar results to using BACE. For example, the 4 variables that are robust in all 30 cases are identical and 63% of the 67 variables continue to be robust at least once. Applying FLS-BMA to the perturbed growth data that is almost perfectly correlated with PWT 6.1 growth and to the perturbed growth and level data yields almost identical results as BACE.

3.3.2.4 General-to-Specific approach

Application of the PcGets approach to the two sets of growth perturbations yields that no variable forms part of the specific model in all cases. Only 3 variables are part of the specific model in 90% of the cases. These variables are log GDP in 1960, Investment Price, and Primary School in 1960, the same 3 variables that were always robust when we applied BACE and FLS-BMA to perturbed data. 30 of the 67 variables are part of the specific model at least once. The point estimates emerging from the PcGets approach show considerable dispersion. For example, the coefficient on log GDP in 1960 varies from -0.015 to -0.006. The first estimate implies a half-life of income per capita differences of 45 years, while the second implies that the half-life is 115 years.

3.3.2.5 Simulations with Short Lists of Variables

Are robust growth determinants as identified by the three agnostic approaches less sensitive to data perturbations when the list of candidate variables is shorter? One way to address this question is to extract a subset of the 67 candidate variables and apply the three agnostic approaches to this shorter list of candidates. Results based on one short list of candidate variables only could, however, be driven by the particular set of candidates that one happens to extract rather than the shortness of the list. We therefore randomly extract 50 short lists of candidate variables and then compare average statistics across these 50 lists with our results using the all-inclusive 67-candidates list. We settle on 50 lists because extracting more did not change our findings. We focus on lists with 15 candidate variables because shorter ones can easily be dealt with in the standard multiple regression framework. We report the results of these experiments in Table VII.

Consider first the BACE approach. In Table V, we find that using BACE only around 10% of the variables that are robust for at least one growth perturbation are robust for all growth perturbations (this share is 11%, when we define robustness as posterior inclusion probability exceeding 10.4% threshold – which is the prior inclusion probability). When we calculate the same statistic for each of the 50 lists of 15 candidate variables and then average across the 50 lists, we get this fraction to be above 40%. This indicates that posterior inclusion probabilities are less noisy when starting with shorter candidate lists. The PCT90/PCT10 posterior inclusion probability average across all candidate variables reveals a similar picture. In the case with 67 candidates in Table V, this average is 4.57. The inclusion probability at the 90-th percentile is therefore on average more than 4 times that at the 10-th percentile. When we calculate the same statistic for each of the 50 lists of 15 candidate variables and then average across lists, the number is 1.66. Hence, on average, the inclusion probability at the 90-th percentile is only 66% greater than that at the 10-th percentile when starting out with 15 candidates.

When we repeat these calculations with the perturbed growth data that is almost perfectly correlated with PWT 6.1 growth, 61% of the variables robust for at least one growth perturbations are robust for all perturbed growth perturbations when averaging across the 50 lists of 15 variables. The corresponding number in Table VI is 26% (or 30% in Table VII, when we define robustness as posterior inclusion probability exceeding prior inclusion probability). Moreover, while the inclusion probability at the 90-th percentile is 2.57 times that at the 10-th percentile when averaging across all candidate variables in Table V, the same statistic averaged across the 50 lists of 15 candidate variables is only 1.32. The average gap between the inclusion probability at the 90-th percentile and that at the 10-th percentile is therefore only 32% in the case with 15 candidates. When we repeat these calculations with the growth and level perturbations, we find very similar results. Moreover, results are very similar using FLS-BMA.

Application of the General-to-Specific approach to shorter lists of candidates also yields more agreement on the important growth determinants than application to the all-inclusive, 67-variable list. For example, recall that starting out with 67 candidate variables yields that no variable was always part of the specific model, independently of which of the two sets of growth perturbations we use. In contrast, when we calculate the fraction of variables which belong to the specific model for each of the 50 lists of 15 candidate variables and average across the 50 lists, the resulting fraction is 28%. With the set of growth perturbations that is almost perfectly correlated with PWT 6.1 growth, this share rises to 48%. Hence, shorter lists of candidate variables lead to a reduced sensitivity of the PcGets approach to data perturbations.

3.4 PWT Growth Determinants Compared

We now examine to what extent robust growth determinants obtained with PWT income data coincide with those obtained with World Bank data or data from the Groningen Growth and Development Centre and The Conference Board.

3.4.1 PWT compared to WDI

The World Bank's World Development Indicators contain estimates of international incomes since 1975. This allows us to compare the 1975-96 growth determinants emerging from the PWT and the WDI.

In Table VIII, the comparison is based on the BACE approach. In this case, the two datasets disagree on 17 of the 28 growth determinants that are robust using one of them (61%). The MAX/MIN inclusion probability ratios also indicate disaccord. Averaging across all variables that are robust according to the PWT, the ratio is 4.4. Averaging across all 67 variables yields that the larger of the two inclusion probabilities exceeds the smaller probability by a factor of 2.4. Compared to the PWT, the WDI income data suggest that religious variables (Fraction Buddhist, Fraction Muslim, Fraction Confucius are robust with the PWT but not the WDI), geography (African Dummy, Latin American Dummy), and colonial origin (Spanish Colony, British Colony Dummy) are less important.

Differences in conditional and unconditional posterior means between the two datasets can be examined using the [MAX-MIN]/MIN[ABS] statistic. The average across variables of this statistic is large in the case of the unconditional mean, whether we average across variables robust using PWT 6.1 (18.32) or across all 67 variables (22.81). These large averages are driven by a few very large [MAX-MIN]/MIN[ABS] statistics, see Malaria Prevalence in 1960s and Spanish Colony for example. The median [MAX-MIN]/MIN[ABS] statistic for the two groups of variable is also sizable however, 3.06 and 2.61 respectively. A researcher will therefore reach rather different conclusions depending on the dataset used. Similar results were obtained when comparing results from FLS-BMA approach.

In Table IX, we compare the important growth determinants according to the WDI and the PWT data using the General-to-Specific approach. Now the two datasets disagree on 13 of the 19 variables that are important using one of them (68%). Interestingly, one of the variables the two datasets agree on using PcGets is the negative effect of log GDP in 1960.

Note that comparing the results for individual variables with the corresponding results for the 1960-1996 period one has to take into account both different time periods, and different country composition of the samples (see Appendix B for the information on countries included in each sample).

3.4.2 PWT Compared to GGCD-TCB

Comparing the robust growth determinants according to the PWT and the GGCD-TCP income data also yields considerable disaccord. In Table X, where we use BACE, the two datasets disagree on 19 of 28 variables (68%). The average MAX/MIN ratio is 5.4 when we focus on the growth determinants that are robust with the PWT and almost 4 when we consider all 67 variables. Broadly speaking, compared to the PWT, the GGDC-TCP data leads to smaller posterior inclusion probabilities for variables related to trade (Primary Exports 1970, Openness Measure 1965-74) and greater inclusion probabilities for variables related to colonization (Colony Dummy, British Colony Dummy) and the weight of the government in the economy (Government Share of GDP in 1960s, Public Investment Share).

Average differences in unconditional posterior means between the PWT 6.1 and GGDC-TCP datasets are very large, whether we consider the average across variables robust using PWT 6.1 (21.81) or across all 67 variables (52.11). These large averages are driven by extreme disaccord regarding the unconditional effect of a few variables, like Defense Spending Share and British Colonial Dummy for example. The median [MAX-MIN]/MIN[ABS] statistic for the two groups of variable are 4.4 and 4.06 respectively and therefore still signal strong disagreement. Results obtained with FLS-BMA are qualitatively similar.

In Table XI, we summarize the results obtained using the General-to-Specific approach implemented in the PcGets package. Now the PWT and the GGDC-TCB income data disagree on 10 of 13 growth determinants

(77%). Interestingly, the PcGets approach yields conditional convergence with the PWT 6.1 but not the GGDC-TCB income data. [6]

3.5 Conclusions

Which are the determinants of economic growth? We have answered this question for different periods using three empirical approaches that treat the proposed factors symmetrically a priori. But if the past is any guide, this answer will undergo substantive changes when income estimates are revised to eliminate errors, incorporate improved national income data, or account for new price benchmarks. At least this is what happened when we used the corrected income estimates in the latest version of the PWT (PWT 6.1) to update the list of 1960-96 growth determinants obtained with the previous version (PWT 6.0) by Sala-i-Martin, Doppelhofer, and Miller (2004). The two datasets yield disagreement on more than half the determinants of 1960-96 growth, and most of the disaccord regards factors that are much debated and key for economic policy. This limited coincidence cannot be attributed to PWT 6.1 income revisions being large by past standards. For example, the correlation between 1960-96 income per capita growth rates in the two databases is 0.98, while the correlation of PWT 6.0 income per capita growth 1960-1985 with the corresponding data from the preceding version (PWT 5.6) is only 0.88.

Minor variations in international income estimates therefore appear to result in substantial changes in the growth determinants emerging from agnostic approaches. To examine this issue further, we apply the three approaches considered to slightly perturbed PWT 6.1 data. For example, in one case we work with a series of 30 growth perturbations that are generated by adding noise to PWT 6.1 income per capita growth. Each of the 30 perturbations is selected to have a correlation between 97.5 and 98.5 percent with PWT 6.1 growth data for 1960-96 (the interval is centered on the 0.98 correlation between the PWT 6.1 and 6.0 growth data). We find that more

[6] When using the GGDC-TCB data we have fewer countries than variables, but this can be handled by all the used approaches. In the BMA schemes of Fernandez, Ley, and Steel (2001) and Sala-i-Martin, Doppelhofer, and Miller (2004) all models larger then the number of observations can be, for convenience, ruled out a priori. Such large models would have very little prior and posterior probability anyway, and some experimentation shows that ruling out large models has a negligible effect on the results. The PcGets algorithm, when facing less observations then variables, starts from all feasible combinations of variables, collects all the resulting terminal models, and then proceeds in the usual way (see section 2).

than half of the candidate variables are robust growth determinants for at least one growth perturbation and non-robust for at least another. Continuous indicators of statistical significance also prove to be sensitive to minor variations in the income data, even when we work with growth perturbations selected to have a correlation between 99 and 99.5 percent with the PWT 6.1 growth data.

We also find that agnostic, all inclusive empirical approaches result in only limited coincidence regarding statistically significant growth factors when using international income estimates obtained with alternative methodologies. For example, although based on the same primary data, PWT and World Bank income estimates still yield disagreement on 8 of 9 growth determinants for the 1975-1996 period using the Bayesian Model Averaging approach and 17 of 28 growth determinants using the Bayesian Averaging of Classical Estimates Approach. Such disaccord is a result of the sensitivity of the agnostic approaches considered to variations in the income data and the often substantial differences in estimates based on alternative methodologies.

That cross-country evidence on the determinants of economic growth is an important guide for researchers and policy makers alike is clear from the many times they refer to it. Our results show that evidence from approaches that treat many explanatory factors symmetrically a priori is sensitive to minor data imperfections. If the income data used in the empirical analysis is imperfect, growth factors that appear statistically or economically significant may therefore actually be insignificant and vice versa.

Our results indicate the importance of developing a better understanding of the sources and statistical properties of the approximation error in the international income data. For an agnostic to take the results emerging from empirical approaches that treat many explanatory factors symmetrically a priori at face value, he/she has to put much weight on the precision on the data.

3.6 Appendix A: Generating Perturbed Income and Growth Data

This appendix describes generation of the artificial datasets, used to study how data revisions can affect the results about determinants of economic growth. We first determine basic characteristics of the revisions of the income data, namely their variance (taking into account heteroscedasticity) and covariance between level and growth data revisions. Then, in experiment I, we generate 30 perturbed datasets adding to the actual data random noise, with the same characteristics. In experiment II we generate 30 datasets perturbed with a smaller noise.

3.6.1 Statistical characteristics of the PWT revision

Our baseline analysis concerns the difference between the PWT6.0 and PWT6.1 datasets, and is performed on the sample of 84 countries for which we have data in both datasets[7] (as well as all explanatory variables). We focus on the variable which is used in the empirical growth studies: real GDP per capita in constant prices, computed with the chain method (PWT mnemonics: RGDPCH), denoted further by *y*. We use the following notation:

Average growth rate of per capita GDP: $\hat{y}_{1960-96}^{PWT6.1} = \frac{\log y_{1996}^{PWT6.1} - \log y_{1960}^{PWT6.1}}{36}$, and analogously for the PWT6.0 dataset.

Revision of the 1960 per capita GDP: $e = \log y_{1960}^{PWT6.1} - \log y_{1960}^{PWT6.0}$

Revision of the 1960-1996 average growth: $\hat{e} = \hat{y}_{1960-96}^{PWT6.1} - \hat{y}_{1960-96}^{PWT6.0}$

First, ε is demeaned, since it involves a change of units from 1985 US dollars to 1996 US dollars. Assuming that there was no revision of the US data, one could also subtract the US inflation between 1985 and 1996, instead of subtracting the mean. The appropriate US inflation index would

[7] We also performed a similar exercise for the difference between PWT5.6 and PWT6.1 datasets, which is even greater. In that case, however, because of data unavailability in the earlier dataset, we had to switch to a different sample period then in the main analysis in the text: 1960-1985.

seem to be the GDP deflator. However, the PWT are based on separate deflators for different GDP components, and it is not obvious how they should be weighted. Subtracting the mean seems to be a reasonable and straightforward alternative which is conservative, i.e. it slightly underestimates the variance of the real revision.

The revisions of level and growth of GDP $\left(\hat{e}, e\right)$ are modeled jointly, reflecting the construction of the dataset: 1960 GDP data are obtained by combining the 1960 information with the GDP in the base year (which is 1985 in PWT6.0 and 1996 in PWT6.1), projected back using real growth rates between 1960 and the base year. This implies a negative correlation between any revision of the growth rate and revision of the 1960 level. Indeed, the correlation of $\left(\hat{e}, e\right)$ is -0.48.

Size of the revisions is negatively related to the level of income, i.e. data for poorer countries is more affected by revisions. This can be seen in figures A1 and A2. To model this, we regress the square of growth and level revisions on the 1960 real GDP (data from PWT6.1). The results are presented in Table A1, and they show that the relationship between the size of the revisions and the level of income is statistically significant.

Table A1. OLS regression of squares of revisions of income data (growth and levels) on the level of income.

Dependent variable:	zhat2	z2
Constant	0.000160	0.61
	0.000048	0.17
logy60pwt61	-0.000019	-0.07
	0.000006	0.02
R2	0.10	0.11
Number of observations	84	84

3.6.2 Generation of a perturbed dataset

In both regressions in table A1 the coefficients on the 1960 GDP are significant. Therefore, when simulating the artificial 'revisions' of the GDP data, we use the fitted values from the above regressions as the variances of the generated noise (denoted by $s^2_{\hat{e},i}$ and $s^2_{e,i}$, where i denotes country). The fitted values turn out to be negative for 17 richest countries in the case of $\hat{e}^2$ and 9 richest countries in the case of e^2. For these countries, we take the variance of the noise to be zero, i.e. we assume that their data will not be subject to further changes.

$$s^2_{\hat{e},i} = \max\left(0, a_{\hat{e}^2} + b_{\hat{e}^2} \log y^{PWT6.1}_{1960,i}\right)$$

$$s^2_{e,i} = \max\left(0, a_{e^2} + b_{e^2} \log y^{PWT6.1}_{1960,i}\right)$$

kth artificial revision of data for country i is constructed as:

$$\hat{e}^k_i = s_{\hat{e},i} h_{1,i}$$

and

$$e^k_i = \left(\frac{r_i}{s_{\hat{e},i}} \hat{e}^k_i + \sqrt{1 - r_i^2}\, h_{2,i} \right) s_{e,i}$$

where $\eta_{1,i}$ and $\eta_{2,i}$ are independent Normal(0,1) random numbers drawn for country i and

$$r_i = \begin{cases} -0.48 \text{ where } \min\left(s^2_{\hat{e},i}, s^2_{e,i}\right) > 0 \\ 0 \text{ otherwise} \end{cases}$$

This procedure ensures that the drawn errors have the same variances as the observed ones, and that their correlation (whenever they are both nonzero) is -0.48.

The ith observation in the kth perturbed dataset is constructed as:

$$\hat{y}^k_{1960-96,i} = \hat{y}^{PWT6.1}_{1960-96,i} + \hat{e}^k_i$$

$$\log y^k_{1960,i} = \log y^{PWT6.1}_{1960,i} + e^k_i$$

We repeat this for all 84 countries in the sample.

3.6.3 Experiments I and II

In experiment I, to ensure that a particular realization of the random errors did not produce an excessively large revision, we check the correlation of each generated dataset with the original PWT6.1 data. The correlation between PWT6.1 and PWT6.0 growth data is 0.977, and between PWT6.1 and PWT6.0 level data it is 0.947. To approximately match this, we discard a perturbed dataset if the correlations between PWT6.1 and the generated data differ from the above correlations by more then 0.0025. We repeat this procedure, until we obtain 30 perturbed datasets with the matched correlations.

In experiment II we repeat the same steps in generating the random noise, but in the end compute perturbed data as:

$$\hat{y}^{k}_{1960-96,i} = \hat{y}^{PWT6.1}_{1960-96,i} + 0.5\hat{e}^{k}_{i}$$

$$\log y^{k}_{1960,i} = \log y^{PWT6.1}_{1960,i} + 0.5e^{k}_{i}$$

In this case, the correlations of the resulting data with the PWT6.1 data are already much higher then correlations between actual datasets: around 0.993 (smallest: 0.990) for growth data and 0.987 (smallest: 0.983) for level data. Therefore, in experiment II we do not discard any datasets.

Another way to assess how big the perturbations of the datasets are is to look at the R^2 in the regression of the perturbed data on the actual PWT6.1 data (and a constant term). The R^2 of these regressions in Experiment I are around 0.954 (smallest: 0.950) for growth data and around 0.898 (smallest: 0.893) for level data. In Experiment II the corresponding R^2 are 0.986 (smallest: 0.979) and 0.974 (smallest: 0.966). The regression of PWT6.0 growth data on PWT6.1 growth data yields R^2=0.954, for levels the figure is 0.896. Therefore, by this criterion also perturbations in Experiment I are similar to the recent revision, while perturbations in Experiment II are much smaller.

Summary statistics about perturbed datasets in both experiments are reported in Table A2.

Table A2: Characteristics of the perturbed datasets.

	growth (1)		level(2)	
	correlation (3)	R2 (4)	correlation (3)	R2 (4)
PWT6.0	0.977	0.954	0.947	0.896
Experiment I				
min	0.975	0.950	0.945	0.893
average	0.977	0.954	0.947	0.898
max	0.979	0.959	0.949	0.901
Experiment II				
min	0.990	0.979	0.983	0.966
average	0.993	0.986	0.987	0.974
max	0.995	0.991	0.992	0.984

Notes:
(1) Average growth rate of real GDP in 1960-1996
(2) Log of the 1960 real GDP
(3) Correlation with PWT6.1 data
(4) R2 in the regression on a constant term and PWT6.1 data

3.7 Appendix B: Information on the common samples of the three dataset comparisons

Sala-i-Martin, Doppelhofer, Miller (2004) dataset, full sample - 88 countries:

Algeria, Benin, Botswana, Burundi, Cameroon, Cent'l Afr. Rep., Congo, Egypt, Ethiopia, Gabon, Gambia, Ghana, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mauritania, Morocco, Niger, Nigeria, Rwanda, Senegal, South africa, Tanzania, Togo, Tunisia, Uganda, Zaire, Zambia, Zimbabwe, Canada, Costa Rica, Dominican Rep., El Salvador, Guatemala, Haiti, Honduras, Jamaica, Mexico, Panama, Trinidad & Tobago, United States, Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Paraguay, Peru, Uruguay, Venezuela, Hong Kong, India, Indonesia, Israel, Japan, Jordan, Korea, Malaysia, Nepal, Pakistan, Philippines, Singapore, Sri Lanka, Syria, Taiwan, Thailand, Austria, Belgium, Denmark, Finland, France, Germany, West, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Turkey, United Kingdom, Australia, New Zealand, Papua New Guinea

Common sample for Sala-i-Martin, Doppelhofer, Miller (2004) and PWT6.1 income data, 84 countries:
Algeria, Benin, Botswana, Burundi, Cameroon, Cent'l Afr. Rep., Congo, Egypt, Ethiopia, Gabon, Gambia, Ghana, Kenya, Lesotho, Madagascar, Malawi, Mauritania, Morocco, Niger, Nigeria, Rwanda, Senegal, South africa, Tanzania, Togo, Uganda, Zaire, Zambia, Zimbabwe, Canada, Costa Rica, Dominican Rep., El Salvador, Guatemala, Honduras, Jamaica, Mexico, Panama, Trinidad & Tobago, United States, Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Paraguay, Peru, Uruguay, Venezuela, Hong Kong, India, Indonesia, Israel, Japan, Jordan, Korea, Malaysia, Nepal, Pakistan, Philippines, Singapore, Sri Lanka, Syria, Taiwan, Thailand, Austria, Belgium, Denmark, Finland, France, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Turkey, United Kingdom, Australia, New Zealand, Papua New Guinea

Common sample for the Sala-i-Martin, Doppelhofer, Miller (2004) variables for 1975-1996 (taken from the same sources as their data for 1960-1996), PWT6.1 income data and World Bank WDI income data, all series for 1975-1996, 86 countries:
Algeria, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cent'l Afr. Rep., Congo, Egypt, Gabon, Gambia, Ghana, Kenya, Lesotho, Madagascar,

Malawi, Mali, Mauritania, Morocco, Niger, Nigeria, Rwanda, Senegal, Sierra Leone, South africa, Togo, Tunisia, Zaire, Zambia, Zimbabwe, Canada, Costa Rica, Dominican Rep., El Salvador, Guatemala, Haiti, Honduras, Jamaica, Mexico, Panama, Trinidad & Tobago, United States, Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Paraguay, Peru, Uruguay, Venezuela, Hong Kong, India, Indonesia, Israel, Japan, Jordan, Korea, Malaysia, Nepal, Pakistan, Philippines, Singapore, Sri Lanka, Syria, Thailand, Austria, Belgium, Denmark, Finland, France, Germany, West, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Turkey, United Kingdom, Australia, New Zealand, Papua New Guinea

Common sample for the Sala-i-Martin, Doppelhofer, Miller (2004) data, PWT6.1 income data and GGDC-TCB income data, 53 countries:
Algeria, Congo, Egypt, Ethiopia, Ghana, Kenya, Morocco, Nigeria, South africa, Tanzania, Canada, Guatemala, Mexico, United States, Argentina, Brazil, Chile, Colombia, Ecuador, Peru, Venezuela, Hong Kong, India, Indonesia, Israel, Japan, Jordan, Korea, Malaysia, Pakistan, Philippines, Singapore, Sri Lanka, Syria, Taiwan, Thailand, Austria, Belgium, Denmark, Finland, France, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Turkey, United Kingdom, Australia, New Zealand

3.8 Bibliography

Acemoglu, Daron, Simon Johnson, and James Robinson, "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution," Quarterly Journal of Economics, 2002, 117 (4), 1231–94.

Alesina, Alberto and Eliana La Ferrara, "Ethnic Diversity and Economic Performance," Journal of Economic Literature, Sep 2005, 43, 721–61.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg, "Fractionalization," Journal of Economic Growth, 2003, 8 (2), 155–94.

Barro, Robert J., "Economic Growth in a Cross Section of Countries," Quarterly Journal of Economics, May 1991, 106 (2), 407–43.

Barro, Robert J., "Democracy and Growth," Journal of Economic Growth, 1996, 1 (1), 1–27.

Barro, Robert J., "Determinants of Economic Growth: A Cross-Country Empirical Study," NBER Working Papers 5698, National Bureau of Economic Research, Inc August 1996.

Barro, Robert J., "Inflation and growth," Proceedings, 1996, (May), 153–169.

Barro, Robert J., "Institutions and Growth, an Introductory Essay," Journal of Economic Growth, 1996, 1 (2), 145–48.

Barro, Robert J and Jong-Wha Lee, "Sources of economic growth," Carnegie-Rochester Conference Series on Public Policy, 1994, 40, 1–46.

Barro, Robert J and Xavier Sala-i-Martin, "Convergence," Journal of Political Economy, April 1992, 100 (21), 223–251.

Barro, Robert J and Xavier Sala-i-Martin, Economic growth, second ed., MIT Press, 2003.

Brock, William A. and Stephen N. Durlauf, "Growth Empirics and Reality," The World Bank Economic Review, 2001, 15 (2), 229–272.

Campos, Julia, Neil R. Ericsson, and David Hendry, “Editors Introduction,” in Julia Campos, Neil R. Ericsson, and David Hendry, eds., General to Specific Modelling, Cheltenham: Edward Elgar Publishing, 2005.

Caselli, Francesco, Gerardo Esquivel, and Fernando Lefort, “Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics,” Journal of Economic Growth, 1996, 1 (3), 363–89.

Ciccone, Antonio and Robert Hall, “Productivity and the Density of Economic Activity,” The American Economic Review, March 1996, 86 (1), 54–70.

Congressional Budget Office, “The Economics of Climate Change: A Primer,” http://www.cbo.gov April 2003.

Diewert, W Erwin, “Superlative Index Numbers and Consistency in Aggregation,” Econometrica, 1978, 46 (4), 883–900.

Doornik, Jurgen A., Object-Oriented Matrix Programming Using Ox, third ed., London: Timberlake Consultants Press and Oxford: www.nuff.ox.ac.uk/Users/Doornik, 2002.

Dowrick, Steve and John Quiggin, “International Comparisons of Living Standards and Tastes: A Revealed-Preference Analysis,” American Economic Review, 1994, 84 (1), 332–41.

Dowrick, Steve and John Quiggin, “True Measures of GDP and Convergence,” American Economic Review, 1997, 87 (1), 41–64.

Durlauf, Steven N., Paul A. Johnson, and Jonathan R. W. Temple, “Growth Econometrics,” in Philippe Aghion and Steven N. Durlauf, eds., Handbook of Economic Growth, North-Holland, 2005.

Easterly, William, “The Middle Class Consensus and Economic Development,” Journal of Economic Growth, 2001, 6 (4), 317–35.

Eurostat, “PPP Methodological Manual,” 2005. draft.

Fern´andez, Carmen, Eduardo Ley, and Mark F.J. Steel, “Model Uncertainty in Cross- Country Growth Regressions,” Journal of Applied Econometrics, September-October 2001, 16 (5), 563–76.

Frankel, Jeffrey A., "Determinants of Long Term Growth," 1997. Asia-Pacific Economic Cooperation Economic Advisers, Background Paper for the Morning Session of the Meeting of the Asia-Pacific Economic Cooperation Economic Advisers, Vancouver, Canada, November 20, 1997.

Grier, Kevin B. and Gordon Tullock, "An Empirical Analysis of Cross-National Economic Growth," Journal of Monetary Economics, September 1989, 24 (2), 259–76.

Groningen Growth and Development Centre and The Conference Board, "Total Economy Database," http://www.ggdc.net August 2005.

Gylfason, Thorvaldur, "Natural resources, education, and economic development," European Economic Review, 2001, 45 (4-6), 847–859.

Hall, Robert E. and Charles I. Jones, "Why Do Some Countries Produce So Much More Output Per Worker Than Others?," Quarterly Journal of Economics, February 1999, 114 (1), 83–116.

Hendry, David F. and Hans-Martin Krolzig, Automatic Econometric Model Selection using PcGets, London: Timberlake Consultants Press, 2001.

Hendry, David F. and Hans-Martin Krolzig, "We Ran One Regression," Oxford Bulletin of Economics and Statistics, 2004, 66, 799–810.

Hendry, David F. and Hans-Martin Krolzig, "The Properties of Automatic Gets Modelling," Economic Journal, 2005, 115, C32–C61.

Heston, Alan, "National Accounts: A brief review of some problems in using national accounts data in level of output comparisons and growth studies," Journal of Development Economics, 1994, 44, 29–52.

Heston, Alan, Robert Summers, and Bettina Aten, "Penn World Table version 6.0," Data Web site: http://pwt.econ.upenn.edu/. December 2001. Center for International Comparisons at the University of Pennsylvania (CICUP).

Heston, Alan, Robert Summers, and Bettina Aten, "Penn World Table version 6.1," Data Web site: http://pwt.econ.upenn.edu/. October 2002. Center for International Comparisons at the University of Pennsylvania (CICUP).

Hoeting, Jennifer A., David Madigan, Adrian E. Raftery, and Chris T. Volinsky, "Bayesian Model Averaging: A Tutorial," Statistical Science, November 1999, 14 (4), 382–417.

International Monetary Fund, "World Economic Outlook," Technical Report 2000. May.

Kormendi, Roger C. and Philip Meguire, "Macroeconomic Determinants of Growth: Cross- Country Evidence," Journal of Monetary Economics, September 1985, 16 (2), 141–63.

Kravis, Irving B., Alan Heston, and Robert Summers, World Product and Income: International Comparisons of Real Gross Products, Baltimore: The Johns Hopkins University Press, 1982.

Kronenberg, Tobias, "The curse of natural resources in the transition economies," The Economics of Transition, 2004, 12 (3), 399–426.

Leamer, Edward E., Specification Searches, New York: John Wiley and Sons, 1978.

Levine, Ross E. and David Renelt, "A Sensitivity Analysis of Cross-Country Growth Regressions," American Economic Review, September 1992, 82 (4), 942–63.

Limongi, Fernando and Adam Przeworski, "Political Regimes And Economic Growth," Journal Of Economic Perspectives, 1993, 7 (3).

Maddison, Angus, The World Economy in the 20th Century, Paris: OECD Development Centre, 1989.

Maddison, Angus and Bart van Ark, "The International Comparison of Real Product and Productivity," Research Memorandum, no. 567 (GD-6), Groningen Growth and Development Centre and The Conference Board 1994.

Maddison, Angus and Bart van Ark, "The International Comparison of Real Product and Productivity," in Angus Maddison, D.S. Prasada Rao, and W.F. Shepherd, eds., The Asian Economies in the Twentieth Century, Edwar Elgar, Cheltenham, 2002, pp. 5–26.

New Zealand Treasury, "Towards Higher Living Standards for New Zealanders: Briefing to the Incoming Government," http://www.treasury.govt.nz/briefings/ 1999.

Papaioannou, Elias and Gregorios Siourounis, "Democratization and Growth," November 2005.

Rodriguez, Francisco and Dani Rodrik, "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence," in Ben S. Bernanke and Kenneth Rogoff, eds., NBER Macroeconomics Annual 2000, Cambridge, MA: MIT Press, 2001.

Sachs, Jeffrey D., "Can Extreme Poverty Be Eliminated?," Scientific American, September 2005.

Sachs, Jeffrey D. and Andrew M. Warner, "Economic Reform and the Process of Economic Integration," Brookings Papers on Economic Activity, 1995, (1), 1–95.

Sala-i-Martin, Xavier, "I Just Ran 2 Million Regressions," The American Economic Review, May 1997, 87 (2), 178–83. Papers and Proceedings.

Sala-i-Martin, Xavier, "I Just Ran Four Million Regressions," November 1997. National Bureau of Economic Research (Cambridge, MA) Working Paper No. 6252,.

Sala-i-Martin, Xavier, Gernot Doppelhofer, and Ronald I. Miller, "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach," The American Economic Review, September 2004, 94 (4), 813–835.

Salazar-Xirinachs, Jose M., "Latin American Trade Policies in 2002 and Beyond: Diagnosis and Prognosis," http://www.sice.oas/org 2001. OAS - Organization of American States, Trade and Information Division.

Sergeev, Sergey, "Measures of the similarity of the countrys price structures and their practical application," Working Paper No. 9, UN ECE, Geneva, Consultation on the ECP, 12-14 November 2001.

Summers, Robert and Alan Heston, "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988," Quarterly Journal of Economics, May 1991, pp. 327–368.

UN Economic Commission for Africa, "Transforming Africa's Economies, Economic Report on Africa 2000," 2001. http://www.uneca.org.

United Nations, "Handbook of The International Comparison Programme," Department Of Economic And Social Development, Statistical Division, Studies in Methods, Series F, No. 62 1992.

van Ark, Bart and Marcel P. Timmer, "PPPs and International Productivity Comparisons: Bottlenecks and New Directions," Joint World Bank-OECD Seminar on PPPs, 30 January - 2 February, Washington D.C. 2001.

Warner, Andrew M., "Once More into the Breach: Economic Growth and Integration," Center for Global Development Working Paper Number 34 December 2003.

Tables

Table Ia: List of potential determinants of Long-term growth used in the empirical analysis

Rank	Variable	Description and Source
	Average growth rate of GDP per capita in 1960-1996	Growth of GDP per capita at purchasing power parities between 1960 and 1996. From Heston et al. (2002).
1	GDP in 1960 (log)	Logarithm of GDP per capita in 1960. From Heston et al. (2002).
2	Primary Schooling in 1960	Enrollment rate in primary education in 1960. Barro and Jong-Wha Lee (1993).
3	Investment Price	Average investment price level between 1960 and 1964 on purchasing power parity basis. From Heston et al. (2001).
4	East Asian Dummy	Dummy for East Asian countries.
5	Population Density Coastal in 1960s	Coastal (within 100 km of coastline) population per coastal area in 1965. From Gallup et al. (2001).
6	Population Density 1960	Population per area in 1960. Barro and Lee (1993).
7	Fraction of Tropical Area	Proportion of country's land area within geographical tropics. From John L. Gallup et al. (2001).
8	Air Distance to Big Cities	Logarithm of minimal distance (in km) from New York, Rotterdam, or Tokyo. From Gallup et al. (2001).
9	Fraction GDP in Mining	Fraction of GDP in mining. From Robert E. Hall and Charles I. Jones (1999).
10	Life Expectancy in 1960	Life expectancy in 1960. Barro and Lee (1993).
11	Fraction Muslim	Fraction of population Muslim in 1960. Barro (1999).
12	Political Rights	Political rights index. From Barro (1999).
13	Primary Exports 1970	Fraction of primary exports in total exports in 1970. From Sachs and Warner (1997).
14	African Dummy	Dummy for Sub-Saharan African countries.
15	Fraction Confucius	Fraction of population Confucian. Barro (1999).
16	Fertility in 1960s	Fertility in 1960's. Barro and Sala-i-Martin (1995).
17	Fraction Population In Tropics	Proportion of country's population living in geographical tropics. From Gallup et al. (2001).
18	Hydrocarbon Deposits in 1993	Log of hydrocarbon deposits in 1993. From Gallup et al. (2001).
19	Fraction Buddhist	Fraction of population Buddhist in 1960. Barro (1999).
20	Timing of Independence	Timing of national independence measure: 0 if before 1914; 1 if between 1914 and 1945; 2 if between 1946 and 1989; and 3 if after 1989. From Gallup et al. (2001).
21	Colony Dummy	Dummy for former colony. Barro (1999).
22	Landlocked Country Dummy	Dummy for landlocked countries.
23	Latin American Dummy	Dummy for Latin American countries.

24	Openess measure 1965-74	Ratio of exports plus imports to GDP, averaged over 1965 to 1974. This variable was provided by Robert Barro.
25	Years Open 1950-94	Number of years economy has been open between 1950 and 1994. From Jeffrey D. Sachs and Andrew M. Warner (1995).
26	Gov. Consumption Share 1960s	Share of expenditures on government consumption to GDP in 1961. Barro and Lee (1993).
27	Fraction Population Over 65	Fraction of population older than 65 years in 1960. Barro and Lee (1993)
28	Real Exchange Rate Distortions	Real exchange rate distortions. Levine and Renelt (1992).
29	Fraction of Land Area Near Navigable Water	Proportion of country's land area within 100 km of ocean or ocean-navigable river. From Gallup et al. (2001).
30	Public Investment Share	Average share of expenditures on public investment as fraction of GDP between 1960 and 1965. Barro and Lee (1993).
31	Fraction Speaking Foreign Language	Fraction of population speaking foreign language. Hall and Jones (1999).
32	Government Share of GDP in 1960s	Average share government spending to GDP between 1960-1964. From Heston et al. (2001).
33	Outward Orientation	Measure of outward orientation. Levine and Renelt (1992).
34	Socialist Dummy	Dummy for countries under Socialist rule for considerable time during 1950 to 1995. From Gallup et al. (2001).
35	Average Inflation 1960-90	Average inflation rate between 1960 and 1990. Levine and Renelt (1992).
36	Absolute Latitude	Absolute latitude. Barro (1999).
37	European Dummy	Dummy for European economies.
38	Religion Measure	Religion measure. Barro (1999).
39	Fraction Population Less than 15	Fraction of population younger than 15 years in 1960. Barro and Lee (1993).
40	Ethnolinguistic Fractionalization	Average of five different indices of ethnolinguistic fractionalization which is the probability of two random people in a country not speaking the same language. From William Easterly and Ross Levine (1997).
41	Spanish Colony	Dummy variable for former Spanish colonies. Barro (1999).
42	Revolutions and Coups	Number of revolutions and military coups. Barro and Lee (1993).
43	Tropical Climate Zone	Fraction tropical climate zone. From Gallup et al. (2001).
44	Population Growth Rate 1960-90	Average growth rate of population between 1960 and 1990. Barro and Lee (1993).
45	Malaria Prevalence in 1960s	Index of malaria prevalence in 1966. From Gallup et al. (2001).
46	Size of Economy	Logarithm of aggregate GDP in 1960.
47	British Colony Dummy	Dummy for former British colony after 1776. Barro (1999).
48	Defense Spending Share	Average share public expenditures on defense as fraction of GDP between 1960 and 1965. Barro and Lee (1993).

49	Terms of Trade Growth in 1960s	Growth of terms of trade in the 1960's. Barro and Lee (1993).
50	Civil Liberties	Index of civil liberties index in 1972. Barro (1999).
51	Fraction Catholic	Fraction of population Catholic in 1960. Barro (1999).
52	Fraction Hindus	Fraction of the population Hindu in 1960. Barro (1999).
53	Land Area	Area in km2. Barro and Lee (1993).
54	Nominal Govertnment GDP Share 1960s	Average share of nominal government spending to nominal GDP between 1960 and 1964. Calculated from Heston et al. (2001).
55	Fraction Protestants	Fraction of population Protestant in 1960. Barro (1999).
56	Higher Education 1960	Enrollment rates in higher education. Barro and Lee (1993).
57	Interior Density	Interior (more than 100 km from coastline) population per interior area in 1965. From Gallup et al. (2001).
58	Population in 1960	Population in 1960. Barro (1999).
59	Public Education Spending Share in GDP in 1960s	Average share public expenditures on education as fraction of GDP between 1960 and 1965. Barro and Lee (1993).
60	Terms of Trade Ranking	Barro (1999).
61	Oil Producing Country Dummy	Dummy for oil-producing country. Barro (1999).
62	English Speahing Population	Fraction of population speaking English. From Hall and Jones (1999).
63	War Particpation 1960-90	Indicator for countries that participated in external war between 1960 and 1990. Barro and Lee (1993).
64	Capitalism	Capitalism index. From Hall and Jones (1999).
65	Fraction Othodox	Fraction of population Orthodox in 1960. Barro (1999).
66	Square of Inflation 1960-90	Square of average inflation rate between 1960 and 1990.
67	Fraction Spent in War 1960-90	Fraction of time spent in war between 1960 and 1990. Barro and Lee (1993).

Note: Variables, definitions, and sources are taken from Sala-i-Martin, Doppelhofer, and Miller (2004), except for the income data, which are updatet to PWT6.1 values. Ordered according to posterior inclusion probability.

Table Ib: List of potential determinants of Long-term growth updated for use in the analysis for 1975-1996

Variable	Description and source	Mean	Standard deviation
Average growth rate of GDP per capita in 1975-1996	Growth of GDP per capita at purchasing power parities between 1975 and 1996. From Heston et al. (2002).	0.0127	0.0202
GDP in 1975 (log)	Logarithm of GDP per capita in 1975. From PWT6.1	8.1756	0.9950
Primary Schooling in 1975	Enrollment rate in primary education in 1975. Barro and Jong-Wha Lee (1993).	0.8379	0.2392
Investment Price	Average investment price level between 1970 and 1974 on purchasing power parity basis. From PWT6.1	78.6897	39.2678
Life Expectancy in 1975	Life expectancy in 1975. Barro and Lee (1993).	59.0151	11.1551
Gov. Consumption Share 1970s	Share of expenditures on government consumption to GDP in 1970-74. Barro and Lee (1993).	0.1579	0.0742
Population Density 1975	Population per area in 1975. Population: PWT6.1, area: Barro and Lee (1993).	164.9146	612.5629
Government Share of GDP in 1970s	Average share government spending to GDP between 1970-1974. From PWT6.1	19.4246	11.8637
Higher Education 1975	Enrollment rates in higher education in 1970-74. Barro and Lee (1993).	0.1081	0.1103
Fraction Population Less than 15	Fraction of population younger than 15 years in 1975. Barro and Lee (1993).	0.3829	0.0886
Nominal Govertnment GDP Share 1970s	Average share of nominal government spending to nominal GDP between 1970 and 1974. Calculated from PWT6.1	17.9384	10.2294
Fraction Population Over 65	Fraction of population older than 65 years in 1975. Barro and Lee (1993)	0.0553	0.0376
Defense Spending Share	Average share public expenditures on defense as fraction of GDP between 1970 and 1974. Barro and Lee (1993).	0.0321	0.0468
Population in 1975	Population in 1975. From PWT6.1	27286.1838	72232.1415
Public Education Spending Share in GDP in 1970s	Average share public expenditures on education as fraction of GDP between 1970 and 1975. Barro and Lee (1993).	0.0338	0.0140
Size of Economy	Logarithm of aggregate GDP in 1975. From PWT6.1	17.2593	1.8466

Note: Variables defined analogously as in Sala-i-Martin, Doppelhofer, and Miller (2004), and taken from the same sources, but corresponding to the 1975-1996 sample. Appendix B reports the countries in this sample.

Table Ic: List of income variables taken from the WDI and GGDC-TCB datasets

Variable	Description and source	Mean	Standard deviation
Average growth rate of GDP per capita in 1975-1996, WDI	Growth of GDP per capita at constant local currency units between 1975 and 1996. From World Bank, World Development Indicators database.	0.0123	0.0202
GDP in 1975 (log), WDI	GDP per capita in 1975, PPP (current international $). From World Bank, World Development Indicators database.	7.3105	1.0252
Average growth rate of GDP per capita in 1960-1996, GK	Growth of GDP per capita, in 1990 US$ (converted at Geary Khamis PPPs) between 1960 and 1996. From GGDC-TCB Total Economy Database.	0.0240	0.0172
GDP in 1960 (log), GK	GDP per capita in 1960, in 1990 US$ (converted at Geary Khamis PPPs). From GGDC-TCB Total Economy Database.	7.9201	0.8886

Note: Appendix B reports the samples on which the means and standard deviations were calculated

Table II: Determinants of long-term growth using BACE. Results with PWT6.1 and PWT6.0 income data compared.

		Results with PWT6.1 income data		PWT6.1-6.0 comparisons		
		InclProb	CondMean	InclProb MAX/MIN	CondMean [MAX-MIN] /MIN[ABS]	UncondMean [MAX-MIN] /MIN[ABS]
1	GDP in 1960 (log)	1.0000	-1.5E-02	1.72	0.78	2.07
2	Primary Schooling in 1960	0.9871	3.5E-02	1.60	0.43	1.30
3	Investment Price	0.9806	-8.5E-05	1.21	0.02	0.18
4	East Asian Dummy	0.7468	1.5E-02	1.14	0.42	0.25
5	Population Density Coastal in 1960s	0.7457	9.5E-06	2.46	0.14	1.80
6	Population Density 1960	0.6834	2.0E-05	8.02	0.58	11.63
7	Fraction of Tropical Area	0.6581	-1.5E-02	1.72	0.16	0.99
8	**Air Distance to Big Cities**	0.4221	-1.5E-06	12.77	0.97	24.10
9	Fraction GDP in Mining	0.2877	5.0E-02	1.96	0.13	1.22
10	Life Expectancy in 1960	0.2577	6.4E-04	1.05	0.26	0.20
11	**Fraction Muslim**	0.2364	1.6E-02	3.49	0.47	4.14
12	**Political Rights**	0.2362	-2.2E-03	6.17	0.37	7.43
13	**Primary Exports 1970**	0.2202	-1.7E-02	3.16	0.32	3.17
14	African Dummy	0.2055	-1.2E-02	1.03	0.22	0.18
15	Fraction Confucius	0.1707	4.8E-02	2.21	0.25	1.76
16	**Fertility in 1960s**	0.1550	-1.7E-02	4.19	0.97	7.25
17	Fraction Population In Tropics	0.1456	-1.4E-02	2.13	0.15	1.46
18	**Hydrocarbon Deposits in 1993**	0.1108	6.2E-04	4.00	0.94	6.77
19	Fraction Buddhist	0.1098	1.7E-02	2.43	0.52	2.69
20	**Timing of Independence**	0.0962	-3.8E-03	5.55	41.02	223.07
	AVERAGE			3.40	2.46	15.08
	MEDIAN			2.32	0.40	1.93
21	Colony Dummy	0.0799	-8.9E-03	1.25	0.11	0.39
22	Landlocked Country Dummy	0.0775	-5.8E-03	3.50	1.57	8.00
23	**Latin American Dummy**	0.0725	-9.1E-03	1.65	0.25	1.06
24	**Openess measure 1965-74**	0.0553	8.1E-03	1.81	0.20	1.17
25	**Years Open 1950-94**	0.0529	8.3E-03	2.35	0.55	2.63
26	**Gov. Consumption Share 1960s**	0.0470	-3.0E-02	3.70	0.65	5.11
27	Fraction Population Over 65	0.0450	1.2E-01	1.78	1.13	2.78
28	**Real Exchange Rate Distortions**	0.0434	-4.6E-05	1.95	0.80	2.51
29	Fraction of Land Area Near Navigable Water	0.0429	6.6E-03	1.99	3.74	6.44
30	**Public Investment Share**	0.0401	-5.0E-02	2.42	0.70	3.10
31	Fraction Speaking Foreign Language	0.0369	4.5E-03	1.53	0.44	1.20
32	Government Share of GDP in 1960s	0.0345	-2.4E-02	2.23	0.40	2.12
33	Outward Orientation	0.0319	-2.7E-03	1.63	0.24	1.02
34	Socialist Dummy	0.0307	4.7E-03	1.73	0.23	1.12
35	Average Inflation 1960-90	0.0292	-7.8E-05	1.60	0.25	1.01
36	Absolute Latitude	0.0291	1.8E-04	1.25	0.22	0.02
37	European Dummy	0.0286	5.8E-03	1.03	13.44	12.96

38	Religion Measure	0.0280	-6.0E-03	1.53	0.61	1.46
39	Fraction Population Less than 15	0.0261	-2.5E-02	1.09	2.23	2.34
40	**Ethnolinguistic Fractionalization**	0.0250	-4.7E-03	4.29	1.66	10.41
41	**Spanish Colony**	0.0244	-3.4E-03	5.07	2.13	14.86
42	Revolutions and Coups	0.0243	-5.5E-03	1.58	0.58	1.49
43	Tropical Climate Zone	0.0226	4.8E-03	1.06	2.10	2.16
44	Population Growth Rate 1960-90	0.0221	-1.8E-01	1.24	20.19	25.18
45	**Malaria Prevalence in 1960s**	0.0220	6.0E-04	5.30	23.10	118.12
46	Size of Economy	0.0219	8.6E-04	1.01	3.06	3.04
47	British Colony Dummy	0.0214	1.9E-03	1.59	1.42	2.84
48	Defense Spending Share	0.0213	3.1E-02	1.80	2.20	4.77
49	Terms of Trade Growth in 1960s	0.0208	-2.2E-02	1.26	2.53	2.93
50	Civil Liberties	0.0199	-7.3E-04	3.19	13.06	43.80
51	Fraction Catholic	0.0193	-2.4E-03	1.45	1.45	2.55
52	Fraction Hindus	0.0191	-1.5E-03	1.96	11.57	21.70
53	Land Area	0.0176	3.4E-10	1.07	0.80	0.93
54	Nominal Govertnment GDP Share 1960s	0.0171	-1.1E-02	2.61	2.97	9.36
55	Fraction Protestants	0.0163	-1.6E-03	2.48	5.18	14.31
56	Higher Education 1960	0.0160	3.1E-03	3.05	21.72	64.29
57	Interior Density	0.0159	-7.2E-06	1.10	1.38	1.63
58	Population in 1960	0.0156	-1.2E-09	1.96	27.07	52.06
59	Public Education Spending Share in GDP in 1960s	0.0154	6.6E-02	1.29	0.75	1.25
60	Terms of Trade Ranking	0.0154	2.1E-03	1.18	3.34	3.76
61	Oil Producing Country Dummy	0.0153	7.2E-04	1.31	5.54	7.55
62	English Speahing Population	0.0152	8.5E-04	1.44	6.55	9.01
63	War Particpation 1960-90	0.0150	-1.2E-03	1.07	2.03	2.04
64	Capitalism	0.0150	3.5E-05	1.12	2.43	2.28
65	Fraction Othodox	0.0148	-3.3E-03	1.09	2.96	3.15
66	Square of Inflation 1960-90	0.0148	-1.3E-08	1.20	44.44	53.65
67	Fraction Spent in War 1960-90	0.0131	8.7E-04	1.11	0.77	0.60
	AVERAGE (all)			2.36	4.33	12.50
	MEDIAN (all)			1.72	0.80	2.69

Notes: Variables in bold denote those that emerge as robust growth factors (posterior inclusion probabilities greater or equal 8%) using one version of the PWT but not the other.

Table III. Determinants of Long-term growth using FLS-BMA, results with PWT6.1 and PWT6.0 income data compared

	Results with PWT6.1 income data			PWT6.1-6.0 comparisons		
		InclProb	CondMean	InclProb MAX/MIN	CondMean [MAX-MIN] /MIN[ABS]	UncondMean [MAX-MIN] /MIN[ABS]
1	GDP in 1960 (log)	1.00	-1.53E-02	1.61	0.76	1.84
2	Primary Schooling in 1960	0.99	3.49E-02	1.54	0.43	1.20
3	Investment Price	0.98	-8.46E-05	1.17	0.02	0.15
4	Population Density Coastal in 1960s	0.77	9.47E-06	2.42	0.14	1.75
5	East Asian Dummy	0.74	1.48E-02	1.17	0.41	0.21
6	Population Density 1960	0.72	2.04E-05	7.23	0.53	10.04
7	Fraction of Tropical Area	0.68	-1.51E-02	1.74	0.16	1.02
8	**Air Distance to Big Cities**	0.44	-1.53E-06	11.23	0.95	20.95
9	Fraction GDP in Mining	0.27	4.74E-02	1.42	0.03	0.46
10	**Political Rights**	0.25	-2.16E-03	5.64	0.42	7.02
11	Life Expectancy in 1960	0.24	6.14E-04	1.05	0.30	0.38
12	African Dummy	0.22	-1.21E-02	1.07	0.21	0.29
13	**Primary Exports 1970**	0.22	-1.69E-02	2.83	0.31	2.70
14	Fraction Muslim	0.21	1.45E-02	2.65	0.37	2.63
15	Fraction Confucius	0.17	4.58E-02	2.35	0.28	2.01
16	Fraction Population In Tropics	0.16	-1.34E-02	1.85	0.16	1.15
17	**Fertility in 1960s**	0.15	-1.67E-02	3.31	0.75	4.81
18	**Hydrocarbon Deposits in 1993**	0.13	6.08E-04	3.91	0.80	6.02
19	Fraction Buddhist	0.12	1.73E-02	2.38	0.48	2.53
20	**Landlocked Country Dummy**	0.09	-5.70E-03	3.48	1.20	6.68
21	**Colony Dummy**	0.09	-8.50E-03	1.17	0.06	0.24
22	Latin American Dummy	0.08	-8.67E-03	1.51	0.28	0.94
	Average			2.85	0.41	3.41
	Median			2.10	0.34	1.80
23	Timing of Independence	0.07	-3.21E-03	3.47	164.68	568.72
24	**Openess measure 1965-74**	0.06	8.11E-03	1.84	0.20	1.20
25	**Years Open 1950-94**	0.06	8.26E-03	2.16	0.49	2.23
26	**Gov. Consumption Share 1960s**	0.05	-3.04E-02	3.52	0.61	4.67
27	Fraction of Land Area Near Navigable Water	0.05	6.08E-03	2.07	3.68	6.55
28	Fraction Population Over 65	0.05	1.09E-01	1.54	0.72	1.64
29	**Government Share of GDP in 1960s**	0.05	-2.56E-02	1.87	0.31	1.44
30	**Real Exchange Rate Distortions**	0.05	-4.64E-05	2.05	0.73	2.56
31	Fraction Speaking Foreign Language	0.04	4.36E-03	1.58	0.48	1.34
32	**Public Investment Share**	0.04	-4.61E-02	3.09	0.93	4.97
33	European Dummy	0.04	5.69E-03	1.22	9.45	11.70
34	Outward Orientation	0.04	-2.75E-03	1.66	0.26	1.09
35	Absolute Latitude	0.03	1.66E-04	1.16	0.25	0.07
36	Revolutions and Coups	0.03	-5.71E-03	1.45	0.57	1.28

37	Religion Measure	0.03	-6.54E-03	1.49	0.73	1.58
38	Socialist Dummy	0.03	4.89E-03	1.47	0.30	0.91
39	Population Growth Rate 1960-90	0.03	-1.68E-01	1.39	12.80	18.14
40	Tropical Climate Zone	0.03	4.71E-03	1.22	2.16	2.41
41	Fraction Hindus	0.03	-1.48E-03	1.49	11.51	16.68
42	**Ethnolinguistic Fractionalization**	0.03	-5.05E-03	4.06	1.43	8.86
43	**Malaria Prevalence in 1960s**	0.03	9.17E-04	4.10	14.73	57.30
44	British Colony Dummy	0.03	1.93E-03	1.50	1.42	2.63
45	Size of Economy	0.02	8.82E-04	1.06	3.49	3.36
46	Fraction Population Less than 15	0.02	-1.48E-02	1.30	2.95	3.54
47	Terms of Trade Growth in 1960s	0.02	-2.02E-02	1.20	2.67	3.01
48	Defense Spending Share	0.02	3.56E-02	1.95	1.72	4.30
49	Civil Liberties	0.02	-9.06E-04	3.20	10.44	35.63
50	Spanish Colony	0.02	-2.70E-03	6.36	2.87	23.64
51	Average Inflation 1960-90	0.02	-3.31E-05	1.01	0.76	0.75
52	Fraction Catholic	0.02	-2.50E-03	1.53	1.35	2.61
53	Square of Inflation 1960-90	0.02	-1.47E-07	1.07	2.86	2.62
54	Nominal Govertnment GDP Share 1960s	0.02	-1.02E-02	2.62	3.17	9.90
55	Land Area	0.02	3.50E-10	1.05	0.71	0.64
56	Interior Density	0.02	-6.93E-06	1.09	1.01	1.18
57	Terms of Trade Ranking	0.02	1.63E-03	1.11	3.98	4.31
58	Public Education Spending Share in GDP in 1960s	0.02	6.26E-02	1.24	0.80	1.24
59	Fraction Othodox	0.02	-2.18E-03	1.03	3.69	3.76
60	Capitalism	0.02	3.03E-05	1.03	0.43	0.39
61	English Speahing Population	0.02	9.39E-04	1.46	6.11	8.46
62	Fraction Protestants	0.02	-1.58E-03	2.48	5.17	14.29
63	Oil Producing Country Dummy	0.02	1.02E-03	1.32	3.31	4.70
64	War Particpation 1960-90	0.02	-1.04E-03	1.08	2.08	2.17
65	Population in 1960	0.02	-1.99E-09	2.00	17.40	33.82
66	Higher Education 1960	0.02	2.00E-03	3.05	32.06	95.76
67	Fraction Spent in War 1960-90	0.02	1.16E-03	1.07	0.69	0.58
	Average			2.22	5.18	15.73
	Median			1.54	0.76	2.61

Notes: Variables in bold denote those that emerge as robust growth factors (posterior inclusion probabilities greater or equal 8%) using one version of the PWT but not the other.

Table IV. Determinants of long-term growth using General-to-Specific approach (PcGets). Results with PWT6.1 and PWT6.0 income data compared

RESULTS WITH PWT 6.1 INCOME DATA

	Coeff	StdError	t-Statistic
Fraction Confucius	0.07313	0.01599	4.57
Fraction Population In Tropics	-0.02102	0.00351	-5.99
Fraction Buddhist	0.03385	0.00752	4.50
Investment Price	-0.00009	0.00002	-3.86

RESULTS WITH PWT 6.0 INCOME DATA

	Coeff	StdError	t-Statistics
Fraction Buddhist	0.02577	0.00777	3.31
Investment Price	-0.00011	0.00002	-4.50
Primary Schooling in 1960	0.03471	0.00623	5.58
Primary Exports 1970	-0.02594	0.00532	-4.87
GDP in 1960 (log)	-0.00987	0.00211	-4.67

Table V: Applying BACE to growth perturbations with correlations between 97.5% and 98.5% with PWT6.1 growth

		InclProb MEDIAN	InclProb 90PCT /10PCT	CondMean [90PCT-10PCT] /MIN[ABS]	UncondMean [90PCT-10PCT] /MIN[ABS]
1	GDP in 1960 (log)	1.00	1.01	0.27	0.28
2	Primary Schooling in 1960	0.95	1.38	0.33	0.83
3	Investment Price	0.98	1.40	0.37	0.85
4	Population Density Coastal in 1960s	0.40	5.97	0.50	7.34
5	Fraction of Tropical Area	0.40	7.43	0.51	10.30
6	Population Density 1960	0.41	6.20	0.39	7.44
7	Air Distance to Big Cities	0.11	5.84	0.46	5.47
8	Fraction Confucius	0.21	8.94	0.96	14.50
9	East Asian Dummy	0.60	3.37	0.50	3.50
10	Fraction Muslim	0.26	10.59	0.78	16.57
11	Political Rights	0.07	11.34	1.10	17.19
12	Fertility in 1960s	0.13	8.96	1.03	15.72
13	Fraction GDP in Mining	0.34	9.21	0.70	12.26
14	African Dummy	0.18	9.93	0.88	18.04
15	Life Expectancy in 1960	0.27	6.39	0.91	9.67
16	Primary Exports 1970	0.29	8.20	0.48	11.43
17	Revolutions and Coups	0.03	3.30	1.60	7.36
18	Fraction Population In Tropics	0.10	12.64	0.98	22.69
19	Real Exchange Rate Distortions	0.04	7.18	2.28	22.89
20	Landlocked Country Dummy	0.04	10.15	1.63	24.31
21	Fraction Buddhist	0.11	9.96	0.87	15.47
22	Years Open 1950-94	0.07	9.55	1.68	23.93
23	Hydrocarbon Deposits in 1993	0.08	9.57	1.39	18.06
24	Latin American Dummy	0.07	12.91	3.19	42.25
25	Timing of Independence	0.06	7.99	2.86	29.10
26	Openess measure 1965-74	0.07	3.58	0.67	5.18
27	Colony Dummy	0.05	5.57	1.36	9.20
28	Public Investment Share	0.05	4.05	1.39	8.19
29	Gov. Consumption Share 1960s	0.07	7.73	1.50	17.87
30	Fraction Population Over 65	0.05	5.98	2.06	16.58
31	Government Share of GDP in 1960s	0.04	4.80	2.43	16.69
32	Absolute Latitude	0.04	4.72	2.03	11.89
33	Fraction Speaking Foreign Language	0.05	3.48	1.28	6.79
34	Ethnolinguistic Fractionalization	0.03	3.63	3.20	12.99
35	Outward Orientation	0.03	3.59	9.11	26.94
36	European Dummy	0.04	4.26	4.50	20.36
37	Religion Measure	0.02	2.09	6.78	13.53
38	Socialist Dummy	0.03	4.17	9.25	43.45
39	British Colony Dummy	0.02	3.72	19.07	52.82

40	Malaria Prevalence in 1960s	0.03	2.91	2.68	3.60
41	Population in 1960	0.02	1.81	2.02	2.08
42	Nominal Govertnment GDP Share 1960s	0.02	2.30	2.79	4.11
43	Tropical Climate Zone	0.02	1.93	8.86	17.44
44	Spanish Colony	0.03	3.52	25.11	84.98
45	Population Growth Rate 1960-90	0.03	2.74	91.07	289.10
46	Fraction Catholic	0.02	2.22	513.81	804.51
47	Fraction Population Less than 15	0.03	2.32	6.81	11.82
48	Size of Economy	0.02	2.21	8.80	18.16
49	War Particpation 1960-90	0.02	1.79	17.41	29.08
50	Defense Spending Share	0.02	2.52	9.39	18.94
51	Oil Producing Country Dummy	0.02	1.60	2.72	2.81
52	Fraction Spent in War 1960-90	0.02	1.61	3.62	4.66
53	Fraction Hindus	0.02	2.54	2.95	4.00
54	Capitalism	0.02	1.92	6.40	10.51
55	Fraction of Land Area Near Navigable Water	0.02	2.25	10.83	20.56
56	Terms of Trade Growth in 1960s	0.02	1.65	7.77	10.25
57	Public Education Spending Share in GDP in 1960s	0.02	1.89	94.12	160.50
58	Land Area	0.02	1.70	18.58	31.51
59	Average Inflation 1960-90	0.02	1.98	9.90	19.38
60	Civil Liberties	0.02	1.82	2.55	2.66
61	Fraction Othodox	0.02	1.38	4.69	6.27
62	Interior Density	0.02	1.63	2.06	3.90
63	Higher Education 1960	0.02	1.52	3.87	7.38
64	Terms of Trade Ranking	0.02	1.43	2.29	2.42
65	Fraction Protestants	0.02	1.63	22.11	35.57
66	English Speahing Population	0.02	1.50	3.42	4.11
67	Square of Inflation 1960-90	0.02	1.25	2.32	2.55
	AVERAGE (all)		4.57	14.63	33.32
	MEDIAN (all)		3.48	2.32	12.99
	AVERAGE (always robust)		2.44	0.37	2.32
	MEDIAN (always robust)		1.39	0.35	0.84
	AVERAGE (robust once or more)		6.14	2.35	14.99
	MEDIAN (robust once or more)		5.90	1.37	13.26

Notes: Variables in small box are robust growth factors (posterior inclusion probability greater or equal 8%) for all 30 growth perturbations. Variables in large box are robust growth factors for at least one growth perturbation.

Table VI: Applying BACE to growth perturbations with correlations between 99% and 99.5% with PWT6.1 growth

		InclProb MEDIAN	InclProb 90PCT /10PCT	CondMean [90PCT-10PCT] /MIN[ABS]	UncondMean [90PCT-10PCT] /MIN[ABS]
1	GDP in 1960 (log)	1.00	1.00	0.16	0.16
2	Primary Schooling in 1960	0.98	1.15	0.27	0.45
3	Investment Price	0.96	1.13	0.23	0.35
4	East Asian Dummy	0.73	1.69	0.18	0.93
5	Population Density 1960	0.53	2.14	0.18	1.37
6	Fraction of Tropical Area	0.56	2.75	0.23	2.33
7	Population Density Coastal in 1960s	0.72	2.04	0.24	1.38
8	African Dummy	0.22	3.63	0.41	4.31
9	Fraction GDP in Mining	0.38	3.57	0.23	3.19
10	Primary Exports 1970	0.23	3.10	0.22	2.41
11	Life Expectancy in 1960	0.24	3.13	0.48	3.62
12	Fraction Muslim	0.21	5.29	0.48	6.49
13	Political Rights	0.17	7.06	0.38	9.12
14	Air Distance to Big Cities	0.24	4.67	0.27	5.28
15	Fraction Confucius	0.21	5.94	0.51	7.59
16	Fertility in 1960s	0.15	4.76	0.49	5.56
17	Fraction Population In Tropics	0.16	3.80	0.32	3.13
18	Fraction Buddhist	0.09	4.42	0.46	5.41
19	Latin American Dummy	0.09	4.97	0.57	6.35
20	Hydrocarbon Deposits in 1993	0.09	4.70	0.58	5.54
21	Timing of Independence	0.07	5.94	0.97	9.59
22	Public Investment Share	0.05	3.33	0.63	4.05
23	Colony Dummy	0.08	4.15	0.74	6.30
24	Landlocked Country Dummy	0.06	4.21	0.59	5.51
25	Gov. Consumption Share 1960s	0.07	3.39	0.76	5.16
26	Years Open 1950-94	0.07	2.38	0.44	2.14
27	Government Share of GDP in 1960s	0.04	2.88	0.81	3.80
28	Openess measure 1965-74	0.08	2.12	0.50	2.15
29	Fraction Population Over 65	0.05	2.69	0.76	3.34
30	Real Exchange Rate Distortions	0.04	4.46	1.03	8.72
31	Fraction Speaking Foreign Language	0.04	2.60	0.63	3.18
32	Socialist Dummy	0.03	2.82	1.36	5.80
33	Absolute Latitude	0.04	1.83	0.72	2.10
34	Ethnolinguistic Fractionalization	0.03	2.69	1.64	4.96
35	European Dummy	0.04	2.40	1.10	4.39
36	Fraction of Land Area Near Navigable Water	0.03	2.40	1.27	4.43
37	Outward Orientation	0.03	2.88	1.67	6.64
38	Revolutions and Coups	0.03	2.97	1.00	4.76
39	Religion Measure	0.02	2.11	1.44	4.14
40	British Colony Dummy	0.02	1.66	1.94	3.96
41	Size of Economy	0.02	2.34	1.71	4.64
42	Fraction Population Less than 15	0.03	1.83	6.36	9.92

43	Tropical Climate Zone	0.02	2.36	1.69	5.17
44	Defense Spending Share	0.02	1.64	2.35	4.48
45	Malaria Prevalence in 1960s	0.02	1.63	2.13	2.25
46	Spanish Colony	0.02	1.90	3.30	6.69
47	Population Growth Rate 1960-90	0.03	1.64	1.99	2.72
48	Terms of Trade Growth in 1960s	0.02	1.54	7.05	9.40
49	Average Inflation 1960-90	0.02	1.70	3.38	4.62
50	Fraction Hindus	0.02	1.77	2.93	3.27
51	Land Area	0.02	1.53	1.57	3.08
52	Fraction Catholic	0.02	1.39	1.97	3.14
53	Fraction Protestants	0.02	1.39	8.15	11.66
54	Capitalism	0.02	1.31	5.20	5.71
55	War Particpation 1960-90	0.02	1.59	5.85	21.91
56	Higher Education 1960	0.02	1.37	5.09	6.12
57	Civil Liberties	0.02	1.54	5.51	5.72
58	Public Education Spending Share in GDP in 1960s	0.02	1.50	5.96	9.76
59	Nominal Govertnment GDP Share 1960s	0.02	1.41	10.12	14.86
60	Oil Producing Country Dummy	0.02	1.34	6.37	7.20
61	Terms of Trade Ranking	0.02	1.24	11.16	13.25
62	Interior Density	0.02	1.29	1.09	2.03
63	Population in 1960	0.02	1.33	2.34	2.36
64	Fraction Spent in War 1960-90	0.01	1.32	5.41	6.54
65	Square of Inflation 1960-90	0.02	1.19	2.32	2.38
66	Fraction Othodox	0.02	1.30	5.33	6.28
67	English Speahing Population	0.02	1.21	2.35	2.31
	AVERAGE (all)		2.57	2.17	5.19
	MEDIAN (all)		2.12	1.09	4.62

Notes: Variables in small box are robust growth factors (posterior inclusion probability greater or equal 8%) for all 30 growth perturbations. Variables in large box are robust growth factors for at least one growth perturbation.

Table VII. Sensitivity of results to growth perturbations in smaller and larger datasets

	67 candidate variables		15 candidate variables(1)	
	Share of always robust(2)	InclProb PCT90/PCT10	Share of always robust(2)	InclProb PCT90/PCT10
Perturbated Growth Data (Experiment I)				
BACE	0.11	4.57	0.46	1.66
FLS-BMA	0.11	4.08	0.43	1.80
General-to-Specific	0.00	-	0.28	-
Perturbated Growth Data (Experiment II - smaller perturbations)				
BACE	0.30	2.57	0.61	1.32
FLS-BMA	0.30	2.44	0.58	1.38
General-to-Specific	0.00	-	0.48	-

Notes:
(1) Average across 50 randomly selected lists of 15 variables
(2) Ratio of the number of variables which are always robust to the number of variables which are robust for at least one growth perturbation. The criterion for robustness in Bayesian approaches is posterior inclusion probability exceeding prior inclusion probability. Prior inclusion probability is 7/67 in the 67 variable dataset and 7/15 in 15 variable datasets. For Gets robustness is defined as inclusion of the variable in the final ('specific') model.

Table VIII: Determinants of long-term growth using BACE. Results with PWT6.1 and World Bank WDI income data compared.

	Results with PWT6.1 income data			PWT6.1-WDI comparisons		
		InclProb	CondMean	InclProb MAX/MIN	CondMean [MAX-MIN] /MIN[ABS]	UncondMean [MAX-MIN] /MIN[ABS]
1	**African Dummy**	0.54	-2.11E-02	17.90	0.70	0.98
2	East Asian Dummy	0.52	2.42E-02	1.91	0.10	1.10
3	Life Expectancy in 1960	0.50	-1.34E-02	2.00	0.17	0.58
4	Fraction GDP in Mining	0.48	6.88E-02	2.30	0.56	2.59
5	**Fraction Confucius**	0.42	7.04E-02	8.47	0.89	15.01
6	Hydrocarbon Deposits in 1993	0.41	1.20E-03	2.43	0.08	1.62
7	Real Exchange Rate Distortions	0.33	-1.31E-04	1.03	0.17	0.14
8	**Fraction Buddhist**	0.29	3.13E-02	14.90	2.35	48.90
9	Nominal Govertnment GDP Share 1960s	0.29	-3.91E-04	1.65	0.27	0.56
10	**Public Investment Share**	0.29	-4.60E-04	5.79	0.72	0.95
11	**Fraction Hindus**	0.27	-1.14E-04	5.45	0.48	0.90
12	**Defense Spending Share**	0.24	9.67E-02	4.75	0.70	7.06
13	Absolute Latitude	0.21	4.11E-04	1.24	0.18	0.47
14	Fraction of Tropical Area	0.20	-1.41E-02	3.61	0.22	0.78
15	**Fraction Population In Tropics**	0.20	-1.94E-02	5.50	0.51	0.91
16	**Investment Price**	0.15	-1.30E-02	6.55	0.84	0.98
17	Openess measure 1965-74	0.10	1.16E-02	2.44	0.01	1.47
18	**Spanish Colony**	0.10	-1.24E-02	5.61	0.97	0.99
19	**British Colony Dummy**	0.10	8.10E-03	1.37	0.36	0.87
20	**Ethnolinguistic Fractionalization**	0.10	-1.75E-02	2.39	0.41	0.75
21	Government Share of GDP in 1960s	0.09	-6.37E-02	2.13	0.19	0.42
22	**Terms of Trade Ranking**	0.09	-2.29E-02	1.80	0.31	0.62
23	**Fraction Muslim**	0.08	1.45E-02	2.25	0.83	3.12
24	Years Open 1950-94	0.08	1.36E-02	3.00	0.08	2.24
25	**Fraction Population Less than 15**	0.08	-9.03E-02	4.57	0.86	0.97
	AVERAGE			4.44	0.52	3.80
	MEDIAN			2.44	0.41	0.97
26	Interior Density	0.06	5.32E-06	1.03	0.15	0.12
27	Population Density 1960	0.06	6.59E-06	1.13	0.12	0.01
28	Public Education Spending Share in GDP in 1960s	0.05	-7.63E-02	2.30	0.50	0.78
29	Population in 1960	0.05	1.48E-01	2.34	5.07	10.52
30	Fraction Protestants	0.05	-1.39E-02	1.62	0.47	0.68
31	Religion Measure	0.05	2.37E-02	1.62	0.74	1.82
32	Fraction Population Over 65	0.04	3.66E-08	1.07	0.16	0.08
33	Fraction Catholic	0.04	-1.11E-02	2.29	1.03	1.01
34	European Dummy	0.04	-1.30E-03	1.97	0.37	0.68
35	Fertility in 1960s	0.04	-1.05E-02	2.31	0.80	0.91
36	Gov. Consumption Share 1960s	0.03	-3.21E-02	1.22	0.13	0.29

37	Average Inflation 1960-90	0.03	-1.56E-04	2.09	0.74	0.87
38	Revolutions and Coups	0.03	-9.65E-03	1.15	0.21	0.31
39	Socialist Dummy	0.03	8.55E-03	1.96	3.57	6.04
40	Fraction Othodox	0.03	-1.96E-02	1.30	0.30	0.46
41	Size of Economy	0.03	-1.53E-03	1.36	0.62	0.72
42	Landlocked Country Dummy	0.03	2.99E-04	1.76	10.06	16.97
43	Air Distance to Big Cities	0.03	7.57E-07	1.62	1.79	3.51
44	**Primary Exports 1970**	0.02	-7.43E-03	1.53	0.59	0.74
45	Political Rights	0.02	1.25E-03	7.86	1.47	1.06
46	Timing of Independence	0.02	-9.05E-04	1.21	0.65	0.71
47	Population Growth Rate 1960-90	0.02	-2.14E-01	1.49	1.46	1.31
48	**Primary Schooling in 1960**	0.02	-7.97E-03	1.44	0.31	0.52
49	**Malaria Prevalence in 1960s**	0.02	-6.39E-05	4.61	0.99	1.00
50	Fraction Speaking Foreign Language	0.02	3.43E-03	4.21	1.07	7.69
51	English Speahing Population	0.02	-6.56E-03	1.38	0.61	0.72
52	Civil Liberties	0.02	-5.73E-03	1.32	2.10	2.45
53	Square of Inflation 1960-90	0.02	-1.24E-06	1.45	0.74	0.82
54	War Particpation 1960-90	0.02	-1.93E-03	1.40	1.17	1.12
55	Colony Dummy	0.02	-5.33E-04	1.22	2.96	2.61
56	Higher Education 1960	0.02	-7.71E-03	1.26	0.52	0.62
57	Land Area	0.02	2.76E-03	1.26	0.13	0.41
58	Outward Orientation	0.02	-2.48E-03	1.62	0.33	0.59
59	Terms of Trade Growth in 1960s	0.02	-6.10E-03	1.20	0.34	0.21
60	Oil Producing Country Dummy	0.02	2.19E-03	1.15	0.88	1.16
61	Fraction Spent in War 1960-90	0.02	-9.53E-03	1.34	0.96	0.97
62	Fraction of Land Area Near Navigable Water	0.02	-2.75E-03	1.04	0.03	0.07
63	Tropical Climate Zone	0.02	2.32E-03	2.20	2.74	7.24
64	Latin American Dummy	0.02	1.37E-10	1.35	4.79	6.83
65	GDP in 1960 (log)	0.02	-1.41E-02	1.21	0.83	0.86
66	Population Density Coastal in 1960s	0.02	8.38E-06	1.08	0.53	0.66
67	Capitalism	0.01	2.68E-04	1.37	2.64	3.97
	AVERAGE			2.78	1.02	2.76
	MEDIAN			1.65	0.61	0.91

Notes: Variables in bold denote those that emerge as robust growth factors (posterior inclusion probabilities greater or equal 8%) using income from one of the datasets but not the other.

Table IX. Determinants of long-term growth using General-to-Specific approach (PcGets). Results with PWT6.1 and WDI income data compared

RESULTS WITH PWT 6.1 INCOME DATA

	Coeff	StdError	t-statistic
Fertility in 1960s	-0.02874	0.00646	-4.45
Nominal Govertnment GDP Share 1970s	-0.00044	0.00013	-3.47
African Dummy	-0.01661	0.00454	-3.66
Terms of Trade Ranking	-0.02491	0.00875	-2.85
Fraction GDP in Mining	0.08117	0.01816	4.47
Fraction Muslim	0.01906	0.00465	4.10
Investment Price	-0.00015	0.00004	-4.22
Life Expectancy in 1975	0.00125	0.00030	4.12
GDP in 1975 (log)	-0.01877	0.00326	-5.77
Fraction Confucius	0.05724	0.01844	3.10

RESULTS WITH WDI INCOME DATA

	Coeff	StdError	t-statistics
Fraction GDP in Mining	0.05908	0.01602	3.69
Fraction Muslim	0.01362	0.00541	2.52
Investment Price	-0.00010	0.00003	-3.46
Life Expectancy in 1975	0.00098	0.00024	4.08
GDP in 1975 (log), WDI	-0.01961	0.00231	-8.48
Fraction Confucius	0.04812	0.01728	2.79
Civil Liberties	0.01415	0.00455	3.11
Population Growth Rate 1960-90	-0.85537	0.18600	-4.60
East Asian Dummy	0.02033	0.00428	4.75
Fraction of Land Area Near Navigable Water	-0.01377	0.00423	-3.26
Timing of Independence	-0.00570	0.00184	-3.11
Openess measure 1965-74	0.01010	0.00374	2.70
Fraction Population In Tropics	-0.01420	0.00515	-2.76
Gov. Consumption Share 1970s	-0.07829	0.01942	-4.03
Defense Spending Share	0.11527	0.03303	3.49

Notes: Variables in bold denote those that emerge as robust growth factors (they are included in the specific model) using income from one of the datasets but not the other.

Table X: Determinants of long-term growth using BACE. Results with PWT6.1 and GGDB-TCN income data compared.

		Results with PWT6.1 income data		PWT6.1-WDI comparisons		
		InclProb	CondMean	InclProb MAX/MIN	CondMean [MAX-MIN] /MIN[ABS]	UncondMean [MAX-MIN] /MIN[ABS]
1	GDP in 1960 (log)	0.99	-1.66E-02	1.43	0.53	1.19
2	**Primary Schooling in 1960**	0.75	3.85E-02	18.00	3.06	72.02
3	**Primary Exports 1970**	0.73	-2.13E-02	16.70	1.80	45.74
4	East Asian Dummy	0.57	1.73E-02	1.70	0.15	0.95
5	African Dummy	0.50	-1.52E-02	1.18	0.49	0.75
6	**Openess measure 1965-74**	0.43	9.41E-03	16.69	2.81	62.63
7	Government Share of GDP in 1960s	0.43	-7.89E-02	2.55	0.48	2.77
8	**Fraction Catholic**	0.36	-1.05E-02	12.84	12.49	148.51
9	Fraction of Tropical Area	0.30	-1.12E-02	2.44	0.27	2.09
10	Life Expectancy in 1960	0.28	8.37E-04	1.17	0.15	0.35
11	Malaria Prevalence in 1960s	0.26	-2.02E-02	1.48	0.05	0.41
12	Ethnolinguistic Fractionalization	0.25	-1.67E-02	2.04	0.44	1.94
13	Years Open 1950-94	0.21	1.06E-02	1.86	0.03	0.91
14	**Land Area**	0.17	1.12E-09	4.35	0.37	4.96
15	**Population Density Coastal in 1960s**	0.15	5.72E-06	5.93	4.57	32.07
16	**Population Growth Rate 1960-90**	0.09	-3.51E-01	3.10	0.79	4.56
17	**English Speahing Population**	0.09	7.75E-03	3.65	3.65	10.69
18	**Interior Density**	0.09	-1.59E-05	3.25	0.66	4.40
19	**Fraction Spent in War 1960-90**	0.09	-1.30E-02	3.21	4.74	17.43
	Average			5.45	1.98	21.81
	Median			3.10	0.53	4.40
20	War Particpation 1960-90	0.08	-5.17E-03	3.52	8.21	26.38
21	Fertility in 1960s	0.07	-8.95E-03	1.19	0.05	0.14
22	**Political Rights**	0.07	1.08E-03	1.22	0.78	1.18
23	Revolutions and Coups	0.07	-8.26E-03	2.75	1.31	5.36
24	European Dummy	0.06	1.56E-02	1.82	4.94	9.84
25	**Fraction Confucius**	0.06	2.35E-02	4.08	0.53	5.24
26	Fraction Muslim	0.06	8.33E-03	1.26	0.21	0.52
27	**Gov. Consumption Share 1960s**	0.06	-3.71E-02	2.28	0.88	3.30
28	Absolute Latitude	0.06	1.90E-04	1.31	0.24	0.62
29	Latin American Dummy	0.06	9.58E-03	1.46	0.70	1.47
30	**Nominal Govertnment GDP Share 1960s**	0.06	-3.54E-02	5.35	0.78	8.50
31	Population Density 1960	0.05	1.18E-05	1.09	2.22	2.12
32	**Investment Price**	0.05	-3.87E-05	11.95	1.64	30.56

33	Fraction GDP in Mining	0.05	3.82E-02	2.18	4.42	10.82
34	Defense Spending Share	0.05	-6.51E-02	1.24	922.92	1146.78
35	Fraction Protestants	0.05	-2.01E-02	1.43	1.99	1.08
36	Oil Producing Country Dummy	0.04	-1.03E-02	2.00	3.43	7.86
37	Fraction Hindus	0.04	-1.31E-02	1.57	0.38	1.17
38	**Fraction Population In Tropics**	0.04	-4.70E-03	2.27	1.87	5.52
39	Spanish Colony	0.04	-4.83E-03	1.51	0.24	0.87
40	Fraction of Land Area Near Navigable Water	0.04	-5.27E-03	2.08	268.83	560.43
41	**Fraction Speaking Foreign Language**	0.04	3.66E-03	7.64	1.07	14.82
42	Capitalism	0.04	-4.32E-04	1.49	2.90	2.27
43	Fraction Buddhist	0.04	7.39E-03	1.29	1.16	1.78
44	Public Education Spending Share in GDP in 1960s	0.04	-8.88E-02	1.30	2.73	1.86
45	Fraction Population Less than 15	0.04	-2.36E-02	1.43	33.35	47.99
46	Size of Economy	0.04	-3.18E-04	1.15	2.28	2.11
47	**Colony Dummy**	0.04	-2.72E-03	24.47	5.35	154.45
48	Socialist Dummy	0.03	4.36E-03	1.03	2.37	2.42
49	Terms of Trade Growth in 1960s	0.03	-3.48E-02	1.53	6.90	11.12
50	Average Inflation 1960-90	0.03	6.73E-05	1.22	0.42	0.74
51	Public Investment Share	0.03	2.85E-02	1.04	2.09	2.05
52	Timing of Independence	0.03	-1.48E-03	1.34	5.68	7.26
53	Real Exchange Rate Distortions	0.03	-2.03E-05	1.44	1.14	2.08
54	Higher Education 1960	0.03	-2.35E-02	1.70	0.17	0.98
55	Fraction Population Over 65	0.03	4.10E-02	1.16	0.98	0.71
56	Religion Measure	0.03	-7.23E-04	1.00	8.62	8.62
57	Air Distance to Big Cities	0.03	1.43E-07	1.56	5.19	7.55
58	Population in 1960	0.03	-3.41E-09	1.30	16.33	12.78
59	Landlocked Country Dummy	0.03	-2.39E-03	1.27	0.83	1.33
60	Terms of Trade Ranking	0.02	-2.51E-03	2.02	5.05	9.20
61	Tropical Climate Zone	0.02	-1.93E-03	1.20	3.55	4.06
62	**British Colony Dummy**	0.02	-7.06E-04	37.98	24.98	911.61
63	Civil Liberties	0.02	4.75E-04	1.46	7.32	10.20
64	Fraction Othodox	0.02	-1.19E-03	1.06	2.69	2.89
65	Square of Inflation 1960-90	0.02	3.36E-07	1.29	0.36	0.75
66	Outward Orientation	0.02	5.16E-04	1.11	0.06	0.04
67	Hydrocarbon Deposits in 1993	0.02	7.47E-05	3.26	8.59	25.71
	Average			3.89	21.14	52.11
	Median			1.56	1.80	4.06

Notes: Variables in bold denote those that emerge as robust growth factors (posterior inclusion probabilities greater or equal 8%) using income from one of the datasets but not the other.

Table XI. Determinants of long-term growth using General-to-Specific approach (PcGets). Results with PWT6.1 and GGDC-TCB income data compared

RESULTS WITH PWT 6.1 INCOME DATA

	Coeff	StdError	t-statistic
Primary Schooling in 1960	0.04106	0.00699	5.87
GDP in 1960 (log)	-0.01729	0.00178	-9.74
Openess measure 1965-74	0.00942	0.00277	3.40
Primary Exports 1970	-0.02792	0.00374	-7.47
Government Share of GDP in 1960s	-0.06206	0.01873	-3.31
African Dummy	-0.01618	0.00359	-4.51

RESULTS WITH GGDC-TCB INCOME DATA

	Coeff	StdError	t-statistics
Primary Exports 1970	-0.02954	0.00507	-5.83
Government Share of GDP in 1960s	-0.19305	0.04247	-4.55
African Dummy	-0.03781	0.00607	-6.23
Fraction Population In Tropics	0.02288	0.00682	3.35
Latin American Dummy	-0.02351	0.00436	-5.39
Gov. Consumption Share 1960s	0.13745	0.04734	2.90
Political Rights	0.00625	0.00091	6.88
Socialist Dummy	-0.02779	0.00508	-5.47
Religion Measure	-0.02981	0.00652	-4.58
Ethnolinguistic Fractionalization	-0.02579	0.00719	-3.59

Notes: Variables in bold denote those that emerge as robust growth factors (they are included in the specific model) using income from one of the datasets but not the other.