Measuring Income with Penn World Table Data: Reaching Consensus?*

Jesus Crespo Cuaresma[a,b,c,d] **Martin Feldkircher**[e]
Bettina Grün[f] **Paul Hofmarcher**[†f] **Stefan Humer**[a]

[a]Department of Economics, Vienna University of Economics and Business (WU)
[b]Austrian Institute of Economic Research (WIFO)
[c]International Institute of Applied System Analysis (IIASA)
[d]Wittgenstein Centre for Demography and Global Human Capital (WIC)
[e]Oesterreichische Nationalbank (OeNB)
[f]Department of Applied Statistics, Johannes Kepler University Linz (JKU)

March 9, 2015

Abstract

The Penn World Tables (PWT) are the most prominent source for international comparable GDP data. Recently, the robustness of empirical results in the empirical economic growth literature to revisions of the PWT data has been challenged. We propose a model framework based on latent variable specifications to derive consensus GDP series based on a panel dataset of income data from the six most recent PWT vintages. This approach allows us to take into account the variability associated with the different PWT revisions in a systematic and consistent way and to quantify the uncertainty surrounding real GDP figures in different vintages of the PWT dataset.

JEL Classification: C99, H99.

Keywords: Gross Domestic Product, Penn World Tables, consensus estimates, Bayesian methods.

*Corresponding Author: *Jesus Crespo Cuaresma*. Address: Welthandelsplatz 1, 1020 Vienna, Austria, Tel: +43(0)131336, Fax: +43(0)131336-728, [✉] *jcrespo@wu.ac.at*. The views expressed in this paper are not necessarily those of the Oesterreichische Nationalbank. We would like to thank participants of an internal research seminar at the University of Salzburg for helpful comments and suggestions. All remaining errors are ours.

†The author's research is supported by the Oesterreichische Nationalbank under the Jubiläumsfond grant 14663.

1 Introduction

The importance of Penn World Table (PWT) data for the advancement of research in development economics is difficult to underestimate. The PWT dataset presents the most comprehensive collection of internationally comparable GDP data available and as such has played a central role in providing empirical results related to the determinants of differences in economic growth across countries. The PWT database provides purchasing power parity (PPP) adjusted GDP (and GDP components) figures, which are essential for conducting sensible cross-country income comparisons. Repeated revisions due to updates of national income data, new price data from the International Comparison Program (ICP) and changes in the underlying methodology to estimate purchasing power parities (PPPs) result in substantial differences in real GDP figures for different vintages of the PWT. The most recent vintage, PWT 8.0, which was made available in 2013, provides yearly data for 167 countries over the period ranging from 1950 to 2011.

The construction of PPP-adjusted GDP relies on extensive price collections carried out in the framework of the ICP. Since 1970 there have been seven waves of these price collection exercises, with the latest one carried out in 2005. Countries participating in a price collection round are labelled *benchmark countries* and the years in which the collections are carried out are referred to as *benchmark years*. The different price data are averaged per product category and combined into so-called basic headings. For these, expenditure data are provided by national statistical offices. These basic headings are aggregated into country PPP levels by means of price measures,[1] which aim to give more weight to goods for which the economy spends most. It is this difference in spending patterns across countries that renders PPP-adjustment calculation so cumbersome as compared to correction of goods for price effects within a country (Feenstra, Inklaar, and M. P. Timmer, 2013). Finally, PPP levels are extrapolated for years other than the benchmark by employing national price dynamics and estimated for non-benchmark countries using regression techniques.

Since the first PWT vintage, several revisions of the methodology for obtaining PPP-adjusted income data have been carried out. Recently, these featured adjustments for quality differences in the underlying goods, the correction of systematic overvaluation in Chinese GDP figures, modifications of the aggregation measures to better accommodate for differences in spending patterns and the treatment of non-benchmark countries and benchmark years (Deaton and Heston, 2010). Feenstra, Inklaar, and M. P. Timmer (2013) provides a detailed description of the method used for the latest PWT dataset (PWT 8.0). The latest vintage of PWT data assesses issues related to the calculation of relative prices of exports and imports and consequently addresses explicitly the distinction between output-oriented and expenditure-oriented real GDP, so as to overcome the critique put forward in Feenstra, Heston, et al. (2009). The latest vintage of PWT data uses a new approach to estimate income in non-benchmark years, an issue which is particularly important for

[1] Other institutions that calculate PPP levels, such as the IMF or the World Bank, use different aggregation measures. For a more detailed account of the differences see Deaton and Heston (2010).

cross-country comparisons over time. Instead of using national price dynamics to extrapolate from the benchmark year, a method is proposed that interpolates data between the seven ICP rounds. PWT 8.0 is thus the first vintage that does not discard price data from previous collections, which is assumed to be a major improvement compared to earlier vintages. Although these adjustments could lead us to believe that, when it comes to PWT data, *the newer the better* (paraphrasing the title of the seminal work by Johnson et al., 2013), the existing literature on the robustness of empirical results to PWT vintages challenges such a view (see Johnson et al., 2013; Breton, 2012).

The sensitivity of empirical work to revisions in international income data has recently become a field of research on its own. Ponomareva and Katayama (2010) re-examine the empirical evidence concerning the effect of GDP growth volatility on income growth put forward in G. Ramey and V. A. Ramey (1995). They show that the results concerning the negative effect of volatility on economic growth found in G. Ramey and V. A. Ramey (1995) are not robust to the use of different vintages of PWT data. While the qualitative results do not appear to depend on the PWT vintage, the size of the estimated effects do depend strongly on the version of the PWT income data used for the analysis. Dowrick (2005) point out that the index measure used by PWT to calculate PPPs systematically overestimates GDP in poorer economies and that this effect is likely to bias the results of studies which deal with the degree of income inequality at the global level, such as Sala-i-Martin (2002). Breton (2012), in addition, argues that the data of the period ranging from 1950 to 1996 in the PWT 7.0 release are unreliable.

Johnson et al. (2013) carry out an extensive replication exercise where they re-estimate several specifications from thirteen prominent empirical studies on economic growth. They show that PWT revisions are quantitatively substantial and the estimates of growth regressions vary remarkably depending on the version of the PWT dataset used. Moreover, the results in Johnson et al. (2013) suggest that studies using high frequency data (e.g., annual data as opposed to averages over longer periods of time) are especially prone to the measurement error inherent to the methodology behind the PWT data. Johnson et al., 2013 conclude that robustness to PWT data revisions appears as a necessary condition to derive policy conclusions from empirical economic growth studies. In the framework of model uncertainty problems in economic growth regressions, Ciccone and Jarociński (2010) investigate the effect of PWT revisions on the robustness of economic growth determinants using the ubiquitous growth dataset by Sala-i-Martin, Doppelhofer, and Miller (2004) and three different revisions of the PWT[2] They find that the estimates of the effect of independent variables on economic growth vary strongly across PWT vintages. For example, using PWT 6.1 data the variable 'investment price' is found to be a key determinant of economic growth, while the variable receives practically no empirical support as a driver of income growth differences if PWT 6.2 data are used. Feldkircher and Zeugner (2012), however, show that

[2] In this paper we focus exclusively on vintage differences based on PPP revisions provided by the PWT database since this source offers the broadest coverage in terms of countries and time span. For an empirical evaluation of growth determinants using PPP-adjusted income data provided by other sources (IMF and World Bank data in addition to PWT) see Hanousek, Hajkova, and Filer (2008).

the instability of results found by Ciccone and Jarociński (2010) can be at least partly related to the methodological approach chosen by the authors. In any case, a significant amount of variation in the results of Feldkircher and Zeugner (2012) can also be attributed to the systematic revisions of the PWT.

In this paper we put forward a latent variable model that allows us to derive *consensus* GDP figures from the different vintages of PWT data. As in the case of Johnson et al. (2013), the focus of our contribution is on the time dimension of the PWT data. In the spirit of latent variable specifications which unveil the source of systematic and idiosyncratic dynamics in repeated panel structures (see, for instance, Grün et al. (2013) for an application to credit ratings), we propose the use of a model that is able to capture cross-country correlation structures among income and allows for clustering the dynamics of groups of economies along flexible long-term trends. The method is able to recover consensus estimates of GDP which incorporate information from all available PWT vintages and should thus overcome the sensitivity issues raised by the literature hitherto. The consensus estimates should thus be seen as a reasonable alternative to using vintage-specific PWT data. The remainder of the paper is structured as follows. Section 2 presents the econometric specification proposed to deal with the estimation of unobserved global, vintage-specific and idiosyncratic effects in a panel of PWT data from different vintages. Section 3 examines the results of the estimation of the model and the derived consensus estimates of income. Section 4 concludes and proposes further paths of research.

2 Econometric approach

In the following we propose a Bayesian probabilistic model which allows to derive a *consensus GDP* if several heterogenous sources of GDP estimates (here PWT vintages) are available. The advantages of such a consensus model are manifold: Firstly, following standard information economic arguments the estimated consensus GDP is more "informative" than the single vintages, as it encapsulates the information of several PWT vintages. Secondly, our model uses latent Dirichlet priors to identify groups of countries whose GDP is driven by a common macroeconomic latent market. This enables us to perform clustering of the countries according to their "long term " GDP trends. Finally, we estimate PWT version and country specific error terms. These errors might be of interest if one aims at validating PPP consensus of the PWT vintages for a country.

2.1 Model

In the following we denote the observed *GDP* of country i in PWT version j at time t, with $Y_{ij}(t)$. Following the literature (Breton, 2012; Johnson et al., 2013) the single observed GDP estimates $Y_{ij}(t)$ might vary for fix i, t between the versions j. Therefore we assume a latent "true" consensus

GDP, $Y_i^*(t)$, which can be estimated from the observed values $Y_{ij}(t)$. That is,

$$Y_{ij}(t) = Y_i^*(t) + \epsilon_{ij}(t), \tag{1}$$

with $\epsilon_{ij}(t)$ denoting the error of PWT version j for country i at time t. According to Equation 1 the only observable variables are $Y_{ij}(t)$ while probabilistic processes, both for the consensus $Y_i^*(t)$ and the error structure $\epsilon_{ij}(t)$ have to be specified.

2.1.1 Specification of the consensus GDP

In order to specify the dynamic structure of the consensus GDPs $Y_i^*(t)$ we assume that for each country i, $Y_i^*(t)$ consists of a convex combination of an idiosyncratic process $\nu_i(t)$ and a group specific latent market factor $f_g(t)$ for $g = 1, \ldots, G$. That is,

$$Y_i^*(t) = \beta_i \nu_i(t) + (1 - \beta_i)\sigma_i f_g(t), \tag{2}$$

with $\nu_i(t)$ denoting the idiosyncratic component of the GDP development and υ_i the long term-mean GDP rate of country i. The dependency from the idiosyncratic component is captured via β_i and $(1 - \beta_i)$ is the dependency from the specific latent market $f_g(t)$. Those markets $f_g(t)$ are multiplied with σ_i, the standard deviation of the $\nu_i(t)$ process, while the $f_g(t)$ are estimated as standardized processes. Those convex combinations of the variances σ_i ensure a reasonable variance decomposition in terms of idiosyncratic GDP development and market development. The priors for the GDP long term means υ_i in equation 2 are $\upsilon_i \sim N(\mu_\upsilon, \tau_\upsilon)$ with difuse priors $\mu_\upsilon \sim N(0, 10^6)$ and $\tau_\upsilon \sim IG(10^{-3}, 10^{-3})$. For the convex combination i.e. the β_i coefficients, we assume that the β_i follow a $Beta(2,2)$ distribution. $Beta(2,2)$ has a mode at 0.5 but also allows for values arbitrary close to 0 and 1. In the following we discuss specifications for $f_g(t)$ and $\nu_i(t)$.

Idiosyncratic GDP component Consensus GDP is the sum of both an idiosyncratic GDP driver, plus a (global) market. For modeling the idiosyncratic component we assume an AR(1) process (centered around 0) on $\nu_i(t)$, i.e.,

$$\nu_i(t) = \gamma_i \nu_i(t-1) + \epsilon_i,$$

with $\epsilon_i \sim N(0, \sigma_i^2)$ and $\sigma_i \sim^2 IG(1,1)$. The factor γ_i is drawn from a $(-1,1)$ truncated diffuse normal distribution.

Market groups G: Next to the idiosyncratic GDP development we consider a finite mixture approach for modeling the market components f_g, i.e. $f = \sum_{g=1}^{G} w_g f_g(\cdot)$. Hereby we wish to estimate both the single mixing distributions f_g as well as the number of groups G. Estimating the optimal number of groups G has attracted a large number of research. Here we follow Rousseau and Mengersen (2011), Ishwaran, James, and Sun (2002), and Frühwirth-Schnatter (2006) and propose an overfitting mixture model, that is we assume that an upper bound $G < G^* < \infty$ is know. As overfitted mixture models come along with some identifiability issues, Frühwirth-Schnatter (2006) presents some Bayesian solutions to this issue. Ideally, in a Bayesian setting, a prior on the maximum number of groups should guard against such an overfitting. In fact, Rousseau and Mengersen (2011) studied the asymptotic behavior of the posterior distribution in over-fitted Bayesian mixture models, i.e., models having more components than actually needed. Moreover, they showed that appropriate priors lead to emptying the redundant groups, i.e., in the Bayesian setting suitable priors can be used to identify the optimal number of clusters by assuming that there exists an conservative upper bound on the number of groups G. Following Rousseau and Mengersen (2011), Dirichlet priors $D(\alpha)$ with small parameters α on the mixture weights w_g are appropriate configurations for estimating the "optimal" number of groups G. Therefore a practical approach is to estimate the model with a large upper bound of groups G^*, a Dirichlet prior with small α parameters on the weights w_g and to check the posterior for small weights, i.e. "empty groups". In our application we set $\alpha_l = 0.001$ for $l = 1, \ldots, 20 = G^*$ and the weights of the clusters are drawn from Dirichlet priors $D(\alpha_1, \ldots, \alpha_{G^*})$.

From Market assignment to consensus market assignment: Although the proposed Dirichlet prior approach allows to estimate the optimal number of clusters, a general problem is the label switching of the latent variables describing the cluster assignments. As a simple example, consider mixture model with two components and data which can also be well separated into two latent groups A and B. In one MCMC chain A and B are represented by group assignments 1 and 2 and the other chain vice versa, i.e., by 2 and 1. It is obvious that from averaging over the two chains we get meaningless results for the group assignments.

In order to circumvent this problem, Gordon and Vichi, 2001 propose consensus clustering, which allows to estimate the consensus cluster assignments given several different cluster assignments (resulting from the MCMC chains). They apply their approach on macroeconomic data of 21 countries. Methodologically, Gordon and Vichi, 2001 use a variant of the Euclidean dissimilarity based on the sum of squared differences of the membership of non-empty classes. Essentially, assuming C independent MCMC chains and the associated group assignments $G^1, \ldots, G^C$ of MCMC chains, where G^c can be interpreted as a $0-1$ matrix with $G^c_{jg} = 1$ if for chain c country j is assigned to market g and 0 otherwise. Further let G the sought least square consensus market assignment, and $P^1, \ldots, P^C$ denote permutation matrices, one aims at minimizing

$$\sum_{c}^{C} \|G - G^c P^c\| ,$$

over all potential group assignment matrices G (i.e. stochastic matrices) and permutation matrices

$P^1, \ldots, P^C$ [3]

Market factors f_g: Several choices of latent processes might be suitable for modeling market components f_g. Inspecting the raw GDP figures suggests that some flexible nonlinear functional approach is required for modeling a common GDP development, f_g. Here we make use of penalized cubic spline regressions (P-splines), where the coefficients are only partly determined by the data but also a penalty term is added to avoid overfitting.This approach assumes that the latent market effect f_g can be modeled by a polynomial spline written in terms of linear combinations of basis spline functions (see Lang and Brezger, 2004). Assuming d knots, we have

$$f_g(t) = \sum_{m=1}^{d} \xi_m B_m(t).$$

The basis functions $B_m(t)$ are nonzero only on a domain spanned by $2 + l$ knots with l denoting the degree of the spline (here $l = 3$). A crucial problem within this framework is the number of knots d. Too many knots may lead to overfitting, while too few knots may generate a function which is not flexible enough to model the data adequately. P-Splines circumvent this problem via penalty terms. Eilers and Marx (1996) propose to use a relatively large number of knots combined with a difference penalty on coefficients of adjacent basis splines and Lang and Brezger (2004) presents a Bayesian setting for this idea. To ensure enough flexibility for the market development equally spaced knots for each 2nd time point are assumed. Along the lines of Lang and Brezger (2004) or Scheipl and Kneib (2009)(p.3535) conventional Bayesian P-Spline smoothing is based on a random walk prior for the (first) differences of the priors for ξ, i.e., $\xi_m = \xi_{m-1} + u_m$ with Gaussian errors $u_m \sim N(0, \tau^2)$, $\tau \sim IG(10^{-3}, 10^{-3})$ and diffuse priors $\xi_1 \propto$ const. Note that in this Bayesian setting the amount of smoothness is controlled via the parameter τ. For a detailed discussion on Bayesian P-Spline smoothing, we refer to Scheipl and Kneib (2009).

The error structure

Next to the model for consensus GDPs $Y_i^*(t)$ we have to discuss suitable models for the errors $\epsilon_{ij}(t)$ in Equation 1. In our model an error structure which allows to capture constant shifts between the estimated consensus GDP and the observed PWT vintage specific estimates is desired.

$$\epsilon_{ij} = \mu_{ij} + \sigma_j, \tag{3}$$

with $\sigma_j^2 \sim IG(10^{-3}, 10^{-3})$. The systematic errors of the different PWT versions, μ_{ij}, can be interpreted as the deviation of the vintage specific estimate for country i from the consensus. Those errors μ_{ij} are drawn from $N(0, \sigma_\mu^2)$ with $\sigma_\mu \sim IG(10^{-3}, 10^{-3})$.

[3] Here, we omit a detailed discussion on consensus clustering, but refer the interesting reader to Gordon and Vichi, 2001; Hornik, 2005.

Note that the μ_{ij} errors capture differences in the parallel shifts of the PWT versions. Economically this means that the effect of a PPP correction between the PWT vintages is estimated in the μ_{ij} values.

3 Data

We apply our dynamic latent GDP model to 4 different consecutive PWT vintages, namely 6.3 to 8.0. PWT vintages 7.1 and earlier versions can be downloaded from https://pwt.sas.upenn.edu/php_site/pwt_index.php, while PWT 8.0 is provided by the University of Groningen, available for download at www.ggdc.net/pwt. Table 1 presents some characteristics of those vintages, like the number of countries contained in each vintage, which varies between 167 for PWT 8.0 and 189 for PWT 7.0/7.1. For all considered PWT vintages, 2005 serves as the reference year for estimating the GDPs. The input prices in PWT are based on the International Comparison Program (henceforth ICP), which aims at collecting comparative price data and PPP levels. The data of PWT 6.3 are based on ICP 2002, PWT 7.0 and 7.1 use ICP 2005, while PWT 8.0 differs from the previous versions that all available price collecting rounds of ICP are used to get PPPs and GDP estimates. In total 6 phases of ICP rounds are used for PWT 8.0 (see http://www.rug.nl/research/ggdc/data/pwt/v80/pwt_80_user_guide.pdf page 8 Table 2). PWT 8.0 incorporates all those ICP rounds although following Feenstra, Inklaar, and M. P. Timmer (2013) (user manual) the 2005 ICP was a great improvement over the earlier versions. Most notably it covers the largest number of countries and also the data collection was more rigorous.

PWT	6.3	7.0	7.1	8.0
From	1950	1950	1950	1950
To	2007	2009	2010	2011
Base Year	2005	2005	2005	2005
Countries	188	189	189	167
ICP	1996??	2005	2005	1970–2005

Table 1: Some summary statistics of the different PWT versions.

4 Results

4.1 The Consensus GDPs

Figure 1 displays the estimated consensus GDPs $Y_i^*(t)$ for the ring countries as well as the GDP estimates of the single PWT versions. Additionally the 95% credible intervals for the consensus estimates are displayed as grey shaded areas. We can clearly infer, that those credible intervals enlarge for period where no data are available, see e.g., Estonia, Slovenia (countries that emerged at the beginning of 1990s). Firstly, figure 1 clearly illustrates the heterogeneity between the single PWT vintages. The highest degree of consensus between the PWT vintages is observed for the US. Out of the ring countries we observe the highest disagreement between the PWT versions for Kenya, Senegal.

4.2 Decomposition of Consensus into idiosyncratic process and market process

A feature of our latent consensus model is that it allows for decomposition of the estimated consensus GDP into an idiosyncratic growth path and a market driven one (see equation 2). Figure 2 shows those decompositions for the ring countries. The gray bars on the top of each figure displays the fraction of the consensus GDP explained by the estimated market $Ma_i(t) = (1-\beta_i)\sigma_i f_g(t)$ for each country. We estimate this fraction as

$$R_i = 1 - \frac{RSS_{Ma_i}}{RSS_{Ma_i} + RSS_{Vi}} \in [0,1],$$

with $RSS_{Ma} = \frac{1}{T}\sum_t (Y_i(t) - Ma_i(t))^2$ denoting the mean residual sum of squares of the differences between the estimated consensus GDP $Y_i(t)$ and the estimated market $Ma_i(t)$. If the estimated market is explains the consensus GDP perfectly R_i becomes 1. Otherwise, if the consensus GDP of a country is mainly explained by idiosyncratic GDP developments this fraction moves towards zero. From figure 2 we can infer that the highest R_i values are observed for Sri Lanka (0.986) and Slovenia (0.9968) (see Table 2).

	R_i		R_i
ZMB	0.381	EGY	0.859
CMR	0.446	HKG	0.868
JOR	0.452	JPN	0.877
KEN	0.601	ITA	0.885
CHL	0.703	GBR	0.893
SEN	0.710	MYS	0.915
ZAF	0.712	DEU	0.931
PHL	0.738	CAN	0.956
BRA	0.759	FRA	0.977
USA	0.818	EST	0.982
OMN	0.824	LKA	0.987
CHN	0.826	SVN	0.997

Table 2: table of R_i for ring countries also displayed as gray bars in figure2

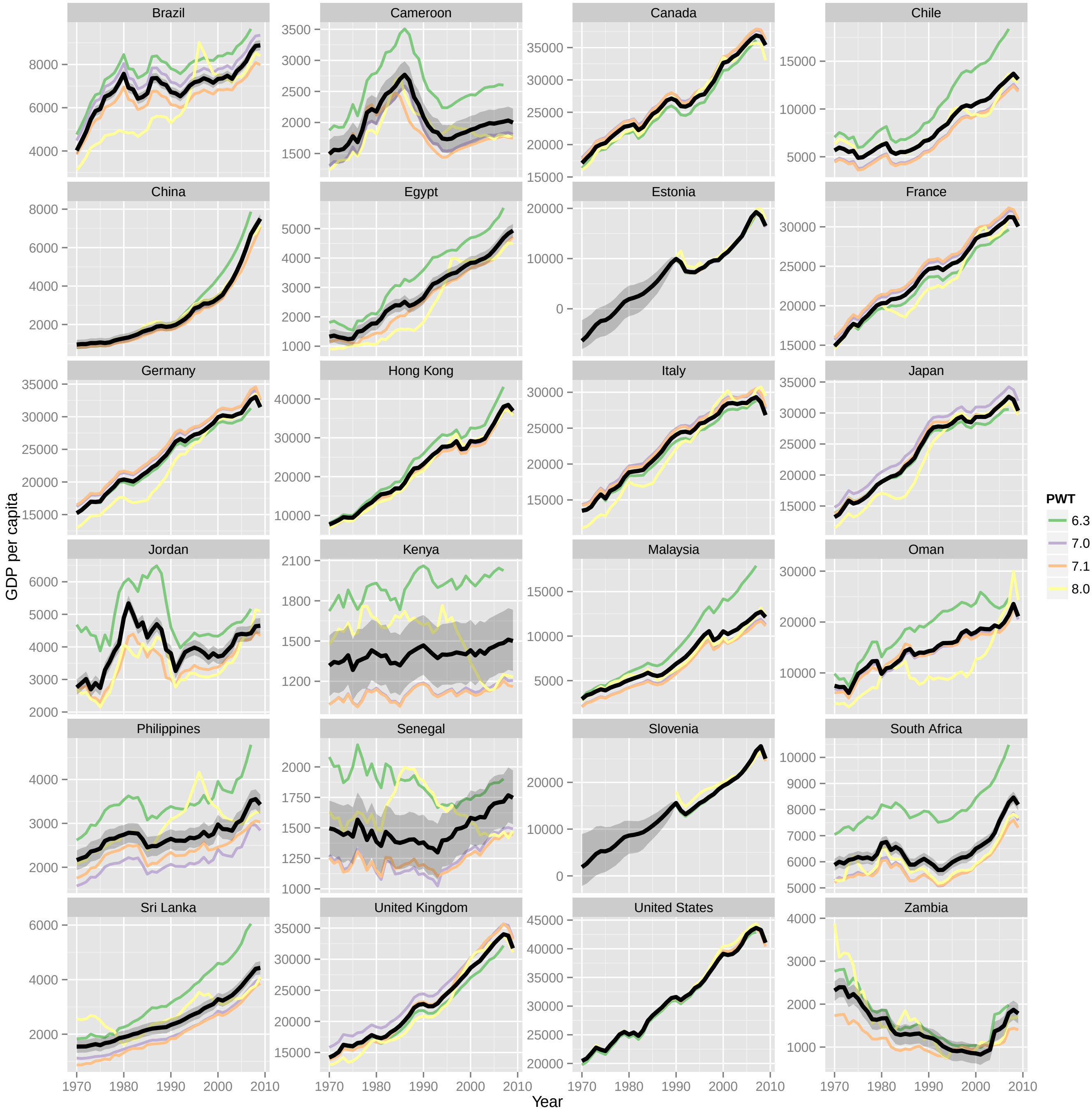

Figure 1: Consensus GDP (black line) and PWT GDP estimates for the Ring Countries. On the y-axis the per capita GDP divided by the countries long term mean GDP (over the different PWT vintages)

4.3 Analyzing the estimated clusters and consensus clusters

We ran three independent MCMC chains for estimating our model whereby we assumed a priori 20 different groups of markets. Following **McCullagh, Bayesian Analysis 2008** data seldom

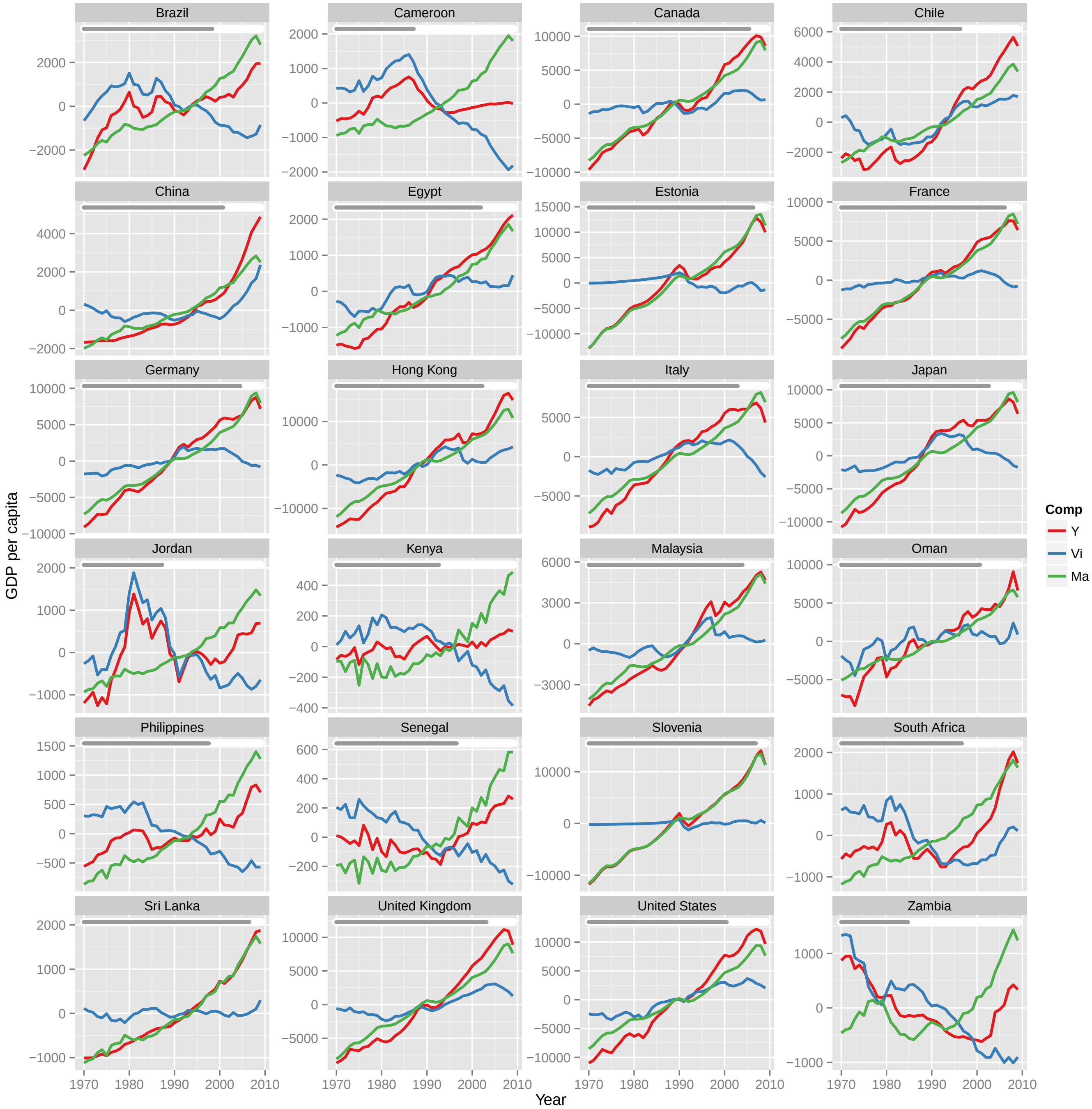

Figure 2: Decomposition of the estimated Consensus GDP in market factor and idiosyncratic part. Estimates for the Ring Countries

contain information on the the number of clusters. Here we used independent chains to get a picture of the clusters. In order to check the sensitivity of the cluster assignment table 3 displays the Allowing up to 20 different clusters a priori, our sampling procedure results between 4 and 9 clusters for each chain.

Before analyzing the estimated consensus cluster based on the cluster assignments of the different MCMC chains, Table 3 presents some descriptive results of the chain wise clustering.

	Chain 1	Chain 2	Chain 3
A	95	111	107
B	50	44	48
C	41	17	22
D	3	6	5
E	0	3	2
F	0	3	2
G	0	3	2
H	0	1	1
I	0	1	0

Table 3: Size of the estimated clusters for the different MCMC chains

We can infer from Table 3 that all chains result in 3 main clusters followed by some smaller ones. The intersection of cluster A between the chains contains 69 countries,indicating that the interpretation of cluster A is similar between the chains. For cluster B we only find 7 countries which are assigned to B in all three chains, but the intersection of cluster C from chain 1 and B form chains two and three contains 27 (out of 41 possible) countries. Again this indicates that cluster C of chain 1 captures similar countries as B for chains 2 and 3.

Consensus clustering Analysis: Estimating a consensus cluster partition out of the single chain cluster-results allows for an easier and more straightforward interpretation. Using the approach proposed in Gordon and Vichi, 2001 we estimate a consensus clustering, by assuming after inspecting the raw cluster results 4 groups of clusters. Table 4 illustrates the frequencies of the cluster assignments.

Cluster 1	Cluster 2	Cluster 3	Cluster 4
3	42	96	48

Table 4: Frequencies of cluster assignments after consensus clustering.

Table 4 illustrates that essentially the consensus clustering results in 3 main clusters, one including 95 countries and two clusters containing 49 respectively 43 countries. Table 5 displays the cluster assignments for the ring countries. [4]

4.4 The consensus deviations μ_{ij}, or validating the PWT vintages

Next to the estimated consensus GDP rates and their decomposition into "global" market trends and idiosyncratic developments we present results for the consensus deviation μ_{ij}. The

[4]Complete clustering results available in the supplemental materials

Cluster	Number	Ring Countries
2	11	GBR JPN SVN EST HKG MYS USA DEU ITA FRA CAN
3	9	CMR EGY KEN SEN ZAF JOR PHL LKA CHN
4	4	ZMB OMN BRA CHL

Table 5: Cluster assignments of the ring countries after consensus clustering over three independent MCMC chains.

interpretation of those errors is twofold: Firstly, by aggregating over the single countries i the mean errors $\frac{1}{N}\sum_i^N \mu_{ij}$ allow to validate which PWT vintages are more optimistic respectively pessimistic given the consensus GDPs. Table 6 summarizes these results. The lowest error is estimated for PWT 8.0 indicating that this version is too optimistic ($\sim$ 462.55$ above the consensus) conditional on our estimated consensus GDPs, while PWT 6.3 seems to be more pessimistic ($\sim$ 302.78$ below the consensus).

	PWT 6.3	PWT 7.0	PWT 7.1	PWT 8.0
1	302.78	59.42	100.36	-462.55

Table 6: Aggregated vintage specific average errors μ_{*j} over all 189 countries in sample

Secondly, Table 7 displays the estimated vintage specific errors μ_{ij} as well as the the errors divided by the country's and version's long term mean.

Secondly, estimating the PPP is a difficult task. If for a given country two PWT vintages differ only through a change in the PPP, the GDP data can be constructed from each other through parallel shifts, which corresponds to adding a time independent vintage and country specific constant. The μ_{ij} capture those effects. Lower (mean absolute) μ_{i*} indicate that a consent on the PPP between the PWT vintages exists, while higher absolute error values might come along with corrections of PPP between the versions. From table 7 we can clearly infer that the lowest mean absolute error is observed for the United States, a result which is expected and desired as the US Dollar serves a the reference currency for estimating PPP.

5 Concluding Remarks

To be written. It will be included in the final submission.

Computational Details

We used MCMC methods, in particular JAGS and its R-packages **rjags** and **code** to estimate our Models. We ran three independent chains each with 10000 iterations whereby the first 5000 were discarded as burn in, and a thinning of 10 was used. This resulted in 500 draws out of the posterior

	PWT 6.3	PWT 7.0	PWT 7.1	PWT 8.0	PWT 6.3	PWT 7.0	PWT 7.1	PWT 8.0	$\lvert\bar{\mu}_{ij}\rvert$
GBR	-1069.59	1642.06	385.43	-957.90	-0.05	0.07	0.02	-0.04	0.04
JPN	-439.39	1593.58	378.49	-1532.68	-0.02	0.06	0.02	-0.07	0.04
SVN	-341.57	-117.50	-186.69	645.76	-0.02	-0.01	-0.01	0.03	0.02
EST	-336.01	-223.10	-73.66	632.77	-0.03	-0.02	-0.01	0.05	0.03
CMR	567.10	-188.56	-205.09	-173.45	0.22	-0.10	-0.11	-0.09	0.13
EGY	685.37	3.91	-264.34	-424.94	0.20	0.00	-0.10	-0.18	0.12
KEN	506.48	-282.22	-312.52	88.27	0.27	-0.25	-0.28	0.06	0.21
SEN	394.25	-246.20	-256.37	108.32	0.21	-0.20	-0.21	0.07	0.17
ZAF	1706.67	-601.23	-664.20	-441.25	0.21	-0.10	-0.11	-0.07	0.13
ZMB	218.30	-14.64	-314.32	110.66	0.13	-0.01	-0.28	0.07	0.12
JOR	996.67	-6.84	-493.79	-496.05	0.20	-0.00	-0.14	-0.14	0.12
OMN	4319.07	-608.67	-585.92	-3124.48	0.23	-0.04	-0.04	-0.29	0.15
HKG	2257.87	-639.03	-942.58	-676.26	0.10	-0.03	-0.04	-0.03	0.05
MYS	1983.71	-837.33	-973.40	-172.98	0.22	-0.13	-0.15	-0.02	0.13
PHL	771.00	-584.71	-359.16	172.87	0.22	-0.27	-0.15	0.06	0.18
LKA	832.50	-437.01	-565.16	169.67	0.25	-0.21	-0.28	0.06	0.20
BRA	956.40	463.54	-559.09	-860.86	0.12	0.06	-0.09	-0.14	0.10
CHL	2412.62	-1115.05	-1244.24	-53.33	0.24	-0.16	-0.18	-0.01	0.15
CHN	264.64	-96.21	-244.25	75.82	0.10	-0.04	-0.10	0.03	0.07
USA	-418.51	49.03	-13.24	382.72	-0.01	0.00	-0.00	0.01	0.01
DEU	-583.76	1026.43	1276.02	-1718.69	-0.03	0.04	0.05	-0.08	0.05
ITA	-656.18	882.51	717.09	-943.42	-0.03	0.04	0.03	-0.04	0.04
FRA	-930.42	783.29	1088.88	-941.75	-0.04	0.03	0.04	-0.04	0.04
CAN	-1113.26	828.40	796.37	-511.52	-0.04	0.03	0.03	-0.02	0.03

Table 7: Deviations μ_{ij} from the estimated consensus GDPs for the single PWT vintages. Columns 1-4 present the raw estimated μ_{ij}, columns 5-8 the errors divided by the countries' and versions' long term mean to make values comparable.

for each chain.

References

Breton, T. R. (2012): "Penn World Table 7.0: Are the data flawed?" In: *Economics Letters* 117 (1), pp. 208–210. issn: 0165-1765. url: http://www.sciencedirect.com/science/article/pii/S0165176512002807.

Ciccone, A. and M. Jarociński (2010): "Determinants of Economic Growth: Will Data Tell?" In: *American Economic Journal: Macroeconomics* 2 (4), pp. 222–46. url: http://www.aeaweb.org/articles.php?doi=10.1257/mac.2.4.222.

Deaton, A. and A. Heston (2010): "Understanding PPPs and PPP-Based National Accounts". In: *American Economic Journal: Macroeconomics* 2 (4), pp. 1–35. url: http://www.aeaweb.org/articles.php?doi=10.1257/mac.2.4.1.

Dowrick, S. (2005): "The Penn World Table: A Review". In: *Australian Economic Review* 38 (2), pp. 223–228. issn: 1467-8462. url: http://dx.doi.org/10.1111/j.1467-8462.2005.00369.x.

Eilers, P. H. C. and B. D. Marx (1996): “Flexible Smoothing with B-Splines and Penalities”. In: *Statistical Science* 11 (2), pp. 89–121.

Feenstra, R. C., R. Inklaar, and M. Timmer (2013): “The Next Generation of Penn World Table”. In: *Nber Working Paper Series* 19255). url: http://www.nber.org/papers/w19255.

Feenstra, R. C., A. Heston, et al. (2009): “Estimating Real Production and Expenditures across Nations: A Proposal for Improving the Penn World Tables”. In: *Review of Economics and Statistics* 91 (1), pp. 201–212. url: http://dx.doi.org/10.1162/rest.91.1.201.

Feenstra, R. C., R. Inklaar, and M. P. Timmer (2013): *The Next Generation of the Penn World Table*. PWT 8.0. Groningen Growth and Development Centre. url: www.ggdc.net/pwt.

Feldkircher, M. and S. Zeugner (2012): “The impact of data revisions on the robustness of growth determinants—a note on ‘determinants of economic growth: Will data tell?’” In: *Journal of Applied Econometrics* 27 (4), pp. 686–694. url: http://ideas.repec.org/a/wly/japmet/v27y2012i4p686-694.html.

Fernandez, C., E. Ley, and M. Steel (2001a): “Benchmark priors for Bayesian model averaging”. In: *Journal of Econometrics* 100 (2), pp. 381–427.

— (2001b): “Model uncertainty in cross-country growth regressions”. In: *Journal of Applied Econometrics* 16 (5), pp. 563–576.

Frühwirth-Schnatter, S. (2006): *Finite Mixture and Markov Switching Models (Springer Series in Statistics)*. 1st ed. Springer. isbn: 0387329099. url: http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20%5C&path=ASIN/0387329099.

Gordon, A. and M. Vichi (2001): “Fuzzy partition models for fitting a set of partitions”. In: *Psychometrika* 66 (2), pp. 229–247. url: http://EconPapers.repec.org/RePEc:spr:psycho:v:66:y:2001:i:2:p:229-247.

Grün, B. et al. (2013): “Deriving Consensus Ratings of the Big Three Rating Agencies”. In: *Journal of Credit Risk* 9 (1), pp. 75–98.

Hanousek, J., D. Hajkova, and R. K. Filer (2008): “A rise by any other name? Sensitivity of growth regressions to data source”. In: *Journal of Macroeconomics* 30 (3), pp. 1188–1206. url: http://ideas.repec.org/a/eee/jmacro/v30y2008i3p1188-1206.html.

Heston, A., R. Summers, and B. Aten (2002): *Penn World Table Version 6.1*. PWT 6.1. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

— (2006): *Penn World Table Version 6.2*. PWT 6.2. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

— (2009): *Penn World Table Version 6.3*. PWT 6.3. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

— (2012a): *Penn World Table Version 7.0*. PWT 7.0. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

— (2012b): *Penn World Table Version 7.1*. PWT 7.1. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

Hornik, K. (2005): “A CLUE for CLUster Ensembles”. In: *Journal of Statistical Software* 14), pp. 65–72.

International Comparison Program (2011): *ICP Handbook*. Methodological Handbook. World Bank. url: http://go.worldbank.org/TP5TQW25P0.

Ishwaran, H., L. F. James, and J. Sun (2002): “Bayesian Model Selection in Finite Mixtures by Marginal Density Decompositions”. In: *Journal of the American Statistical Association* 96 (456), pp. 1316–1332.

Johnson, S. et al. (2013): "Is newer better? Penn World Table Revisions and their impact on growth estimates". In: *Journal of Monetary Economics* 60 (2), pp. 255–274. issn: 0304-3932. url: http://www.sciencedirect.com/science/article/pii/S0304393212001377.

Lang, S. and A. Brezger (2004): "Bayesian P-Splines". In: *Journal of Computational and Graphical Statistics* 13 (1), pp. 183–212.

McCullagh, P. and J. Yang (2014): "How many clusters?" In: *Bayesian Analysis*), pp. 1–19.

Neary, J. P. (2004): "Rationalizing the Penn World Table: True Multilateral Indices for International Comparisons of Real Income". English. In: *The American Economic Review* 94 (5), pages. issn: 00028282. url: http://www.jstor.org/stable/3592828.

Ponomareva, N. and H. Katayama (2010): "Does the version of the Penn World Tables matter? An analysis of the relationship between growth and volatility". In: *Canadian Journal of Economics* 43 (1), pp. 152–179. issn: 1540-5982. url: http://dx.doi.org/10.1111/j.1540-5982.2009.01567.x.

Ram, R. and S. Ural (2013): "Comparison of GDP Per Capita Data in Penn World Table and World Development Indicators". English. In: *Social Indicators Research*), pp. 1–8. issn: 0303-8300. url: http://dx.doi.org/10.1007/s11205-013-0284-0.

Ramey, G. and V. A. Ramey (1995): "Cross-Country Evidence on the Link between Volatility and Growth". In: *American Economic Review* 85 (5), pp. 1138–51. url: http://ideas.repec.org/a/aea/aecrev/v85y1995i5p1138-51.html.

Rousseau, J. and K. Mengersen (2011): "Asymptotic behaviour of the posterior distribution in overfitted mixture models ". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 73 (5), pp. 689–710.

Sala-i-Martin, X., G. Doppelhofer, and R. Miller (2004): "Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach". In: *American Economic Review*), pp. 813–835.

Sala-i-Martin, X. (2002): *The Disturbing "Rise" of Global Income Inequality*. NBER Working Papers 8904. National Bureau of Economic Research, Inc. url: http://ideas.repec.org/p/nbr/nberwo/8904.html.

Scheipl, F. and T. Kneib (2009): "Computational Statistics and Data Analysis". In: *Computational Statistics & Data Analysis* 53 (10), pp. 3533–3552.

The World Bank (2013): *International Comparison Program*. url: http://go.worldbank.org/PQ5ZPPYSY0.

World Bank (1993): *Purchasing Power of Currencies – Comparing National Incomes Using ICP Data*. Technical Note. The World Bank. url: http://siteresources.worldbank.org/ICPINT/Resources/270056-1255977254560/PurchasingPowerofCurrencies-1993.pdf.

— (2008): *Global Purchasing Power Parities and Real Expenditures – 2005 International Comparision Program*. Methodological Handbook. The World Bank. url: http://siteresources.worldbank.org/ICPINT/Resources/270056-1255977254560/6483625-1337016259587/2005ICPReport_FinalwithNewAppG.pdf.