Essays on Macroeconomics in Africa

Lionel Roger

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

Nottingham, March 2018

Abstract

This thesis comprises three self-contained essays on topics surrounding economic growth and macroeconomic performance in sub-Saharan Africa (SSA). The first two address aspects of measurement error in macroeconomic data from SSA: the first essay (Chapter 2) explores the potential of nighttime luminosity as a means of improving GDP estimates, the second (Chapter 3) assesses the potential impact of changes in the data on empirical results from time series analysis. The third essay (Chapter 4) investigates exchange rate dynamics in Zambia, and how these pass through to consumer prices.

The first essay (Chapter 2) explores heterogeneity in the nexus between nighttime light emissions and economic activity, a relationship that is increasingly exploited with the aim of reducing measurement error in growth estimates of countries where the data is considered weak. I show that the elasticity between nighttime lights and economic activity varies substantially across countries, and that this variation has a systematic component which could have implications for empirical results. Using the elastic net method in order to isolate the relevant factors, I find that 55% of the variation can be explained by observable factors. When tracing economic growth in Africa since 1992 using luminosity and accounting for the explained share of parameter heterogeneity, I find no evidence of an 'African Growth Miracle' as described by, e.g., Young (2012). However, I do find evidence that countries that recently revised their GDP figures (like Nigeria and Ghana) had a tendency to report inflated growth rates for recent years. This is consistent with Jerven (2014)'s hypothesis of purely 'statistical growth'.

The second essay (Chapter 3) explores the inconsistencies across different versions and sources of national accounts data (three versions of the Penn World Table and the World Development Indicators), and their impact on macroeconomic inference from time series analysis. I use the statistical framework developed by Juselius, Møller and Tarp (2014) and assess the robustness of their conclusions regarding the long-run impact of foreign aid on economic growth in 36 SSA countries that are generally considered to have low statistical capacity, and where sources on macroeconomic variables sometimes disagree strongly. The results of this exercise are mixed: When I apply the Cointegrated VAR models precisely as developed

by JMT to the new datasets, results remain robust for about two thirds of the countries. Once I re-specify the time series models based on the respective data (as different data will sometimes imply different lag lengths, cointegration ranks, etc.), using the same statistical methodology as JMT, this often leads to more substantial changes in the conclusions.

The third essay (Chapter 4) investigates the dynamics between the exchange rate and consumer prices in a resource-rich setting in a case study of Zambia. Using a combination of short-run sign- and zero-restrictions, I identify relevant global and domestic shocks in a structural VAR (SVAR). The results suggest that the pass-through of the exchange rate to consumer prices (ERPT) depends greatly on the shock that originally caused the exchange rate to fluctuate. While, for instance, the price of copper is the most important driver of the exchange rate, the fluctuations caused by it tend to affect prices only moderately (an ERPT of ca. 7%). On the other hand, exchange rate fluctuations caused by monetary shocks come with a much higher pass-through of up to 25%.

Acknowledgements

I am deeply indebted to my supervisors Oliver Morrissey and Markus Eberhardt for the invaluable guidance and support they gave me throughout the past couple of years. With their infallible sense of priorities and feasibility, they never failed to put me back on track whenever I got lost in the bewildering abyss research can sometimes turn into. Their commitment dates back to the Masters, extends to the process of securing funding for the PhD, and has continued without interruption until this day.

I would also like to thank the other members of the Nottingham School of Economics. The academic staff, for their excellent comments and suggestions. The administrative staff, for making things run so smoothly and maintaining a familiar atmosphere despite the challenges of centralisation. And my fellow students, who are a fantastic bunch of people.

I thank my friends back home, who always made me feel close regardless the distance, and my friends in Nottingham, for expanding my conception of 'home'. A special mention is due to Vasilis, with whom I have walked through the ups and downs of postgraduate life since the very first day of it.

None of this PhD or anything else in life would have been possible without the support of my family, so once more, I thank them. Finally, I thank Martina for coffee every morning, and just so much more.

The financial support of the ESRC (award number 1511835) and the Nottingham School of Economics' Graduate Teaching Fellowship were greatly appreciated. The World Bank Group financed part of the work that went into chapter 4, and gave me the opportunity to travel to Zambia to discuss my research with local stakeholders. All views are my own.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Over the course of the past decade, there has been a tremendous shift in the perception of African economies. Where front pages once referred to Africa as the 'hopeless continent', headlines now read 'Africa rising' or 'aspiring Africa' (The Economist, 2000, 2011, 2013). Where scholarly debate once revolved around the 'bottom billion' (Collier, 2008) and 'Africa's growth tragedy' (Easterly and Levine, 1997), one of the main controversies in recent years has been whether or not talking about an 'African growth miracle' (Young, 2012) is an exaggeration.

Indeed, the turnaround of African economies seems remarkable. According to the World Bank's World Development Indicators (WDI), per capita incomes in sub-Saharan African (SSA) countries had been in decline almost without interruption for two decades between 1975 and 1994, falling on average by 1.3% every year. But this gloomy episode was followed by a spectacular recovery, with incomes increasing by about 2.0% on average every year between 1995 and 2014; paired with population growth, total GDP growth then averaged 4.5% annually over this period in SSA, compared to a global growth rate of 3.0%. Judging by these numbers, the euphoria that has surrounded the recent debate on the African growth performance seems natural.

Simultaneously, however, a different debate on African GDP figures erupted. After 'Africa's growth tragedy' had just been supplanted by the 'African growth miracle', 'Africa's statistical tragedy' (Devarajan, 2013) was now making the headlines. Sparked by Morten Jerven's critical accounts of how many statistical agencies in SSA were operating (Jerven, 2013c), especially regarding the accuracy of data on which GDP was estimated, growth statistics like those reported above were met with increasing scepticism. Given the low capacity, occasionally paired with political opportunism, official statistics from African countries may have been grossly misleading for decades. For some places, there did not even exist any official estimates of economic performance for much of their recent history. Aggregate databases such as the WDI or the Penn World Table (PWT) often

reported numbers nonetheless, filling gaps using a variety of techniques of inter- and extrapolation often based on strong if not heroic assumptions. Efforts to harmonise income figures across countries, and differences across data suppliers in the methodologies employed to do so, further added to the confusion surrounding those numbers.

The first two chapters of this thesis are dedicated to assessing the degree and the implications of the shortcomings of sub-Saharan African statistics. Chapter 2 considers if the growth rates suggested by GDP are consistent with the increase in economic activity suggested by emissions of light at night, as a means of assessing the reliability of GDP growth data. Chapter 3 turns to implications: are the inferences from studies of growth performance over time sensitive to the source of GDP data employed? Both chapters focus on GDP over a relatively long period for a relatively large sample of countries. In contrast, Chapter 4 considers macroeconomic policy analysis in one country, investigating the effect of the price of copper (the major export) on the exchange rate and inflation in Zambia, allowing for monetary responses. The analysis shares the application of time series econometric techniques with Chapter 3, but has a more specific focus of addressing current policy issues in macroeconomic management.

The first essay (Chapter 2) explores the recent African growth performance (1992-2013) using human-made light emissions at night, as recorded by satellites originally intended to detect cloud coverage. Due to the intimate connection between economic activity and lights emissions, these data have become a popular proxy for economic growth where official data is considered weak (Chen and Nordhaus, 2011; Henderson et al., 2012). Of course, luminosity is not a perfect proxy for GDP and cannot replace sound national statistics, but it can be a valuable complement. The feature that makes lights unique in their usefulness as an indicator of economic performance is that they are credibly independent of countries' statistical capacity or political meddling in the compilation of national accounts.

Chapter 2 aims to identify and address an important methodological issue in the use of luminosity data. Implicitly or explicitly, all of the existing literature relies on the assumption that the nexus between lights and economic activity is identical for all countries. I show that this is not the case, and that, across countries, the amount of light that is emitted with every incremental unit of GDP can differ substantially. If this is not accounted for, the lights proxy has a tendency to misleadingly suggest substantial revisions to growth rates that were correct in the first place. As a remedy, I suggest a method that uses machine learning to identify the determinants of the lights-GDP relationship, incorporating geographical and economic structure properties of countries that influence the degree to which economic activity translates into the emission of luminosity. On this basis, growth

rates based on luminosity that take into account the explained part of the heterogeneity across countries can be computed. This offers the potential to improve on existing methods, and will be tested and evaluated through a simulation exercise and then used for assessing the recent growth performance of SSA.

The second essay (Chapter 3) focusses on the implications of inconsistencies across GDP datasets for empirical research. In addition to the concern that official national statistics for SSA countries may have serious inaccuracies, GDP and growth estimates from different sources are not entirely consistent, and in some cases exhibit large divergences that cannot be explained simply by differences in way the data are constructed. This begs the question of the reliability of conclusions from analysis of African growth performance: would inferences be the same using different sources of data? The chapter will contribute to a growing body of literature that is concerned with assessing the robustness of empirical results to changes between GDP datasets, where many established findings turn out to be worryingly sensitive in this respect (e.g., Ponomareva and Katayama, 2010; Ciccone and Jarociński, 2010; Atherton et al., 2011; Johnson et al., 2013). Two findings of this literature in particular motivate my study. First, the results of studies that look at lower income (non-OECD) countries appear to be generally more sensitive to the choice of dataset. Second, studies that exploit data at a higher frequency tend to be more fragile in this respect than studies that consider long-run averages. Time series studies relying on annual GDP data from SSA countries would then be expected to be particularly susceptible to changes in the dataset.

The aim of the chapter is to assess if varying the source of GDP (and other macroeconomic) data affects the results of time series analysis of SSA growth. To this end I revisit the study by Juselius et al. (2014, henceforth JMT) on the macroeconomic impacts of aid in SSA countries, applying the same cointegrated vector autoregressive (CVAR) framework but with four popular alternative macroeconomic datasets (three different vintages of the PWT, including that employed by JMT, and WDI). The analysis will involve two steps. First, I apply the 36 country-specific time series models developed by JMT to the new data without any modification; this is a replication where only the data are altered. While this provides an assessment of the effect of altering data (where in principle each series is measuring the same phenomenon), a guiding principle of CVAR is that the specification and testing should be adapted to capture time series features of the data. To this end, the second step will focus on the subset of 12 countries for which data are available from all four sources over the full period. In the spirit of 'letting the data speak freely' (Hoover et al., 2008), I will specify an appropriate CVAR model for each country-dataset combination, applying the same statistical

criteria as JMT, and complementing it with historical and institutional knowledge for the appropriate modelling of extraordinary events. This can be considered a replication of the method, with the objective of assessing if inferences on the macroeconomic impact of aid in SSA are sensitive to the data source employed.

The growth rates for the recent economic boom of SSA countries mentioned above do not end in 2014 by coincidence. In the past three years, large parts of SSA appear to have lost their economic momentum. According to WDI, per capita incomes still rose by 1.8% in 2014, but only by 0.3% in 2015. In 2016, the last year for which the WDI currently reports growth rates, GDP per capita decreased for the first time in almost twenty years, by a staggering 1.4%.

The causes for this unfortunate backslide are not fully understood, but sharp drops in commodity prices certainly contributed (e.g., Rodrik, 2018). Many economies in SSA are still highly undiversified, and exports are often dominated by a single commodity. For instance, oil typically accounts for 90% of export earnings in Angola, Nigeria, the Republic of Congo, Guinea, Equatorial Guinea, and Sudan; gold accounts for 75% of Malis exports; and Diamonds for more than 60% in Botswana. This commodity-dependence makes those countries particularly vulnerable to external shocks, and poses some peculiar problems when it comes to macroeconomic management, monetary policy in particular. The third and last essay of this thesis addresses this issue through a case-study of Zambian exchange rate and consumer price dynamics, a country where more than 70% of export earnings typically come from copper.

Specifically, we investigate the nexus between the price of copper, the exchange rate, and consumer prices in the period between 1995 and 2014 in the framework of a structural vector autoregression (SVAR) model. In order to identify relevant macroeconomic shocks without making overly restrictive assumptions, we use a combination of sign- and zero-restrictions on the short-run dynamics of the variables. To overcome some pervasive econometric problems associated with estimating parameters in an under-identified system (the result of using sign-restrictions), we will employ state-of-the-art Bayesian estimation techniques put forward by Baumeister and Hamilton (2015). We follow Forbes et al. (2015) in estimating differential exchange rate pass-through (ERPT). Here, the idea is that the transmission between exchange rate fluctuations and consumer prices – the ERPT – can vary greatly depending on the shock that originally caused the exchange rate to fluctuate. The analysis will allow assessment of the macroeconomic effects of commodity prices and although specific to Zambia has implications for macroeconomic management in commodity-dependent SSA countries.

Chapter 2

Blinded by the Light? Heterogeneity in the Luminosity-Growth Nexus and the 'African Growth Miracle'

2.1 Introduction

In 2011, Shanta Devarajan, then the World Bank's Chief Economist for Africa, declared 'Africa's statistical tragedy' (Devarajan, 2013): While the continent finally seemed to have overcome its 'growth tragedy' (Easterly and Levine, 1997), the trouble was that no one could actually be quite sure about it. Poor statistical capacity, paired with the possibility of politically motivated misreporting of economic performance led to widespread concerns about the reliability of GDP figures, especially from sub-Saharan African countries (Jerven, 2013c). The amplitude of the problem is prominently illustrated by the case of Nigeria, where, after an extensive revision of the statistical office's methodology, GDP figures were revised upwards by 60% overnight. Indeed, since 2000, at least 14 African countries have performed a similar move, with revisions averaging around +20% of the initial GDP estimates (NNBS, 2013).

While improving future estimates of economic performance does come with its own challenges (financial, institutional and political), it is a relatively straightforward task from a technical point of view: compliance with international reporting standards, more comprehensive and regular surveys across the economy, and appropriate processing of the data can be implemented where there exists political will and resources are made available. What is inherently more difficult to rectify, however, is information on past growth performances: If Nigeria is now 60%

richer than we used to think, when exactly did this growth come about? Was there indeed an 'African Growth Miracle' in the 1990s and 2000s that escaped traditional measurement methods, as implied by Young (2012)? Or did African countries already start into the 1990s with a substantially higher level of income, and much of the growth that has been attributed to recent periods had actually already happened much earlier in history (e.g., Jerven, 2015)?

The fundamental issue here is that any effort to correct past growth rates using conventional methods will be likely to suffer from at least some of the same deficiencies that also undermined the reliability of the original estimates: Surveys of the economy – provided there is a will for them to be made available – may not have been regularly collected. If beliefs about the structure of the economy at the time of data collection were false, the way in which samples were constructed may not have been representative of the actual structure of the economy. Historical tax records may be inaccurate and by construction only cover the formal sector, which in low-income countries typically accounts only for a small share of the economy. And in case there has been any politically motivated tinkering with the data, data recovered from statistical archives may itself have been subject to manipulation. Statistically speaking, it seems impossible to obtain estimates of past growth rates with errors that are uncorrelated with those of the originally reported growth rates.

Perhaps the most promising and innovative avenue to address this problem is the use of historical satellite imagery, specifically man-made light emissions at night. Several studies have established that there exists a high correlation between economic activity and luminosity, and that lights can therefore serve as a surprisingly powerful proxy of GDP (e.g., Chen and Nordhaus, 2011; Ghosh et al., 2010). Crucially, it can convincingly be argued that the measurement error in light emissions at night is unrelated to the error associated with the figures reported by national statistical offices. Henderson et al. (2012) offer a statistical framework that exploits this property in order to construct estimates of GDP growth with reduced overall measurement error. Combining growth estimates based on luminosity and reported GDP data, they present revised average growth rates for the period between 1992 and 2006 for countries with low statistical capacity. While they cannot discern any systematic pattern of over- or under-reporting in these countries, the suggested revisions are very large for individual economies: In Nigeria and Angola, their growth estimates based on luminosity suggest annual growth rates about half as high as the official ones: 1.92% instead of 4.04% in Nigeria, and 3.88% instead of 6.99% in Angola. On the other hand, some notoriously disappointing growth performances are radically corrected upwards: In Côte d'Ivoire, the Democratic Republic of Congo, and Burundi, the change in luminosity sug-

gests growth rates about 3 percentage points higher than official figures. Based on assumptions about the amplitude of the measurement error in official data, the authors then report weighted averages of the official growth rates and those obtained from luminosity. For countries with supposedly bad data, they suggest an optimal weight of about 50% for the luminosity measure.

The argument put forward in this paper is that the relationship between luminosity and GDP varies across countries, contrary to one of the central assumptions of the literature that exploits this relationship. Under these circumstances, the suggested revisions for individual countries may be highly misleading. To illustrate this, consider Poland and Thailand, two countries that are very different in many respects, but with a similar growth performance over the past two decades and data that is generally considered reliable: While in both countries, GDP approximately increased by 230% between 1992 and 2013, the luminosity emitted by Poland increased by a factor of 2.5, while that of Thailand increased by a factor of almost 5. One possible explanation is of course that at least one of these countries has vastly mis-measured or mis-reported their GDP figures. Perhaps a simpler explanation is that, for some reason, GDP and luminosity interacted differently in these places. For instance, Poland started the period with a much higher level of income, it has a very different location (nearer the pole), a different economic structure, more people living in urban areas, etc.

The aim of this study is to obtain estimates of past GDP growth rates that take into account the heterogeneity in the relationship between luminosity and GDP that can be explained based on such country characteristics. To this end, we introduce an approach that consists in two steps: In a first step, we will estimate country-specific coefficients determining the relationship between GDP and lights[1] – that is, our regression will allow for slopes to vary freely between countries. In a second step, we seek to put some structure in the relationship between luminosity and GDP: we now treat the coefficients obtained in the first exercise as the dependent variable, and seek to identify the part of their variation that can be explained by observable country-characteristics. This allows us to make predictions of GDP based on luminosity values that take into account cross-country heterogeneity in the relationship between these variables, to the extent that this heterogeneity can be traced back to generalisable relationships.

Intuitively speaking, the reasoning behind this two-step procedure is that, while the first step (estimation of country-specific slopes) does indeed allow for full heterogeneity in terms of lights-GDP elasticity across countries, the resulting

[1]To clarify: Our study is concerned with GDP *growth*, not levels of GDP. For the latter, lights are a relatively poor predictor, that is, they do not predict cross-country variation well (Henderson et al., 2012). We do, however, predict relative GDP levels *within* countries, from which growth rates are then derived. See section 2.3.2 for details.

estimates have no value when it comes to revising the reported growth rates. Instead, any predictions of GDP growth based on lights using the parameters estimated in the first step simply tend to replicate the reported growth rates, whether these were mis-reported / mis-measured or not. The second step serves the purpose of identifying the systematic part of the variation between them. We can then predict *expected* elasticities between GDP and lights conditional on any country's characteristics. These conditional elasticities will then form the basis for our predicted GDP growth rates: By allowing for elasticities to vary between countries based on their characteristics, we acknowledge the fact that luminosity doesn't behave the same in every country of the world, and avoid suggesting large revisions where these are unwarranted. But by restricting the heterogeneity to the systematic component that is generalisable across countries, and based on 'legitimate' factors, we avoid the simple replication of potentially misreported growth rates.

One fundamental difficulty we encounter when attempting to discern structure in the elasticities between lights and GDP is that, unlike for most well-researched economic phenomena (say, economic growth), there does not exist any theoretical framework to offer guidance as to which variables may matter, and in what way. That is, any attempt to specify a model that is parsimonious from the outset would necessarily be based on relatively arbitrary decisions, and risk the omission of important predictors. However, in view of predictive accuracy, a certain degree of parsimony is necessary in order to avoid issues of over-fitting. To overcome this problem, we employ elastic net regularisation, and leave-one-out (LOO) cross-validation as a means of obtaining optimal predictive properties. The intuition behind this is simple: At the core, the elastic net is the ordinary least squares estimator (OLS), augmented with penalty terms that have the role of shrinking parameters (or to drop variables, where parameters are shrunk to zero). The severity of these penalty terms is determined by parameters (α and λ) that can be adjusted (*tuned*), typically with the aim of finding the model that yields the most accurate predictions. The role of cross-validation is then to iteratively divide the existing sample into a training-set (used to estimate the model) and a test-set (used to test its predictions), so as to assess every model's *out-of-sample* predictive power. In other words, it finds the model that is most *generalisable* beyond the narrow sample the estimator is presented with. We exploit this property in order to identify the share of the variation in the estimated lights-GDP coefficients that can credibly be attributed to observable country characteristics, and, crucially, is not a result of mis-reported or mis-measured GDP data.

As the asymptotic and finite sample properties of our procedure are unknown *a priori*, we test and demonstrate its potential on simulated data. We create a setup

in which some countries measure and report their GDP accurately, while others systematically misreport it. Luminosity series based on the *true* growth rates are then generated for every country. Crucially, we introduce variation across countries in the elasticity at which GDP induces luminosity. This variation is driven by a range of variables, some of which are simultaneously determinants of GDP growth. The results suggest that, indeed, our suggested methodology has the potential to improve the accuracy of GDP growth estimates based on luminosity. This is mainly because, as opposed to estimators that assume a single elasticity for all countries, our approach is less prone to suggesting revisions where the reported data was correct in the first place. When applying our methodology to the real data, this property appears to hold as well: Our estimator suggests substantially smaller revisions to growth rates where the official data is recognised to be more reliable (as suggested by the Penn World Table's grades of data quality, ranging from A to E). The alternative estimator that imposes a single slope applies larger corrections throughout country grades, suggesting equally large revisions for the countries with the supposedly worst data quality as for those with the best.

We then use our estimates to assess the growth performance of sub-Saharan African countries between 1992 and 2013, and derive estimates of GDP growth rates based on changes in luminosity for each of the African countries in our sample. We find no indications of an 'African Growth Miracle' that would have gone unnoticed by official statistics, as implied by Young (2012): While official growth figures and luminosity based estimates diverge substantially for individual countries, there does not appear to be any particular directionality in this discrepancy. On aggregate across the continent, the official growth rates are quite well aligned with those obtained from the luminosity proxy. On the other hand, it appears as though countries that recently revised their GDP figures after a prolonged period of time (since 2000 and after at least 10 years; this applies to Botswana, Ethiopia, Ghana, Niger and Nigeria) had a tendency to report inflated growth rates for recent years. Between 2003 and 2013, these countries reported annual growth rates of 7.5% on average, while our luminosity based estimates suggest an average growth rate of 5.4%. This pattern is consistent with Jerven (2014)'s hypothesis of 'statistical growth'. According to this narrative, recent improvements in statistical capacity would have revealed a lot of economic activity that had previously gone unnoticed by statistical authorities, and spuriously been accounted for as recent economic growth. Finally, we consider growth rates of individual countries across three sub-periods. The general pattern here is that the most extreme growth performances – stellar growth episodes in individual countries or cataclysmic recessions – tend to appear more moderate seen through the luminosity proxy.

We do consider our results to be an improvement over existing lights-based

estimates of growth rates, but emphasise that – like all results in this literature – they are best considered to be of suggestive nature. Lights are inevitably an imperfect proxy for economic activity, and idiosyncratic differences in the relationship between lights and GDP beyond what we are able to discern in this study, paired with randomness due to measurement error, may be driving substantial shares of the remaining discrepancy between our predictions and reported growth rates. The type of conclusion that we wish to draw from this exercise is not that country X should revise its historical growth rates by precisely amount Y. Instead, we consider our results suggestive of broad patterns, and as a valuable complement to other sources of information, such as traditional types of data and historical records. Furthermore, we stress the importance of the observation that a substantial share of the discrepancy between reported growth rates and those inferred from luminosity can be closed once we take into account the factors that drive this relationship: National accounts data may come with deficiencies, but the risk of overstating those deficiencies is as real as the risk of ignoring them.

The remainder of this paper proceeds as follows: section 2.2 discusses the previous literature on night lights. Section 2.3 describes our methodology. Section 2.4 applies our methodology to simulated data in order to test and illustrate its basic properties. The empirical data we employ and any transformations to it are described in section 2.5. Section 2.6 substantiates the observation that there exists a sizeable amount of heterogeneity between countries in the lights-GDP relationship. In section 2.7, we bring the method to the actual data, and in section 2.8 we apply our estimates to recent growth rates from African countries. Section 2.9 concludes.

2.2 Literature review

Night time luminosity as a proxy for economic activity is generally employed for two purposes: First, because of the resolution at which the data is available, it allows researchers to move beyond the usual administrative units (typically countries) when investigating the determinants and dynamics of economic development (e.g., Alesina et al., 2012; Michalopoulos and Papaioannou, 2013; Hodler and Raschky, 2014). Second, as it is inherently independent of the measurement or reporting error by statistical agencies, it has a potential to augment these statistics and to reduce overall measurement error (Chen and Nordhaus, 2011; Henderson et al., 2012). It is the latter type of application this study is concerned with. This section will provide a brief overview of the relevant literature.

To the best of our knowledge, the link between economic activity and luminosity emitted into space was first noted by Croft (1978). However, the data,

collected by weather satellites of the National Oceanic and Atmospheric Administration (NOAA), were not systematically stored and made available for research until relatively recently. The first studies that systematically investigated and exploited the lights-GDP relationship therefore only emerged in the early to mid-2000s, with Sutton and Costanza (2002), Ebener et al. (2005), Doll et al. (2006) and Sutton et al. (2007) each aiming to provide estimates of incomes at the sub-national level. Ghosh et al. (2009, 2010), while still devoting much of their analysis to sub-national distributions of economic activity, shift the focus to augmenting reported national accounts data; in particular, they provide luminosity based estimates that are meant to incorporate informal economic activity not reflected in official statistics.

A more formalised statistical treatment followed by Chen and Nordhaus (2011) and Henderson et al. (2012). In both studies, the central aim is to augment official income data at the country level with the information derived from luminosity. In the presence of measurement error in official income (alternatively, output) data, and with luminosity as a strong predictor of incomes with errors uncorrelated to those in official statistics, there must be some linear combination of official figures and those derived from lights emissions that minimises the overall measurement error. In Henderson et al. (2012)'s notation:

$$\hat{y}_i = \lambda z_i + (1 - \lambda)\hat{z}_i$$

where z_i are the reported official growth figures, $\hat{z}_i$ is the luminosity based proxy, and $\hat{y}_i$ is the new synthetic measure of income (output) growth in country (or region) i. The weight on the luminosity based proxy, $(1 - \lambda)$, must then be chosen such that, on expectation, the overall measurement error is minimised.

In their version of the analysis, Chen and Nordhaus (2011) derive such optimal weights both for output levels and long-term growth rates, separately for countries with different levels of statistical quality (A–E as classified in older versions of the Penn World Tables (PWT)). Their findings suggest that luminosity adds value for countries with low statistical quality (ratings D and E), and that its value (as quantified by the optimal weights $(1 - \lambda)$, or θ in their notation) is typically higher when considering long-run growth rates rather than output levels. The optimal weights they derive for the luminosity proxy reach just over 30% for long-run growth rates in countries rated D, and less for other countries. Henderson et al. (2012) focus on growth rates entirely. Instead of using the A–E rating from the PWT, the authors rely on the World Bank's rating of statistical capacity and divide countries into only two categories, *good data* and *bad data* countries (*bad data* being defined as scoring less than 3 out of 10 in the statistical capacity

rating). The optimal relative weights of lights and GDP for good and bad data countries crucially rely on assumptions made about the signal to variance ratio in reported GDP growth rates from *good data* countries. The authors' preferred choice, assuming 90% signal to variance in good data countries, implies a weight of about 52% for lights for the 30 countries they qualify as having bad data, and still 15% for those with good data.

Based on these weights, Henderson et al. (2012) compute the optimal GDP growth rates for the 30 *bad data* countries in their sample (optimal in the sense of minimising the expected total measurement error). The first important observation is that they do not find evidence of systematic over- or under-reporting of growth rates across *bad data* countries. Instead, the discrepancies between official growth rates and the lights-proxy average around zero. However, many of the revisions suggested by their results are substantial: In Burundi, where official figures suggest a negative average growth rate of -0.71% annually between 1992 and 2006, the lights proxy suggests positive growth of 2.89% annually; weighted optimally, the suggested average annual growth rate is 1.13%. At the other end of the spectrum (and focussing on the African countries in the sample), Angola's growth rate is revised downwards from 6.99% annually to 5.38% (with lights suggesting 3.88%), and Nigeria's from 4.04% to 2.94% (with lights suggesting 1.92%).

A handful of studies have since been dedicated to gaining a deeper understanding of the relationship between economic activity and lights. Using county-level GDP series from Brazil, India, the United States and Western Europe, Bickenbach et al. (2016) show that at a sub-national level, there exists large variability in the relationship between lights and GDP. Within all of the economies they consider, they find that the estimated elasticity between lights and GDP differs substantially across regions. For instance, when regressing growth in lights on growth in GDP, they report a small but significant positive coefficient on lights of 0.09 in East India, but a *negative* significant one in West India (-0.14). For Brazil, their estimates across five regions range between 0.37 (Centro Oeste) and 0.04 (Sul). Similar discrepancies are observed in the USA and Western Europe, where the reported GDP data are likely to be particularly accurate. The authors explore several avenues to restoring parameter consistency (mainly through the inclusion of a range of control variables, and restricting the sample in a number of ways), but none of these succeed. They conclude that, at the regional level, luminosity does not appear to be a useful proxy of economic growth. However, as the authors note, their findings do not necessarily translate to the national level, where parameters may still be constant.

Wu et al. (2013) explicitly seek to identify factors affecting the relationship between lights and GDP at the country level. Their analysis differs in one important

aspect from the bulk of the literature: Instead on regressing lights on GDP, which studies that are concerned with proxying for GDP typically do, they employ lights as the dependent variable and regress a number of potential determinants on it. Their choices of variables derive from a rudimentary theoretical framework that models lights as a normal consumption good. On this basis they hypothesise that, beyond incomes, luminosity may be affected by factors such as the share of agriculture, savings, and latitude. Empirically, their model translates into a standard linear regression with luminosity on the left hand side, and each of its hypothesised determinants (including GDP) on the right hand side. GDP per capita, latitude, the degree of spatial agglomeration and the savings rate are found to be significant determinants of luminosity. Additionally, they separate GDP into its agricultural and non-agricultural component, finding that non-agricultural production is a much stronger driver of luminosity than agricultural production.[2] In a conceptually similar way, Levin and Zhang (2017) investigate the correlates of luminosity in urban areas using the newer and higher resolution VIIRS luminosity data.[3] They identify numerous variables (anthropogenic and physical) that affect the luminosity of urban areas. The strongest and most persistent correlates beyond incomes are found to be snow cover (especially during winter months), rents from fossil fuels, latitude and road density.

Studies of this type deliver useful insights as to what drives luminosity beyond only GDP; it is important to highlight, however, that effectively they investigate the determinants of *luminosity*, as opposed to the determinants of the *relationship* between luminosity and GDP. The study at hand seeks to specifically address the latter. Furthermore, beyond simply exploring factors that may alter the way in which economic growth translates into changes in luminosity, we will offer an empirical framework to improve the predictive power of the lights-based GDP proxy in the presence of this type of heterogeneity.

[2]Keola et al. (2015) make a similar observation about agriculture not being reflected as distinctly by luminosity as other sectors, but to a more problematic degree than that observed by Wu et al. (2013). In order to overcome the issue, they suggest combining luminosity data with data on landcover, where growth in agricultural production is approximated based on changes in cultivated land.

[3]The recently launched VIIRS series is superior to the DMSP-OLS series employed in this study in a number of respects (Dai et al., 2017). However, data is only available from 2012 onwards, making the investigation of historical growth rates impossible.

2.3 Methodology: Addressing heterogeneity in the relationship between GDP and lights

The literature that exploits luminosity as a proxy for economic activity typically assumes a relationship (implicitly or explicitly) where economic activity generates lights at some constant elasticity (e.g., Chen and Nordhaus, 2011; Henderson et al., 2012; Keola et al., 2015). This elasticity is generally assumed to be the same across all countries, an assumption this study aims to relax.

2.3.1 Conceptual framework

Our version of the framework therefore describes a basic relationship where economic activity (Y) generates luminosity (L) at some *country-specific* elasticity β_i:

$$L_{it} = Y_{it}^{\beta_i} * exp(\varepsilon_{it}).$$

The relationship is perturbated by an error term ε_{it}, which we note for later has the structure $\varepsilon_{it} = \epsilon_i + \epsilon_t + \epsilon_{it}$, that is, the error term is a composite of time-specific perturbations in year $t = 1, ..., T$ (e.g., because of degrading and replaced lights-sensors over time), country-specific perturbations in country $i = 1, ..., N$ (e.g., different levels of light emission at the baseline), and an idiosyncratic error term ϵ_{it}. Generally, $E[\varepsilon_{it}] = E[\epsilon_i] = E[\epsilon_t] = E[\epsilon_{it}] = 0$. The country-specific elasticity β_i quantifies the extent to which economic activity translates into luminosity for each country. Taking logarithms, equation 2.3.1 can be rewritten in linear form:

$$\ell_{it} = \beta_i y_{it} + \varepsilon_{it}$$

with $\ell = ln(L)$ and $y = ln(Y)$. Since we are interested in predicting GDP from luminosity, we rearrange this to

$$y_{it} = \gamma_i \ell_{it} - \gamma_i \varepsilon_{it}, \qquad \gamma_i = \frac{1}{\beta_i}.$$

Since $E[\varepsilon_{it}] = 0$, and ℓ_{it} is fixed, $E[Y_{it}] = exp(\gamma_i \ell_{it})$, for some given inverse elasticity γ_i in country i. This relationship can then be exploited in order to approximate (true) GDP Y in places where there is uncertainty whether the reported data is accurate, or where such data is unavailable. Note that the focus lies in growth rates (i.e., relative changes of Y over time *within* a given country), so the absence of an intercept will have no bearing on the results of interest. We consider the inverse elasticities γ_i to be a function of a set of determinants that vary at the country level:

$$\gamma_i = f(\Phi_i, \Psi_i, \eta_i)$$

where Φ_i are observable determinants, Ψ_i are unobservable determinants, and η_i is a random variable.[4] We deliberately do not attribute a specific functional form to $f(.)$, reflecting the fact that there is little theoretical understanding on how exactly luminosity and GDP interact, and that, from an economic point of view, there is little interest in describing it *per se*. Instead, our interest lies in exploiting the aggregate manifestation of this relationship for predictive purposes. To this end, we will seek to capture the share of the variation in γ that can be attributed to the observables Φ, while remaining widely agnostic about the underlying mechanisms.

2.3.2 Empirical approach: A procedure in two steps

The broad strategy that we adopt consists in two steps: First, as γ_i are unknown, we obtain estimates $\hat{\gamma}_i$ from a fixed effects regression that allows for country-specific slopes (equation 2.1 below). Second, we employ the flexible elastic net estimator to discern the part of the variation in the $\hat{\gamma}_i$s that can be attributed to observables (Φ_i) in a way that is generalisable across countries. This will allow us to predict *expected* inverse elasticities $\tilde{\gamma}_i$, conditional on observable country characteristics: $\tilde{\gamma}_i = E[\gamma_i|\Phi_i]$. These can, in turn, be used to predict Y_{it}, on the basis of which we can derive estimated growth rates.

Step 1: Estimating naive country-specific slopes ($\hat{\gamma}_i$): We obtain our first stage estimates, $\hat{\gamma}_i$ by estimating the following equation using least squares:

$$z_{it} = \sum_{i=1}^{N} \gamma_i(\ell_{it} \times D_i) + \alpha_i + \alpha_t + \varepsilon_{it} \tag{2.1}$$

where z_{it} and ℓ_{it} are the logs of reported GDP figures and luminosity in country i at time t respectively, D_i are country dummies, and α_i and α_t are country and year fixed effects. The time fixed effects are mainly necessary in order to account for differences in sensor sensitivity across years (see data description in section 2.5). The country fixed effects serve to absorb differences in levels across countries that are constant over time, such as different baseline levels of brightness and differences in units (GDP will be included in local currency units). In order to estimate the parameters $\hat{\gamma}_i$, we therefore exploit the temporal variation of luminosity and GDP

[4]For all intents and purposes, Ψ_i and η_i can be considered the same thing. The differentiation is merely to indicate that we neither believe that we can empirically capture all factors that do affect the relationship between lights and GDP, nor that it is entirely deterministic.

within each country.

Step 2: Explaining $\hat{\gamma}_i$ using the elastic net: The main challenge in our second step is to select and weight relevant predictors of $\hat{\gamma}$ from a large set of candidate variables. At the core, this is because of the inherent lack of knowledge about $f(.)$: we have very little guidance when it comes to selecting the variables to include in this predictive exercise (potential elements of Φ, labelled Φ^*). Econometrically, this creates a situation with a large number of variables compared to the number of observations (N countries). In order to avoid over-fitting – that is, in order to identify the determinants that apply in a generalisable manner and beyond our narrow sample – we make use of elastic net regularisation (Zou and Hastie, 2005). The elastic net estimator is suited to situations with large amounts of explanatory variables compared to the number of observations, and has been shown to perform very well in selecting the relevant predictors. It minimises the following criterion:

$$L(\lambda_1, \lambda_2, \boldsymbol{\delta}) = |\hat{\boldsymbol{\gamma}} - \boldsymbol{\Phi^*\delta}|^2 + \lambda_2|\boldsymbol{\delta}|^2 + \lambda_1|\boldsymbol{\delta}|_1 \tag{2.2}$$

with

$$|\boldsymbol{\delta}|^2 = \sum_{j=1}^{p} \delta_j^2,$$

$$|\boldsymbol{\delta}|_1 = \sum_{j=1}^{p} |\delta_j|$$

.

where $\boldsymbol{\delta}$ is the vector of coefficients (weights) attributed to the elements of $\boldsymbol{\Phi^*}$, which in turn is the set of potential determinants of γ we take into consideration.[5] $\hat{\boldsymbol{\gamma}}$ is the dependent variable, that is, a vector containing the N $\hat{\gamma}_i$s obtained from estimating equation 2.1 above. Equation 2.2 is best discussed by considering the terms seperately. The first term, $|\hat{\boldsymbol{\gamma}} - \boldsymbol{\Phi^*\delta}|^2$ is the square of the residuals, reflecting the same minimalisation problem as in OLS. The second term, $\lambda_2|\boldsymbol{\delta}|^2 = \lambda_2 \sum_{j=1}^{p} \delta_j^2$, is a penalty term, penalising for large coefficients. It is in fact the same term as in a ridge regression (Hoerl and Kennard, 1970). The third term, $\lambda_1|\boldsymbol{\delta}|_1 = \lambda_1 \sum_{j=1}^{p} |\delta_j|$, is another penalty term, and corresponds to the one employed in a lasso regression (Tibshirani, 1996). In fact, Zou and Hastie (2005) show that the minimisation of $L(\lambda_1, \lambda_2, \boldsymbol{\delta})$ can be expressed as the optimisation problem

[5] Ideally, the weights $\delta = 0$ for elements of Φ^* that are not part of Φ.

$$\hat{\boldsymbol{\delta}} = arg\ \underset{\boldsymbol{\delta}}{min}|\hat{\boldsymbol{\gamma}} - \boldsymbol{\Phi}^{\boldsymbol{*}}\boldsymbol{\delta}|^2, \qquad s.t.\ \alpha|\boldsymbol{\delta}|_1 + (1-\alpha)|\boldsymbol{\delta}|^2 < s \text{ for some } s \tag{2.3}$$

with $(1-\alpha) = \lambda_2/(\lambda_1 + \lambda_2)$. In that sense, the elastic net penalty is a convex combination of the lasso and the ridge penalty, with α determining the weight between the two. In fact, both the lasso and the ridge estimator are nested in the elastic net, as $\alpha = 1$ corresponds to the lasso, while $\alpha = 0$ corresponds to ridge. Both α and λ (or, via algebraic detours, s) will then be calibrated using n-fold cross-validation in order to minimise the out of sample prediction error (as measured by the root mean squared error; see appendix A.6 for a more detailed discussion).

The elastic net estimator is particularly well suited for our application for a number of reasons. First, it performs particularly well in the presence of large numbers of predictors compared to the number of observations; in our empirical application, we will include 57 variables for 129 observations (cf. section 2.7). Second, as opposed to ridge, it has the lasso's characteristic of selecting variables (setting weights of poor predictors to zero). This facilitates the interpretation of the output.[6] Third, compared to the lasso, the results and predictions from the elastic net tend to be more stable, especially if some of the independent variables are strongly correlated. Fourth, lasso and ridge – the main contenders for the choice of estimator – are nested in it, so if any of them is superior in terms of out-of-sample predictive power, it will be selected ($\alpha = 0$ or $\alpha = 1$).[7]

Based on the estimated weights $\boldsymbol{\delta}$ and the known values of $\boldsymbol{\Phi}^{\boldsymbol{*}}$, we can then obtain our predicted parameters $\tilde{\gamma}_i = E[\gamma_i|\Phi_i^*]$ and fit model 2.1 using $\tilde{\gamma}_i$ rather than $\hat{\gamma}_i$. The result is, for each country, a full time series of predicted GDP values based on luminosity emissions, where luminosity enters with a weight that is consistent with the country's characteristics ($\tilde{Y}_{it} = E[Y_{it}|\Phi_i^*, \ell_{it}]$). Growth rates are then derived from the annual changes in these predicted values in the usual manner ($\tilde{g}_{it} = \frac{\tilde{Y}_{it} - \tilde{Y}_{it-1}}{\tilde{Y}_{it-1}}$).

[6] Although we do not focus on the interpretation of the output *per se*, as our focus is on predictive power. See Mullainathan and Spiess (2017) for a discussion of the interpretability of output from machine learning procedures in general.

[7] Another class of variable selection procedures we considered were step-wise *General to Specific* approaches, e.g. Hoover and Perez (2004), which are common in the cross-country growth literature. The general procedure there would be to run conventional OLS regressions, and to iteratively eliminate variables based on some measure of explanatory power, e.g., their t-ratios. We found these to be very sensitive even to small changes (like the order in which the variables are included) in terms of selected variables as well as in terms of predicted values. Overmore, they had a strong tendency to overfit.

2.3.3 Alternative estimators

While the main focus of our analysis lies on the results obtained using the two-step procedure outlined above, we do use alternative modes of estimation for descriptive purposes, for comparability with the earlier literature, and as a robustness check. We briefly introduce and discuss the respective equations in this subsection, all of which will be estimated using least squares.

In order to explore some patterns of the heterogeneity of γ in section 2.6, we will divide our sample into G groups that are relatively homogeneous with respect to selected characteristics. We then impose a single slope coefficient within each of these groups, estimated via

$$z_{it} = \sum_{i=1}^{G} \gamma_g(\ell_{it} \times D_g) + \alpha_i + \alpha_t + \varepsilon_{it}, \tag{2.4}$$

where D_g are dummy variables indicating membership to group g. We also estimate a conventional fixed effects model that imposes a single inverse elasticity γ for all countries:

$$z_{it} = \gamma x_{it} + \alpha_i + \alpha_t + \varepsilon_{it}. \tag{2.5}$$

This is the model employed by the bulk of the literature, including Henderson et al. (2012) in their baseline estimations. Throughout the paper, equation 2.5 will be treated as our main benchmark specification. As a robustness check and for comparability, our results in section 2.8 will also be derived on the basis of this specification.

Finally, note that when Henderson et al. (2012) proceed to inferring long-run growth rates to augment the GDP data, they estimate a substantially simplified cross-sectional equation. In the context of their study, this maintains tractability of the statistical framework, and enables them to compute optimal weights for the official data and the lights predictions (see section 2.2). We do not engage in such computations, but rather seek to improve the accuracy of the predicted growth rates directly. Nevertheless, we present basic results (estimates of long term growth rates over the entire period) inferred from their simplified model. To this end, we follow them in averaging our variables over the first and last two years in the sample period (1992/93 and 2012/13), and take the log-difference between these periods. The estimated equation is then

$$z_i^{LR} = \alpha + \ell_i^{LR} + \varepsilon_i, \tag{2.6}$$

where the superscript LR indicates long run growth rates as described.

2.4 Simulation exercise

The asymptotic and finite sample properties of the methodology outlined in section 2.3 are unknown. This is partly due to the econometrically unconventional challenge that we are facing: We effectively seek to reduce measurement error in our target variable (GDP growth) using a proxy that has an independent measurement error (lights); at the same time, we wish to account for heterogeneity in the relationship between this proxy and the measure of interest. Moreover, our methodological strategy partly relies on machine learning techniques that have no analytical solution. In order to assess the basic properties of our suggested methodology, this section will therefore present a simple simulation exercise. Section 2.4.1 provides a brief description of the data generating process (DGP), which is elaborated in appendix A.2. Sections 2.4.2 and 2.4.3 illustrate different estimation strategies (long difference, fixed effects with common slope, and our two-stage procedure described in section 2.3) on a single simulated dataset. In section 2.4.4, we repeat this exercise under 625 different parameter combinations on 31,250 simulated datasets in a simple Monte Carlo type simulation, and discuss the key insights about our estimator's performance.

2.4.1 Data generating process

Our data generating process (DGP) is designed to reflect the econometric challenges we are facing, while being tractable in its statistical properties. The simulated dataset needs to feature the following series: (i) a true measure Y ('GDP') (ii) an indicator plagued with measurement error, Z ('reported GDP'), (iii) a proxy for Y that has a measurement error independent of that in Z, ℓ ('lights'), and (iv) determinants (analogous to Φ) and potential determinants (Φ^*) of how Y and L interact in every unit of observation i ('country'); that is, determinants of γ. We design our DGP to exhibit the following features:

1. The resulting series are designed to have econometric properties similar to the observed data; this concerns time series behaviour, orders of magnitudes and distributions of growth rates, and measurement error.

2. We introduce a set of countries with accurately reported data, and a set of countries with poorly reported data. The latter category is modelled to systematically report biased growth rates; the bias in any bad data country's growth rates is determined at random according to a uniform distribution. Depending on the specific parameters, this distribution can imply a tendency for bad data countries to over-report or under-report growth, or for the bias to be centred around zero (no directional bias overall).

3. The elasticity with which y translates into ℓ, γ, is designed to vary between countries based on a set of determinant variables; our estimator will crucially face the challenge of discerning those determinant variables from irrelevant variables.

4. As a potential source of confusion for the estimator, we introduce some overlap between the determinants of y and the determinants of the elasticity γ; for instance, investment levels may both affect GDP growth and the lights-GDP elasticity.

Table 2.1: Variable parameters in DGP

Parameter	Description	Example	MC min	MC max
N_{bad}	Number of bad data countries	20	10	30
SD_ϵ	Error in lights-GDP relationship	μ_g	$\mu_g/0.5$	$\mu_g/2$
Range misrep.	Range of misreporting	$[-0.02; 0.03]$	$[-0.05; 0]$	$[0; 0.05]$
SD_γ	Randomness in inverse elasticities	0.05	0	0.1

Notes: This table summarises the parameters chosen for the DGP both in the illustrative example above, as well as the range iterated though in the Monte Carlo type simulation. All parameters are iterated through over an equally spaced grid with 5 values ranging from MC min to MC max. μ_g is the respective country's mean growth rate.

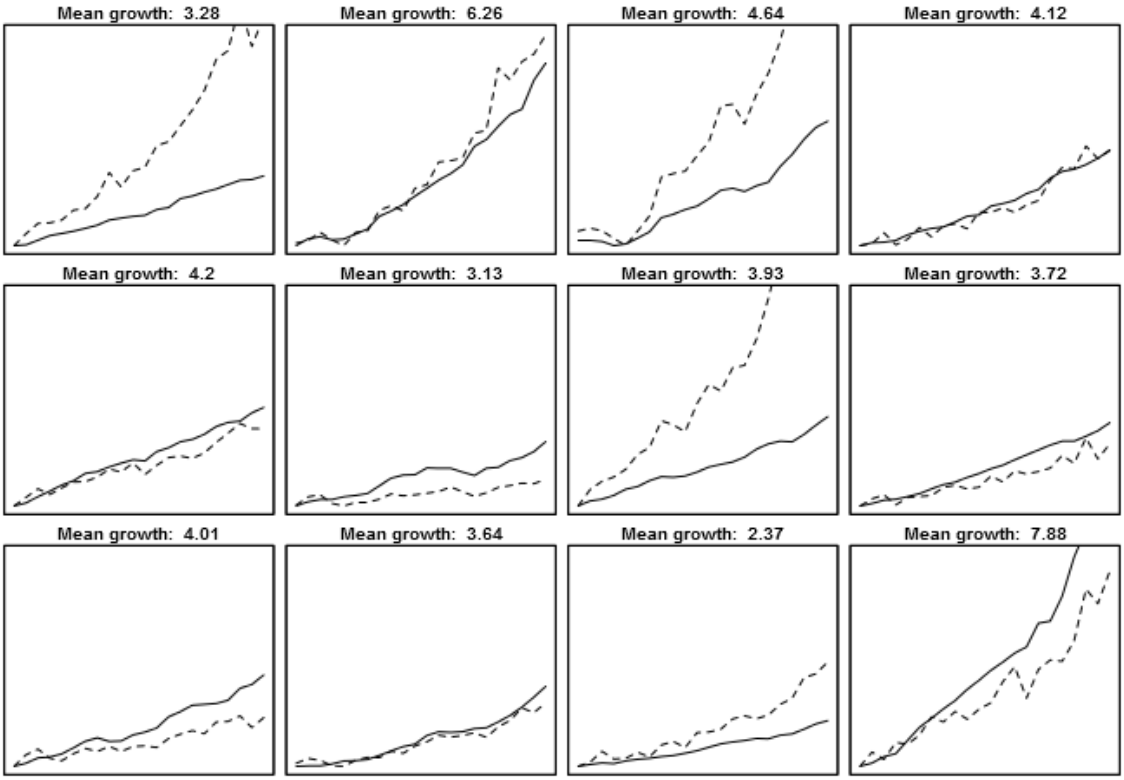
Appendix A.2 provides a detailed summary of the construction of the series (i)–(iv) designed to fulfil criteria 1–4. The dataset generated for this exercise comprises $N = 150$ units ('countries') over $T = 22$ periods ('years'); figure 2.1 illustrates the resulting GDP and luminosity series (solid and dashed lines respectively) by plotting them for 12 randomly selected units of the dataset we use to illustrate our procedure in sections 2.4.2 and 2.4.3. The key parameters are reported in table 2.1, along with the range of parameters that will be explored in section 2.4.4.

2.4.2 Benchmark estimations

We compare the predictive qualities of our two-step estimator to two more conventional estimators that have been employed in the literature, specified in equations 2.6 and 2.5 in section 2.3.3. Table 2.2 reports the estimated coefficients and prediction errors for the single simulated dataset we analyse in this section for illustrative purposes across different methods.

Long differences Henderson et al. (2012) obtain their estimates of long-run growth rates (average annual growth between 1992/93 and 2005/06) from a simplified, cross-sectional regression (equation 2.6 in section 2.3). On the left hand

Figure 2.1: 12 simulated GDP and lights series



Notes: The solid lines depict simulated GDP series, the dashed lines the corresponding proxy L as generated with the DGP described above for 12 randomly selected units.

Table 2.2: Predictive performance of estimators (example)

	Long diff.	FE (single slope)	Two-stage
Estimated γ (true: 1.05)	0.24	0.29	0.57
Mean prediction error of average growth rate			
Good data countries	1.01%	1.03%	0.66%
Bad data countries	1.15%	1.18%	1.04%
Total	1.03%	1.05%	0.70%

Notes: The estimated γ in the two-stage method refers to the mean across the N $\hat{\gamma}_i$s. The prediction errors are mean absolute deviations from the true average growth rate in each country across the entire period.

side, they include the log-difference of GDP between the first and the last period, on the right hand side an intercept and the log-difference of luminosity between the first and the last period. Based on this regression, they then predict the long-run growth rate for every country based on luminosity. They then report annualised growth rates derived from these values.

This drastic reduction of the dataset is necessary for their statistical framework to be applicable (or remain tractable), where one key ambition is to derive optimal

weights for the luminosity based series and official GDP data. In this study, we are not concerned with deriving such optimal weights, but rather with improving the lights-based estimates directly by addressing slope heterogeneity. The framework we suggest to this end crucially relies on the temporal variation *within* countries over time, and we cannot (and need not) base our estimates on a purely cross-sectional dataset. However, for direct comparison, we estimate Henderson et al. (2012)'s specification on our simulated data. The estimated inverse elasticity $\hat{\gamma}^{LR}$ is 0.24 (compared to a true value of 1.05 on average). The predicted average annual growth rates are, on average, off by about 1.03% in good data countries, and by 1.15% in bad data countries. This is an improvement for the bad data countries, where the reported GDP growth deviates from its true value by 1.54% on average in this dataset. However, the prediction error is of a similar amplitude for countries where the data is perfectly accurate, effectively deteriorating growth estimates in these places.

Fixed Effects with single slope Next, we turn to the panel fixed effects estimation with a common slope $\hat{\gamma}$, equation 2.5. This is most similar to what most of the empirical literature that uses night lights as a proxy of economic activity employs (e.g., Keola et al., 2015; Bickenbach et al., 2016); in Henderson et al. (2012), the baseline estimates are derived from this specification. As it is better comparable to our specification with country-specific slopes, we consider this specification our main benchmark. In order to derive growth rates from the resulting estimates, we then produce predictions in levels, derive annual growth rates ($\frac{GDP_t - GDP_{t-1}}{GDP_{t-1}}$), and average these over all periods for each country (see section 2.3.2).

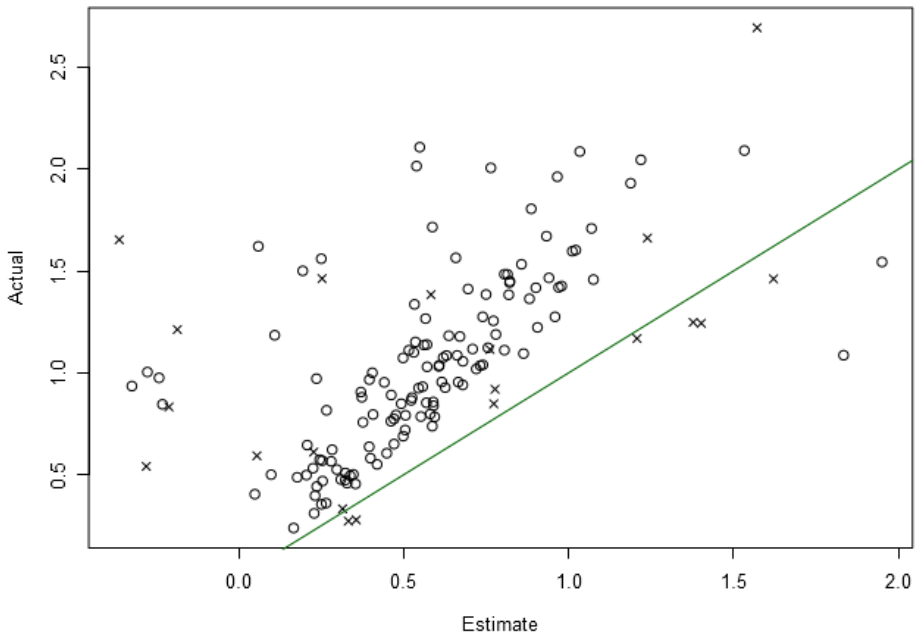
The estimated coefficient $\hat{\gamma}$ is 0.29, again substantially smaller than the true average coefficient of 1.05. In the presence of measurement error, we would indeed expect the measured coefficient to be smaller than the structural one due to attenuation bias. It is worth noting that a large portion of the bias still persists even when we remove every source of measurement error from our DGP. In this particular example, the absence of any measurement error in the DGP (leaving everything else constant) leads to an even *less* accurate estimate of γ (0.25). This is consistent with the finding that – where the heterogeneity in the slope is correlated with the independent variable (as it is in our DGP, and as we suspect it to be the case in the empirical data), a substantial aggregation bias can result from erroneously imposing a single slope across all units of observation (Ul Haque et al., 1999). Note that the direction of the bias depends on the sign of the correlation between the independent variable and the slopes, and may as well be positive.

From a predictive perspective, the fixed effects estimates with a common slope are roughly at par with the long differences: On average, they are off by 1.03% in

good data countries, and by 1.18% in bad data countries. As with the long differences method, this is an improvement for bad data countries, but with revisions being almost equally substantial for good data countries.

2.4.3 Two-step procedure

Figure 2.2: Actual versus estimated γ (simulated data)



Notes: Coefficients obtained from estamating equation 2.1 on the simulated data plotted against the true coefficients. ∘ indicate good data countries, × are bad data countries. The green line indicates $\hat{\gamma}_i = \gamma_i$. Attenuation bias from measurement error means the estimated values are systematically below the actual ones; in the absence of measurement error, the green line and the datapoints would entirely coincide.

Step 1: Estimating naive country-specific slopes ($\hat{\gamma}$): In the first step, we aim to derive country-specific slopes for each of the countries in the dataset. These estimates are naive in the sense that they take the reported data at face value, irrespective of whether these estimates are biased by any mis-measurement or misreporting of GDP. We obtain our (naive) country-specific slopes $\hat{\gamma}_i$ from estimating equation 2.1 – the fixed effects regression from above, augmented by $\ell_{it} \times D_i$, that is, N interaction terms between lights and the country dummies. Figure 2.2 plots the estimated coefficients against the true values from the simulated data. The circles indicate good data countries, the crosses indicate bad data countries as defined above. The green diagonal line indicates $\hat{\gamma}_i = \gamma_i$, that is, the closer an estimate is to the line, the more precise it is. Attenuation bias due to measurement error biases the estimates down compared to their true values, that

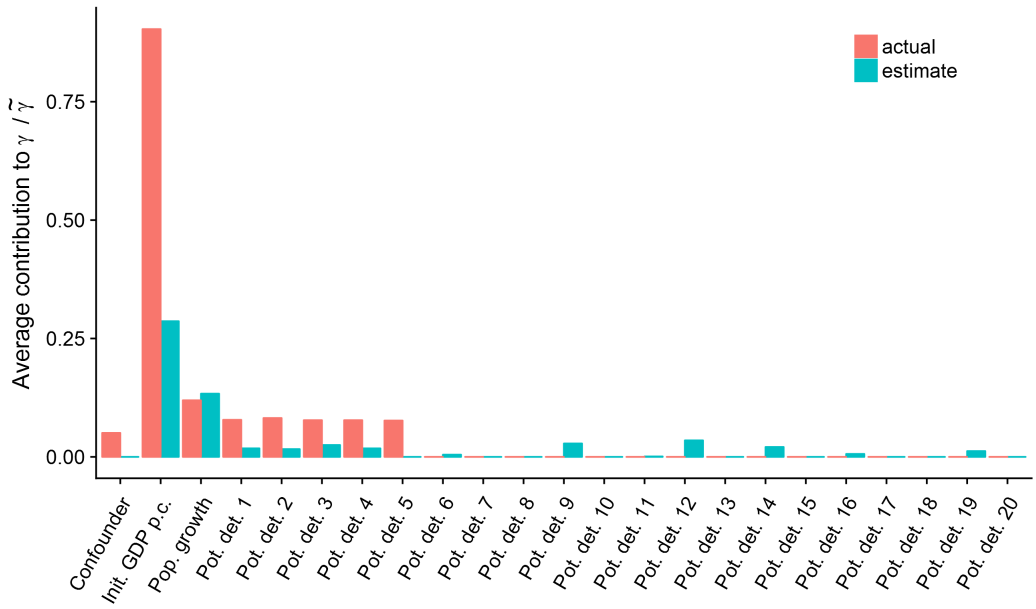
is, they are plotted left of the green line in figure 2.2. At least in this example however, this bias is weaker than the aggregation bias occuring where a single slope is erroneously assumed, like in the estimators above. Note also that, due to measurement error, a few $\hat{\gamma}_i$ are negative, even though there are no true negative values of γ.[8]

On average, the luminosity coefficients obtained from estimating equation 2.1 are 0.57, much closer to the true average of 1.05 than both methods that impose a common slope (setting all measurement error to zero in our DGP would yield $E[\hat{\gamma}] = E[\gamma] = 1.05$ in this example, with all estimated slopes being accurate; there is no aggregation bias). Using the country-specific $\hat{\gamma}$ as estimated to predict GDP growth from luminosity does, however, by construction not have much potential to improve estimates from bad data countries. Rather, it essentially reproduces reported growth rates: For the countries with accurately reported GDP, the predicted growth rates are only 0.26% off compared to the true (and reported) growth rates. For countries with systematically misreported GDPs (bad data countries) it is off by 1.22% on average, which is worse than the previously discussed estimators. The aim of of the next section will be to discern the systematic component in the variation in γ and improve the predictions for countries where data is poor.

Step 2: Explaining $\hat{\gamma}_i$ using the elastic net In order to avoid simple replication of the reported growth rates by taking the country-specific slopes at face-value, we will now employ the elastic net estimator in order to discern the systematic component in the variation in slopes between countries. In the DGP described in section 2.4.1 and specified in appendix A.2, γ had 8 determinants in total: 3 of them are simultaneously determinants of economic growth, and 5 others are unrelated. We also generated another 15 random variables that do not actually affect γ, but that we will consider *potential* determinants. All of these variables are included as independent variables in the elastic net estimator, with $\hat{\gamma}$ as the dependent variable. As discussed in section 2.3 and further in appendix A.6, the parameters of the elastic net estimator are then tuned to attribute weights to the variables that minimise the out-of-sample prediction error. The aim of this second step estimation is therefore to discern the relevant predictors Φ from the pool of po-

[8]It should be noted that the possibility of negative coefficients points at a limitation of the conceptual framework adopted by most of the literature, including the study at hand: If the coefficients obtained are strictly interpreted as the inverse elasticity $\gamma = \frac{1}{\beta}$ with $L = Y^{\beta}$, then a positive $\hat{\gamma}$ near zero would imply a structural elasticity β near *positive* infinity. A negative $\hat{\gamma}$ that approaches zero, however, would imply a structural elasticity β that approaches *negative* infinity. We note this as a caveat from the narrative point of view, and where this is central to the analysis it would be desirable to introduce a functional form that does not have such a drastic discontinuity. In the study at hand, however, the purpose is purely predictive, with the relevant parameter being γ.

Figure 2.3: Actual and estimated determinants of γ



Notes: The bars depict the absolute contributions of each of the determinants of γ. The red bars indicate the actual contributions, known from our DGP, the blue bars are the contributions to the predicted $\tilde{\gamma}$, based on the elastic net estimation. Note that the typical importance of initial GDP appears inflated in this representation, as the mean is driven by some outliers due to the exponential distribution we chose for the variable.

tential predictors Φ^*, and to attribute optimal weights to each of them. These can then be employed to derive estimates of γ_i *conditional on country-characteristics* Φ_i, which we label $\tilde{\gamma}_i$.

Figure 2.3 depicts the average composition of the γs in red, and the estimated contributions in blue. This is quantified by multiplying the estimated weights $\hat{\delta}$ with the values of Φ_i^* for each country, and then averaging these across countries. If the elastic net was doing a perfect job in identifying the predictors and their respective weights, the bars would therefore be identical. However, as we introduced measurement error in our DGP, and the estimates of γ_i, $\hat{\gamma}_i$, are only approximations of the true values (figure 2.2) the elastic net will naturally only capture parts of the structure. In this example, 6 out of 8 elements of Φ are correctly identified, and 3 variables are falsely attributed non-negligible weights. Note that the accuracy of these predictions strongly hinges on the amplitude of the diverse sources of measurement error in our DGP: Setting them to zero completely aligns true and estimated determinants; with increasing error, the elastic net attributes smaller and smaller weights to the determinants, and – as a tendency – converges to the unconditional mean. The R^2 of the second stage in this illustrative example is 50%, slightly lower than the 55% we will obtain in our empirical application.

Discerning the determinants of γ, however, is mainly an auxiliary goal in our strategy. Ultimately, we aim to improve the predicted GDP growth rates from

changes in luminosity. Indeed, it appears as though we achieve this goal: In the present example, the predicted GDP growth rates for good data countries are, on average, 0.70% off when compared to the true growth rates; for good data countries, the mean error is about 0.66%, a substantial improvement over the other estimators. For countries with poor data the improvement is still present but less substantial, with predictions now deviating from actual growth rates by 1.04% on average.

2.4.4 A simple Monte-Carlo simulation

The example presented above refers to a single iteration using a particular combination of parameters. While any artificial DGP will only provide a very stylised view on the matter, we will now repeat the exercise above using several parameter combinations for the generated data. In order to keep computing time at reasonable levels and the exercise tractable, we focus our attention on four key parameters.[9] For each parameter, we iterate through a grid of 5 equally spaced values, which leads to a total of 625 parameter combinations. Each of these we repeat 50 times, leading to 31,250 iterations in total. The parameters we vary are:

- Number of countries with flawed data (N_{bad}): When deriving the determinants of γ, our setup treats all countries equally, which avoids any *a priori* assumptions about the quality of their data. The larger the number of bad data countries, the more influential these observations become and potentially undermine the precision of our estimates.

- Error in lights-GDP relationship (SD_ϵ): The precision of any estimates will obviously be influenced by the signal-to-noise ratio in our proxy variable (luminosity). We therefore vary the standard deviation of the error term ϵ_{it} in the relationship between lights and GDP.

- Bias in bad data countries (*Range misrep.*): Misreporting is modelled as the systematic misreporting of growth figures within a country, where the amplitude of the misreporting is a (uniformly distributed) random number. We shift the limits of this distribution, so as to cover the hypothetical case where bad data countries generally *under*report growth rates (e.g, their statistical offices fail to pick it up), the case where they *over*report (e.g., they

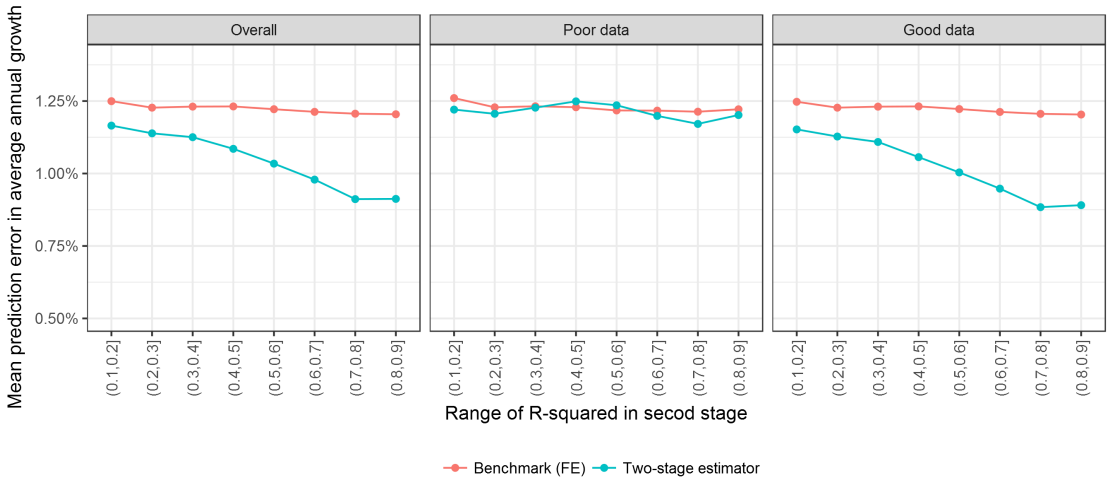
[9]The elastic net is a computationally demanding procedure, and one iteration of the procedure outlined above (including data generation, estimation and prediction) takes about 1.4 seconds to carry out on the computer that we have at our disposal (including λ-tuning, but excluding α-tuning). In order to obtain reliable indicators of the estimators' expected performance given any parameter combination, we run 50 repetitions per parameter combination. In total, the simulation takes approximately 12 hours to run in this specification.

tend to manipulate figures upwards), and the case where misreporting has no direction in particular.

- Randomness in γ (SD_γ): The (inverse) elasticity of the relationship between GDP and lights is unlikely to be fully deterministic, and not all of the determinants are necessarily observable. Our method crucially relies on the predictability of γ, so we will assess its performance under different values of it.

The general pattern that emerges is that, in countries where the reported data are accurate, our two-stage procedure is less prone to suggest false corrections. Where the reported data are poor, the results are more ambiguous, and overall, the estimators appear to be more or less on par in these cases. Especially as the bias in mis-reporting becomes more systematic (that is, on expectation more different from 0), its performance deteriorates, while the benchmark estimator assuming a single slope γ remains relatively unaffected by this: In these cases, our estimator has a tendency of being too 'lenient', and increasing parts of the bias from mis-reporting are reflected in the $\tilde{\gamma}_i$s.[10] Appendix A.3 compares the performance of the estimators across the range of parameters specified in table 2.1.

Figure 2.4: Mean absolute prediction error by R^2 of second stage



Notes: The plotted values are obtained from sorting iterations into bins by their R^2 in the second stage (explanatory power of the elastic net), and then taking the mean absolute prediction error within each of these bins.

Irrespective of the underlying parameters, we observe that the quality of our

[10] One promising avenue for future research would be to explore weighting country observations by their statistical capacity or a proxy thereof, that is, giving reliable data more weight when deriving the systematic component of the variation in $\hat{\gamma}$. This would come at the cost of making *a priori* assumptions about the quality of the data, but potentially enhance predictive accuracy.

predictions from the two-step procedure strongly hinges on the predictability of our parameter γ_i, or more precisely of its estimate $\hat{\gamma}_i$: The larger the proportion of the inverse elasticity that we can explain based on Φ^*, the more precise the subsequent predictions of growth rates, conditional on luminosity. Figure 2.4 plots the average prediction error in annual growth rates by the range of R^2 in the second stage, that is, how much of the variation in $\hat{\gamma}_i$ is captured by the elastic net. The values are obtained by grouping iterations of the simulation into bins, according to the R^2 of the elastic net in the second stage. The prediction errors from the two-stage procedure are plotted in blue. For comparison, the average prediction errors from the corresponding iterations obtained from the fixed effects estimator assuming a common slope γ is plotted in red. For bad data countries, the two-stage procedure does not substantially outperform the benchmark; instead the measures are more or less on par. For good data countries, however, it does outperform the fixed effects estimation even where very little (10-20%) of the variation in the $\hat{\gamma}_i$ is captured; fewer false corrections are applied to countries countries with accurate data. As the explanatory power of the second stage goes up, so does the relative advantage of the two-stage procedure.

Overall, this exercise suggests that there is a potential for the two-stage estimator to improve the predictions based on luminosity by capturing some of the heterogeneity, in particular by avoiding false corrections. Note also that, compared to alternative approaches in the literature, we do not need to make any *a priori* assumptions about countries' data quality in order to obtain that pattern (unlike approaches that compute optimal weights depending on the assumed data quality). In the empirical application that follows, we will report both the estimates resulting from our procedure, as well as more conventional estimates based on the assumption of a single slope. Reassuringly, the big picture of the results remains similar irrespective of the estimator employed.

2.5 Data

The data we employ in our empirical application can be divided into two datasets: The primary dataset consists in the series on GDP (Y) and luminosity (L), our main variables of interest. The second dataset contains the potential determinants of this relationship (Φ^*). This encompasses a wide range of geographical and socio-economic variables at the country level.

Main dataset: Lights and GDP

Lights: The series on luminosity we employ is derived from the version 4 of the DMSP-OLS nighttime lights time series provided by the US National Oceanic and Atmospheric Administration (NOAA). The data span from 1992 to 2013, and come in 30 arc second grids, corresponding to a spatial resolution of about 1km^2 at the equator. It covers the surface of the earth between -65 and 75 degrees latitude, which corresponds to almost the entirety of the earth's inhabited land. For each year, each of the ca. 700 million grid cells is attributed a luminosity value labelled Digital Number (DN), ranging from 0 (very dark) to 63 (very bright). These values are the result of pre-processing by NOAA staff, who combine cloud-free imagery collected throughout the year, and remove glare from solar light, moonlit data, as well as features from the aurora (northern lights).

One major issue with the use of luminosity data is the contamination with gas flaring. In areas where residual gas from the production of petroleum-production is burned, large flames almost constantly illuminate the night sky; in the DMSP-OLS data, even moderately sized plants can lead to the top-coding (DN values of 63) of several kilometres square. In line with most of the literature, we therefore remove gas flaring areas as identified by Elvidge et al. (2009). However, these areas tend to be large and sometimes go far beyond the spots where the flares do induce excess luminosity. In some areas (Nigeria in particular), this leads to a loss of substantial parts of the country's surface. In order to mitigate the problem, we use MODIS data on landcover (version 5.1) to identify agglomerations (defined as adjacent built-up areas larger than 10 km^2) in the gas-flaring areas, and include them with a buffer of 50km.

Top-coding is generally an important issue with the DMSP-OLS data: the scale is capped at 63, and a large share of the lit cells take on this value. This can potentially come with a substantial loss of information at the top end of the distribution (cells that are in reality brighter than what is required for a DN value of 63). A number of issues follow from this: First, urban centres often only consist of lit cells. Any increase in luminosity in these places will not be reflected in the cells' DN, and small decreases will go unnoticed where luminosity is high (larger by some margin than the threshold for being top-coded). Second, particularly densely lit countries (e.g., rich and densely populated ones) may be disproportionately affected by top-coding, altering the relationship between GDP and measured luminosity in these places. Our methodology aims at capturing such effects, by deriving country-specific elasticities based on countries' properties such as population density and income per capita. However, this effect is taken to the extreme in areas that are very small, like city-states and small island countries.

For this reason, and in line with most of the literature, we exclude countries with a surface area smaller than 5,000 km^2. This affects 29 countries, all of them small island states and city-states.

Especially where the interest lies in growth rates, we must be concerned with the consistency of luminosity measurements over time. Because the sensitivity of the sensors declines over time, and satellites are replaced, the reported luminosity data is indeed not immediately comparable across years. Broadly speaking, the literature offers two ways of addressing this issue: The most common way, employed for instance by Henderson et al. (2012), is to include year fixed effects in the regressions, absorbing global fluctuations in luminosity. Indeed, most of the time–inconsistencies appear to be global, and mostly consist in shifts in levels. Another option is to calibrate the luminosity data beforehand, especially where the nature of the estimation does not lend itself to the inclusion of year fixed effects. For instance, Tanaka and Keola (2017) propose a methodology for such adjustments using presumably reliable GDP data from OECD countries as a benchmark to calibrate luminosity values. Elvidge et al. (2009) seek to identify spots on earth that arguably emit constant levels of luminosity as a benchmark for calibration (and choose Sicily). In our application, the inclusion of fixed effects appears to be the econometrically more tractable option. Moreover, we can thereby avoid making any strong assumptions about certain places' constant emissions of luminosity, or the accuracy of their GDP figures.

In 12 out of 22 years of the DMSP-OLS data, there do actually exist two series, as satellites were in orbit at the same time and registered luminosity in parallel. In line with Henderson et al. (2012) and most of the other literature, we average across satellites in these years. The luminosity values we employ in our series are expressed in $DN/cell$ and computed for every country and year. To this end, we take the sum of the luminosity values (DN) on the territory of each country, and divide it by the overall number of cells contained in it.[11] Country borders are obtained from the GADM Global Administrative Areas database version 2.8. Where countries' territories have changed over time, we follow the World Development Indicators in how we treat them. Where country GDP series have retrospectively been split (e.g., Yugoslavia), we use these series and compute luminosity using current borders. Where the GDP series is incomplete, we include those years for which it is available. Depending on the exercise, countries with incomplete data are excluded from the analysis (this is the case for the estimation of country-specific slopes using equation 2.1).

[11] This follows a convention in the literature of considering the density rather than the absolute level of luminosity emitted by each country. While this may be somewhat sensible from an intuitive perspective, it has no bearings on our empirical results as the area of a country is fixed over time, and the scaling will be fully absorbed by country fixed effects.

Finally, when discussing luminosity data, it is worth noting that the DMSP-OLS series has effectively been superseded by the VIIRS series. As shown by Dai et al. (2017), the latter is superior in many respects: It has higher (effective) resolution, is more consistent across years, and less plagued by issues of top-coding. However, the purpose of our investigation is to recover historical growth rates, and VIIRS data is available only from 2012 onwards.

Gross Domestic Product: The GDP series are obtained from the World Bank's World Development Indicators 2015 (WDI). As we are interested in real GDP growth rates, we use the series in constant local currency units (labelled *NY.GDP.MKTP.KN* in the WDI), that is, adjusted for inflation but without any transformations aiming at rendering income values internationally comparable. This is to minimise any distortions to growth rates stemming from adjustments for purchasing power or exchange rates (see Deaton and Heston, 2010); furthermore, it is the series with the widest coverage. The lack of international comparability in levels is no concern, as our analysis does not touch on cross-sectional GDP levels, and differences in units will be fully absorbed by country fixed effects.

Both luminosity and GDP values are included as logarithms in order for the coefficients to be interpretable in terms of elasticities and to conform to the methodological framework outlined in section 2.3.

Auxiliary dataset: Potential determinants of γ

Our secondary dataset consists in factors that are potential determinants of the (inverse) elasticity between GDP and lights, labelled Φ^* in the methodological framework. Its purpose is to derive the key determinants of γ, and then to predict expected values of $\tilde{\gamma}$ that are conditional on these factors for each country. As in the present setup, γ varies across countries but is assumed to be constant over time, this dataset is cross-sectional. Time-varying variables will therefore need to enter in some aggregated form. Depending on the variable, we take averages over the period 1992-2013, or initial values. Furthermore, for the lack of theoretical knowledge of the lights-GDP nexus, it is unclear from the outset which transformation of any given variable is the most sensible one, and has the strongest predictive power for $\hat{\gamma}$. We therefore include the variables in up to three transformations: As square roots (where this is numerically possible), in levels, and squared. Similarly, the choice of the variables themselves cannot follow any rigid theoretical understanding of the relationship of interest, and a parsimonious specification cannot be motivated on strong theoretical grounds (see section 2.3). The included variables are therefore partly motivated from earlier studies concerned with the relationship

between GDP and lights (especially Wu et al., 2013; Levin and Zhang, 2017; Keola et al., 2015; Bickenbach et al., 2016), and partly on speculative grounds. We then leave it to the elastic net estimator to select the variables with a strong predictive power for γ across and beyond the sample, and to attribute appropriate weights.

Table 2.3: Potential determinants of γ (Φ^*)

	Transformations			
Variable	Level	Square	Root	Aggregation
Geographic (non-anthropogenic)				
Snow cover	✓	✓		Mean
Absolute latitude (centroid)	✓	✓	✓	First
Surface area	✓	✓	✓	First
Geographic (anthropogenic)				
Population density	✓	✓	✓	Mean
Population growth	✓	✓		Mean
Urban population (share of total)	✓	✓	✓	Mean
Forest area (share of total)	✓	✓	✓	Mean
Agricultural land (share of total)	✓	✓	✓	Mean
Economic				
GDP level	✓	✓	✓	First
GDP per capita	✓	✓	✓	First
Investment (% GDP)	✓	✓	✓	Mean
Consumption (% GDP)	✓	✓	✓	Mean
Agriculture (% GDP)	✓	✓	✓	Mean
Industry (% GDP)	✓	✓	✓	Mean
Services (% GDP)	✓	✓	✓	Mean
Manufacturing (% GDP)	✓	✓	✓	Mean
Agricultural growth (relative)	✓	✓		Mean
Industrial growth (relative)	✓	✓		Mean
Services growth (relative)	✓	✓		Mean
Manufacturing growth (relative)	✓	✓		Mean
Fossil fuel revenues (% GDP)	✓	✓		First

Table 2.3 lists all included factors and the included transformations. Most variables are obtained from WDI, with the following exceptions: Snow cover is obtained from the National Snow and Ice Data Center; it corresponds to yearly average snow cover (as a share of the land surface area), aggregated from monthly data and within the borders provided by the GADM database. The absolute latitude is the latitude of the centroid of a country, computed by the author based on GADM borders. Furthermore, the sectoral growth variables (agricultural, industrial, services and manufacturing growth) are relative contributions – they express the share of total growth that is attributable to the respective macro-sector, that is, they are not growth rates *per se*.

2.6 Heterogeneity in the relationship between GDP and lights

Henderson et al. (2012), as well as the bulk of the studies using a methodology similar to theirs, estimate a model of the type specified in equation 2.5 or 2.6. On this basis, they estimate an inverse elasticity $\hat{\gamma}$ (see section 2.3), finding a value of about 0.28. Throughout their analysis, this coefficient is assumed to be constant across countries, and serves as a basis for assessing the accuracy of reported growth rates in countries with poor statistical capacity.[12]

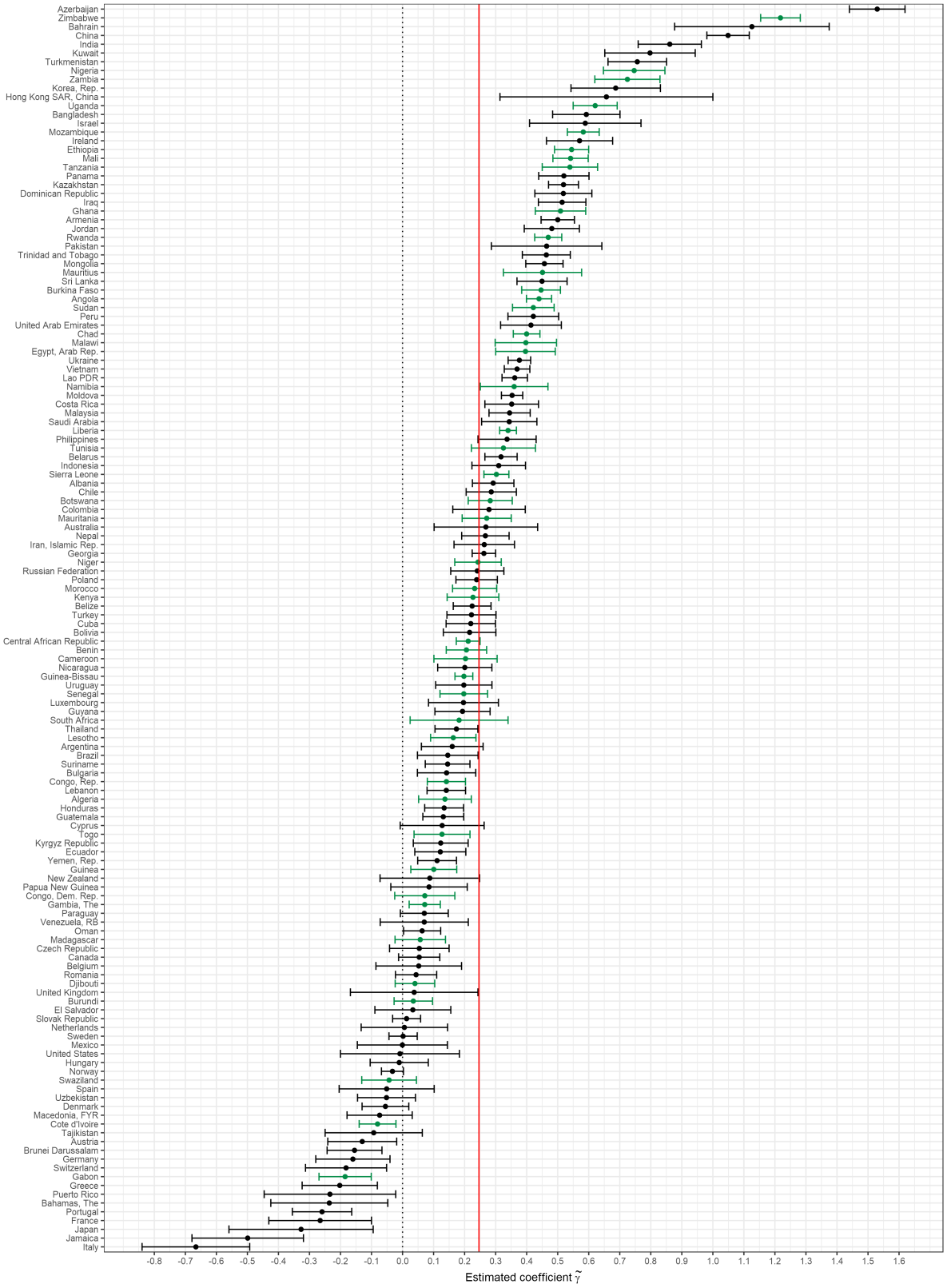
Figure 2.5 plots the coefficients obtained from estimating equation 2.1, that is, a twoway fixed effects regression augmented by N $\ell_{it} \times D_i$ interaction terms. This allows us to obtain country-specific estimates of the slope coefficient γ_i. While each of these estimated slopes is based on a small sample of 22 observations (1992-2013), the strength of the relationship is such that most of the obtained coefficients are statistically significant from zero at the 5% level (114 out of 142).[13] The figure plots the country-specific coefficients in descending order of the point estimate; the precise values are reported in tabular form in appendix A.4. The errorbars around the point estimates indicate 95% confidence intervals based on standard errors clustered at the country level computed following MacKinnon and White (1985). In view of our later application, the bars associated with coefficients for African countries are highlighted in green. On average, the estimated $\hat{\gamma}$ is 0.25, which is marginally lower than the 0.28 obtained by Henderson et al. (2012), and lower than what we obtain when estimating equation 2.5 on the sample at hand (0.30, see section 2.7). But, crucially, few of the coefficients do actually correspond to that mean: in 109 out of 142 countries, the coefficient is statistically significantly different from the mean at the 5% level. At the extremes, Azerbaijan has the highest $\hat{\gamma}_i$ at 1.53, and Italy has the lowest at -0.67.

Naturally, the dispersion of the coefficients is the result of a conflation of factors. First, the OLS coefficient has its own variance by construction, and even under perfect parameter homogeneity one would expect some variation across coefficients obtained from different random samples, even more so in the presence of measurement error and where there is an element of randomness in the relationship between the variables. However, as suggested by the fairly narrow confidence intervals, this can only be a small part of the story. Second, there may be struc-

[12] In order to derive optimal weights and compute lights-adjusted growth measures, Henderson et al. (2012) rely on equation 2.6 and exclude high-income countries specifically because their γ appears to differ.

[13] As the luminosity data only starts in 1992, earlier studies had to rely on sample sizes that were prohibitively short in order to rely on the time dimension to estimate elasticities based merely on within country variation.

Figure 2.5: Estimated $\hat{\gamma}$ in 142 countries



The plotted estimates are obtained from a two-ways fixed effects regression with countryspecific slopes as specified in equations 2.1. Mean point estimate of $\hat{\gamma}$: 0.25 (red line). Mean SE: 0.05. Standard Errors are clustered at the country-level following MacKinnon and White (1985). The errorbars indicate 95% confidence intervals. Bars highlighted in green correspond to African countries.

tural differences between countries in the relationship between luminosity and economic activity that are driven by *unobservable* factors (Ψ); say, a culturally driven preference for bright streets. Third, structural differences across countries that are driven by *observable* factors (Φ) may lead to differences in the elasticities. And finally, systematic mis-reporting or mis-measurement of GDP figures will be reflected in different estimated elasticities across countries. The aim of this study is then to identify the observable factors Φ that drive the relationship between GDP and lights, compute *expected* true coefficients for each country conditional on these observed factors, and derive luminosity-based growth rates based on those conditional, country-specific coefficients.

Visually inspecting figure 2.5 already appears to be suggestive of some patterns. Looking at the very top of the distribution, there appears to be a high number countries where petroleum production accounts for a large share of the economy: with Azerbaijan, Bahrain, Kuwait, Turkmenistan and Nigeria, 5 of the 8 countries with the highest estimated inverse elasticities $\hat{\gamma}$ are major oil producers. Beyond misreporting and measurement error, this is compatible with a number of explanations. For one thing, while we do remove areas that have been identified as contaminated by gas-flaring by Elvidge et al. (2009), it is likely that some areas with gas flares still remain (see section 2.5). Any growth-induced increase in luminosity may then be dwarfed by the brightness emitted by gas flares irrespective of other economic activity, resulting in a low elasticity of lights with respect to GDP, and therefore a high inverse elasticity γ_i. Alternatively, the relative cheapness of electricity may imply that even at lower stages of development, luminosity is already high, and further increases in GDP do not induce much additional luminosity. At the bottom of the distribution of $\hat{\gamma}_i$s, on the other hand, there appears to be a bunching of highly developed economies: 10 out of 34 OECD members are among the 20 countries with the lowest $\hat{\gamma}$. The average reported per capita income in these places in the year 2000 was 24,865$ US, compared to 9,048$ US in the remaining 122 countries (WDI 2015).

Figure 2.6 depicts the results from a grouped regression of the form specified in equation 2.4. The plotted coefficients $\hat{\gamma}_g$ are those on the $\ell_{it} \times D_g$ interactions, and indicate the estimated inverse elasticity as obtained for each of the sub-groups of countries. The compositions of the groups are summarised in table 2.4, with details in appendix A.1. Grouping 1 uses a k-means sorting algorithm to divide the countries into two distinct groups (see appendix A.1), based on the share of their GDP that that is attributed to exports of fossil fuels (oil, coal, or natural gas). The resulting groups are unequal in size, with a group of 25 countries that have an average share of fossil fuels in GDP of 37.6% across the sample period, and a group of 135 countries where fossil fuels make a very small contribution to GDP,

Table 2.4: Composition of the groups (all values are group means)

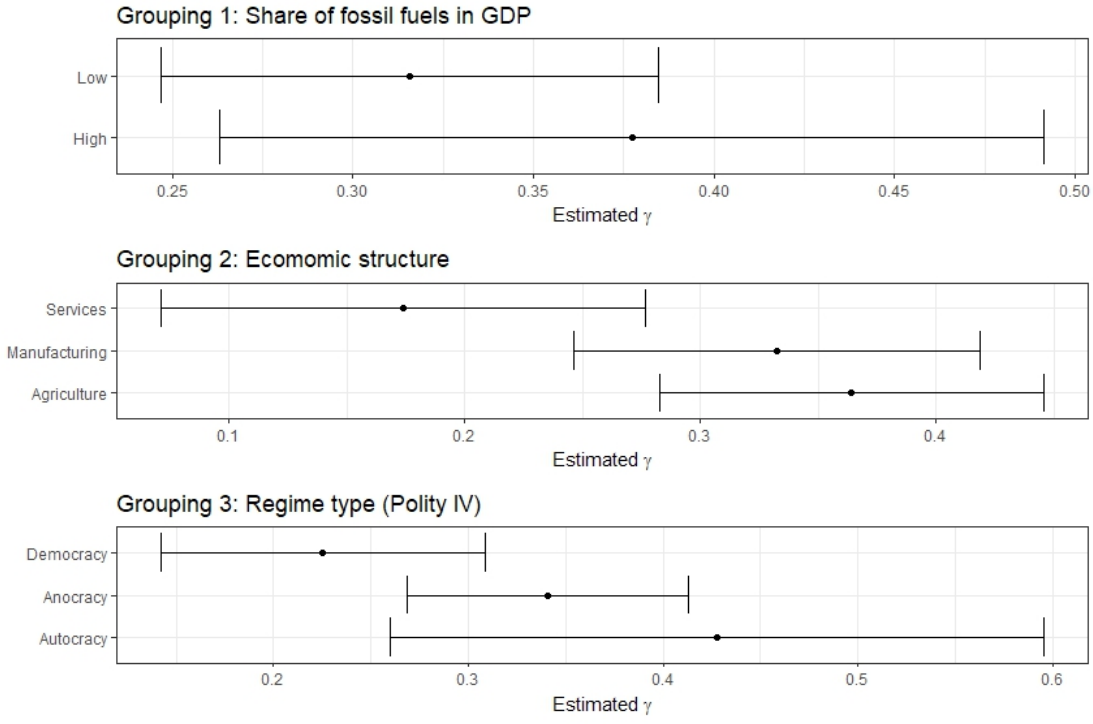
Grouping 1: Share of fossil fuels in GDP		
Group	Fossil fuels (%GDP)	N
1	37.6	25
2	1.6	138

Grouping 2: Economic structure				
Group	Agric. (%GDP)	Manuf. (%GDP)	Services (%GDP)	N
1	26.3	6.8	38.4	33
2	15.6	17.4	49.9	65
3	2.8	13.5	63.6	58

Grouping 3: Regime type (Polity IV)		
Group	PolityIV	N
1	-8.0	16
2	0.3	65
3	8.5	75

Notes: Groupings 1 and 2 are obtained from a k-means clustering algorithm, grouping 3 follows the Polity IV classification whereby any country with a rating below -6 is classified as an autocracy, any above 6 as a democracy, and those in between are labelled anocracies. See appendix A.1 for details.

Figure 2.6: Estimated $\hat{\gamma}$ by groups of countries



Notes: The plotted estimates are obtained from a two-ways fixed effects regression augmented by $\ell_{it} \times D_g$ interaction dummies, as specified in equation 2.4. Standard Errors are clustered at the country-level following MacKinnon and White (1985). The errorbars indicate 95% confidence intervals. Details on the construction of the groups can be found in appendix A.1.

on average 1.6%. As it would be expected from what we observed in figure 2.5, the point estimate of $\hat{\gamma}_g$ is larger for the fossil fuel producing countries (0.38) than for those with little reliance on fossil fuels (0.32). However, the precision of these estimates is too low in order for these differences to reach statistical significance.

Grouping 2 is again based on k-means clustering, where countries are sorted into clusters that are similar in terms of their economic structure, as measured by the relative shares in GDP of the main sectors (agriculture, manufacturing, and services, averaged over the sample period). The resulting groups are (1) a group of 33 countries with a particularly high agricultural share in GDP (26.3% on average, compared to 15.6% and 2.8% in groups 2 and 3), (2) a group of 65 countries with a relatively even sectoral composition, and the largest average share of manufacturing (17.4%), and (3) a group of 58 countries dominated by services (63.6% of GDP, on average); the latter group comprises all 32 OECD members. In line with what is suggested by figure 2.5, group 3 exhibits a much lower $\hat{\gamma}_g$ (0.17) than the two groups of countries that rely more on agriculture (0.36) and manufacturing (0.33).

Grouping 3 divides the countries according to their regime type, as suggested by the Polity IV index (averaged over the sample period; Marshall and Jaggers, 2002, updated version from 2014). Polity IV rates countries' political regimes on a scale that ranges from -10 (fully autocratic) to 10 (fully democratic). In order to form groups, we apply the same threshold values employed by the Polity IV project, whereby any country with an overall rating below -6 counts as an autocracy, any country with a rating above 6 counts as a democracy, and any country with a rating between these values counts as an 'anocracy'. Again, the estimated standard errors are too large in order for statistically significant differences to emerge at the 95% level. But the point estimate for democracies (0.23) is, in economic terms, substantially lower than that for anocracies (0.34) or autocracies (0.43). This highlights another factor that may be contributing to the dispersion of (inverse) elasticities: systematic misreporting. The observed pattern suggests that autocracies emit less luminosity for every increase in reported GDP. This is consistent with a situation where autocracies have a tendency to over-report their growth rates. In fact, Magee and Doces (2015) draw precisely this conclusion based on luminosity data.

Of course, there is some overlap between the groups of countries discerned above and we have thus far not made any attempts at identifying the underlying factors that drive differences in the $\hat{\gamma}_g$: For instance, rich economies tend to rely more on services, and be more democratic (see appendix A.1). The analysis that follows will therefore seek to isolate the relevant drivers of the elasticity between GDP and lights, and to provide estimates of actual growth rates that take these

factors into account. Note that, in order to avoid replication of systematically misreported growth rates, e.g. by undemocratic regimes, only factors that are arguably 'legitimate' drivers of the lights-GDP relationship (that is, not suggestive of misreporting, like regime type) will be considered in the predictive exercise in section 2.8.

2.7 Explaining the relationship between GDP and lights

2.7.1 Estimating the relationship between GDP and lights

In what follows, we will establish the relationship between GDP and lights that our predictions later on will be based on. In line with the discussion above, we do so in three different ways: First, for reference, we estimate the long difference as specified in equation 2.6; as in Henderson et al., we conflate the first and last two observations in the sample period (1992/93 and 2012/13). Second, we estimate the panel FE model as specified in equation 2.5. And third, we estimate country-specific slopes in a fixed effects regression augmented by N $\ell_{it} \times D_i$ interaction terms, as in equation 2.1. Table 2.5 summarises the results obtained from these regressions. Columns 1 to 3 summarise the cross-sectional long difference equation (equation 2.6), columns 4-6 the panel regressions (equation 2.5), and column 7 refers to the panel regression with country-specific slopes. As the latter is only estimated for countries where observations are available for all 22 years, we report the results for the other models using the same sub-sample for reference (columns labelled 'Compl.'). For the models with a single coefficient (all except column 7), we also report the results obtained when excluding OECD countries (columns labelled 'No OECD'). This is to create comparability to Henderson et al. (2012) who exclude rich economies from their benchmark estimations, as they find their $\hat{\gamma}$ to be smaller.

Focussing on the long differences reported in columns 1-3, the sample is naturally limited to those countries where luminosity and GDP values are available in the years 1992/93 and 2012/13. Using the maximally available sample of 145 countries, the estimated coefficient is 0.45. Restricting the sample to the 113 non-OECD countries does not substantially reduce this estimate (0.43). In fact, the impact of restricting the sample by only three countries (Finland, Iceland, Myanmar) in order to match those employed for the estimation of country-specific coefficients has a larger impact, reducing the estimate to 0.42. Either way, the differences between the three estimates are small and within about half a standard

Table 2.5: Regression output for different models

	Long difference			Panel fixed effects			Individual
Sample	Max	No OECD	Compl.	Max	No OECD	Compl.	Compl.
ln(lights)	0.45***	0.43***	0.42***	0.30***	0.31***	0.30***	0.25†
	(0.05)	(0.06)	(0.05)	(0.04)	(0.04)	(0.03)	(0.05)†
R^2	0.36	0.32	0.36	0.20	0.20	0.22	0.53
Adj. R^2	0.36	0.32	0.36	0.15	0.15	0.17	0.48
N	145	113	142	166	132	142	142
T (avg.)				21.16	21	22	22
Obs.	145	113	142	3512	2773	3124	3124

Notes: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. †Average of country-specific coefficients (standard errors). The results for individual countries are depicted in figure 2.5 and listed in appendix A.4. The standard errors of all panel estimations (columns 4-7) are clustered at the country level. R^2 and adjusted R^2 are within-R^2 based on a FE model where the FE have been differenced out, which explains why they are much smaller than the values typically reported in the literature (see main text).

error of either estimate, therefore statistically negligible. Note that the values of the estimated coefficients lie above the 0.32 obtained by Henderson et al. (2012). This is likely to be due to differences in the sensitivity of satellite sensors across time, which in the cross-sectional setup cannot be absorbed through time fixed effects. The magnitude of the coefficient may then be sensitive to the choice of sample period, without however affecting the predictive qualities within the respective period.

Turning to the estimates obtained from panel fixed effects regressions (equation 2.5), reported in columns 4 to 6, these estimates are very much in line with the bulk of the literature, where γ is typically estimated at around 0.3 for the global sample (Henderson et al., 2012; Chen and Nordhaus, 2011; Keola et al., 2015). The reported standard errors are clustered at the country level following MacKinnon and White (1985). Across the sub-samples, there is very little variation in the coefficient, with the exception that it slightly increases to 0.31 as OECD countries are excluded from the estimation (column 5). Indeed, our inspection of figure 2.5 in section 2.6 had suggested a bunching of OECD members among the countries with the lowest $\hat{\gamma}$, and we also found a lower value for economies that rely heavily on services (figure 2.6).

Compared to most of the literature, consider also the substantially smaller R^2 and adjusted R^2. This is due to computational differences: The within-R^2 typically reported is based on the model with time fixed effects included as dummy variables, and accounts also for the predictive power of those terms. We report the R^2 based on the same model but differencing out the time fixed effects; this is computationally identical, but the resulting R^2 does not encompass the fixed

effects' explanatory power, which we deem more informative.[14]

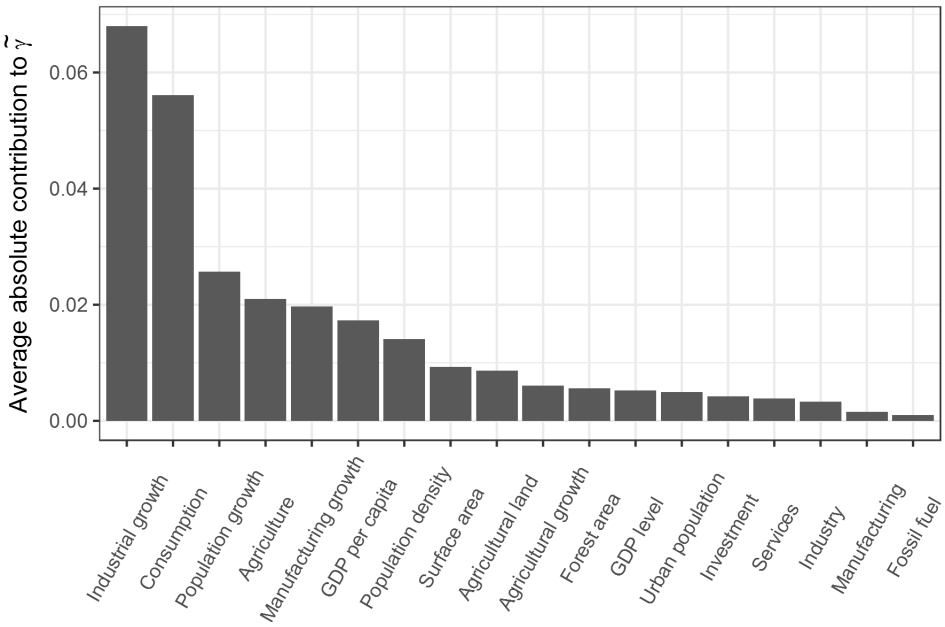
Column 7 reports the aggregate results from the estimation of equation 2.1. The reported coefficient of 0.25 corresponds to the mean point estimate across all countries in the sample. The standard error, 0.05, is the mean estimated standard error for each of these coefficients. The point estimates of the coefficients themselves range between -0.67 (Italy) and 1.52 (Azerbaijan) with a standard deviation of 0.31 between them. Figure 2.5 in section 2.6 plotted the complete distribution of the measured coefficients including their 95% confidence intervals, and appendix A.4 reports them in tabular form. The R^2 is greatly increased when compared to the other models, even when accounting for the inclusion of additional explanatory variables (adjusted R^2), which once more underlines the importance of slope heterogeneity in this context. The estimated average $\hat{\gamma}_i$ of 0.25 is smaller than the one estimated under the assumption of a common γ; the most direct comparison is with column 6, the fixed effects estimate with a single slope (0.30). As discussed in section 2.4, this discrepancy can essentially stem from two sources. First, measurement error biases OLS estimates towards zero, more so in smaller samples. As each of the coefficients we estimate is essentially based on 22 observations, this may induce a slight decrease in the estimated γs. Second, an aggregation bias can occur in a fixed effects estimation where slope heterogeneity is falsely assumed, as we argue is the case (Ul Haque et al., 1999). The direction of this bias depends on the sign of the correlation between the coefficient (γ_i) and the explanatory variable (ℓ), and can therefore not be determined *ex ante*.

2.7.2 Predicting country-specific $\tilde{\gamma}_i$ and growth rates

As noted in sections 2.3 and 2.4, taking the country-specific $\hat{\gamma}$s as obtained from estimating equation 2.1 defies the purpose of our investigation. Instead of providing an indication of potential mis-measurement or mis-reporting of GDP, the resulting predictions would merely replicate the reported long-run growth rates, along with some noise. In order to circumvent this issue, we will now – analogous to section 2.4.3 with the simulated data – seek to identify the systematic component of the variation in $\hat{\gamma}$. To this end, we employ the elastic net estimator as described in section 2.3.2, using the variables described in section 2.5 as potential determinants of $\hat{\gamma}$, Φ^*. Figure 2.7 plots the average absolute contributions of the variables to the predicted coefficients $\tilde{\gamma}_i$, for those variables that have been attributed non-zero weights by the elastic net. Note that, for the ease of representation, the graph aggregates the contributions of the variables in levels and their non-linear trans-

[14] This is related to differences in the software packages employed. Most studies rely on Stata, which reports a within- and a between R^2 that includes the fixed effects' contribution. We use the `plm` package for R for our computations, which differences the fixed effects out.

Figure 2.7: Contribution of determinants to $\tilde{\gamma}$



Notes: The bars depict the absolute contributions of each of the determinants of $\tilde{\gamma}_i$, on average across all countries. Non-linear transformations and the variables in levels are conflated here for readability. 'Industrial growth' etc. correspond to *relative* contributions of the sectors to total GDP growth, that is, they do not correspond to actual growth rates.

formations (squares and square roots). Similar to figure 2.2 in section 2.4.3, the depicted values are average absolute contributions: for each country, we multiply the value of a variable with the respective estimated coefficient ($\hat{\delta}$) – the variable's contribution to $\tilde{\gamma}_i$ – and then average these values across all countries.

When considering the determinants as elicited by the elastic net, it is important to bear in mind the limitations that are inherent to most methods of machine learning. While in predictive terms, these methods can often dramatically outperform traditional statistical methods, great care must be exerted when interpreting their output. As shown by Mullainathan and Spiess (2017), the resulting selection of variables and weights can be quite unstable, even across models that have similar predictive power. For the interested reader, we note however that among the largest six contributors, (i) the share industrial growth, population growth and the share of agricultural in GDP all *increase* $\tilde{\gamma}$, while (ii) the share of consumption and levels of GDP per capita *decrease* $\tilde{\gamma}$. For the share of manufacturing growth in total growth, the contribution is ambiguous, with the variable in levels entering with a negative sign, and the squared variable entering positively. The estimated coefficients are reported in appendix A.7. We are interested primarily in the predictive properties; in this respect, our variables can explain 55% of the variation in $\hat{\gamma}$ (as measured by the R^2), and the out-of-sample mean squared error at the optimal combination of α and λ is 0.07 (see appendix A.6 for the cross-validation

of α and λ).

We can now use the *expected* elasticities $\tilde{\gamma}_i$ conditional on country-characteristics Φ_i^* and luminosity values from each country to predict GDP levels, and then derive growth rates from the relative changes in these levels (see section 2.3.2). The discrepancy between the figures thereby obtained and the officially reported growth rates (according to WDI) is then what we call the 'suggested revision' (as in, suggested by the lights proxy). Figure 2.8 plots the amplitude of the revisions that are suggested by our two-stage method, compared to those obtained from a fixed effects estimation that assumes a single γ. It disaggregates this effect by the statistical quality of the countries as rated in the Penn World Table (see Summers and Heston, 1991), with grades ranging from A (best, e.g., USA) to E (worst, e.g., Chad).[15] Two things can be noted: First, our estimator generally applies smaller corrections to the countries' average annual growth rates than the fixed effects estimates (0.91 pp vs. 1.15 pp across the entire sample). Second, its estimates suggest substantially smaller revisions on average for countries with supposedly better data quality, and larger corrections to places with supposedly less reliable data. This pattern is of course desirable, but almost entirely absent in the alternative estimates based on a single slope: there, the suggested revisions are equally large across levels of statistical quality. We consider this observation indicative of the idea that our estimates are less prone to suggest undue revisions in places where the reported data is in fact accurate, a property that is also suggested by our simulation exercise in section 2.4.

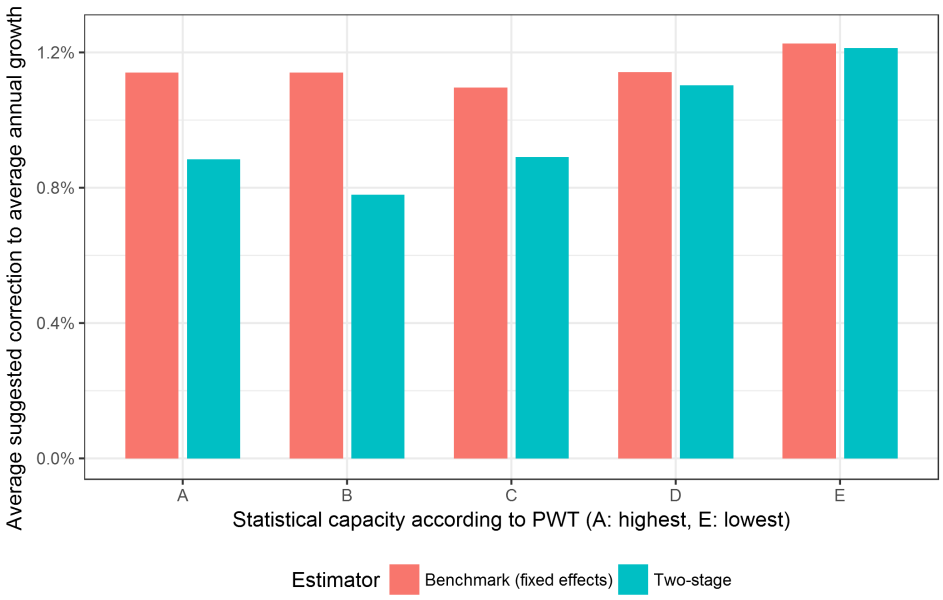
2.8 Application: Economic growth in Africa, 1992-2013

We now turn to the assessment of recent growth performance in sub-Saharan Africa. First, we will briefly review the academic debate, and then present our own results. Note that in our own analysis, we do not engage in any formal hypothesis testing in the statistical sense of the word. Instead, we seek to provide well-grounded indicative evidence of growth patterns in SSA and whether these are reflected in the luminosity data; our focus is therefore on the point estimates.[16]

[15] Recent vintages of the PWT do not report such country grades anymore. An alternative measure of statistical capacity is provided in the WDI, but this measure is not provided for most high-income countries. Furthermore, the PWT estimates date from the beginning of our sample period, and should therefore give a good representation of the initial quality of statistics.

[16] The correct estimation of standard errors for coefficient resulting from estimators with regularisation terms, such as the elastic net we employ, is an open debate. For the elastic net, there exists no analytical solution to derive standard errors. One possible avenue would be to bootstrap the errors, but this does not overcome some fundamental conceptual issues. See, e.g., Dezeure et al. (2015) for a more comprehensive discussion of the issue.

Figure 2.8: Average corrections by statistical capacity



Notes: The bars describe the average correction suggested by lights based on the respective estimator. While FE is indiscriminate as it makes similar corrections across countries of different statistical capacity, the twostage estimator discriminates in the sense that growth rates of poor data countries are corrected more.

2.8.1 An 'African Growth Miracle'?

Perhaps the most optimistic view on recent economic developments in SSA comes from Young (2012), who summarises his findings as the 'African Growth Miracle'. Based on increases in assets owned by households observed in Demographic and Health Surveys (DHS) since 1987 and using education as a proxy for income, he estimates that real household consumption per capita in sub-Saharan Africa has been growing at an annual rate of 3.4% to 3.7% between 1991 and 2004. That is, up to four times the rate reported in national accounts figures. From a methodologically very different perspective, Pinkovskiy and Sala-i Martin (2014a,b) arrive at similarly optimistic conclusions. Combining growth rates from national accounts with distributional data derived from household surveys, and assuming a log-normal distribution of incomes, they find that poverty in Africa had been falling at a much higher rate then previously thought. While optimistic with regards to poverty, unlike Young (2012), they do not challenge aggregate figures on GDP or consumption.[17]

Not all research shares this optimism. Harttgen et al. (2013), for instance, strongly question the empirical foundations of Young's 'African Growth Miracle'.

[17]In fact, Pinkovskiy and Sala-i Martin (2014b) explicitly compare the accuracy of national accounts data versus survey data as a measure of true income using light emissions at night as a third and independent measure of income. They conclude that national accounts data consistently outperform aggregate income estimates derived from household surveys.

They highlight that Young's conclusions rest on a number of assumptions that are unlikely to be met in reality. Crucially, Young assumes constant income elasticities for assets, ignoring changes in preferences or relative prices. Using educational attainment as a proxy for income, his results also rest on the assumption of constant returns to education, ignoring the existence of a substantial body of literature indicating that this is not the case. Beyond these methodological objections, Harttgen et al. (2013) question the validity of asset based measures of consumption in general: They show that, empirically, there is almost no correlation between the two measures. Growth in asset ownership would therefore be more likely to reflect lower disposal rates (new items last longer), or shifts in preferences and relative prices, than unobserved growth in consumption. They conclude that the traditional view, whereby African economies stagnated until 1995 and then started to grow, cannot convincingly be challenged on these grounds. In contrast to what Young finds using changes in asset ownership, those income or consumption measures that are derived immediately from survey means typically indicate lower growth than those reported in national accounts data. Mediating between Young (2012) and Harttgen et al. (2013), Johnston and Abreu (2016) note that asset based indexes are best used as a proxy for changes in wealth, rather than for changes in income. While Young's calculations may not be a fair challenge of national accounts data, they would however reveal substantial increases in household wealth in wide parts of the continent.

Jerven (2014) reviews the evidence on the growth performance across Africa over the past two decades, and reaches the conclusion that recent growth rates are likely to have been over-reported. This is because GDP levels as such are typically *under*estimated, as witnessed by recent revisions of GDP figures in many countries across SSA, e.g., 63% in Ghana or 60% in Nigeria. As statistical capacity is increasing, economic activity is being registered that previously went unnoticed. This 'statistical growth', as Jerven labels it, appears to be particularly prevalent in the years that precede a GDP rebasing. Because rebasing is a process that spans over several years, the direction of future GDP revisions can become clear long before the actual rebasing is declared; it may then be in the best interest of statistical offices and incumbent governments to attribute much of the newly measured economic activity to recent growth. Jerven emphasises that there is likely to be substantial heterogeneity in the misreporting of recent growth figures, and that countries with recent revisions in GDP levels are more likely to have inflated recent growth figures than those where no such revisions took place.

Although they do not explicitly focus on African countries, Henderson et al. (2012)'s examination of growth patterns based on changes in luminosity still contributes to this debate. 24 out of 30 countries for which they present revised growth

rates (the countries with the lowest statistical capacity rating by the World Bank) are African. The suggested revisions are negative for 13 of these countries, and positive for 11; on average, their results suggest a moderate upwards correction of about 0.12 pp annually between 1992 and 2006 for the 24 African countries with the lowest statistical capacity. The aim of the statistical exercise in this paper is to re-assess and refine these estimates, accounting for the fact that the relationship between luminosity and GDP growth may indeed differ across countries. Furthermore, we refine the analysis by considering growth patterns within sub-periods, and across relevant groupings of countries.

2.8.2 The aggregate picture

We now assess the sub-Saharan African growth experience from 1992 to 2013 in the light of our estimates from an aggregate persepective. Besides looking at the continent as a whole, we will disaggregate it into relevant groups in two ways. First, we consider growth experiences of coastal, landlocked and resource-rich countries separately; this typology by 'opportunity group' was introduced by (Collier and O'Connell, 2009), and has arguably dominated and structured much of the debate about African growth in recent years (see Thorbecke, 2015, for a comprehensive overview of typologies). If there was evidence that across these groups, there has been a substantial bias in reported GDP figures, this could be considered an indication that at least parts the recent discourse have been flawed. Second, we sort countries into groups according to whether they recently updated their base years after a prolonged period of time, in order to examine whether there is evidence supporting Jerven (2014)'s hypothesis of 'statistical growth'.

We start by looking at the African continent as a whole, to assess whether there is any support for an 'African Growth Miracle' as declared by Young (2012) in the lights data. Figure 2.9 plots the trajectory of real GDP across three sub-groups of African countries in our final sample, as well as across all these countries as a whole. The reported values are the unweighted averages of real GDP levels, normalised to their 1992 values. The solid blue line is derived directly from the official values, as reported in the WDI 2015. The dotted red line reports the predictions based on luminosity under the assumption of a common elasticity between lights and GDP across all countries (the predictions obtained from equation 2.5, or the model reported in column 6, table 2.5). The dashed green line plots the predictions obtained from our two-step method with heterogeneous slopes as described in section 2.3 and 2.4, and therefore accounts for differences in the relationship between luminosity and GDP across countries.

Panel (a) presents the most aggregate picture of the evolution of real GDP

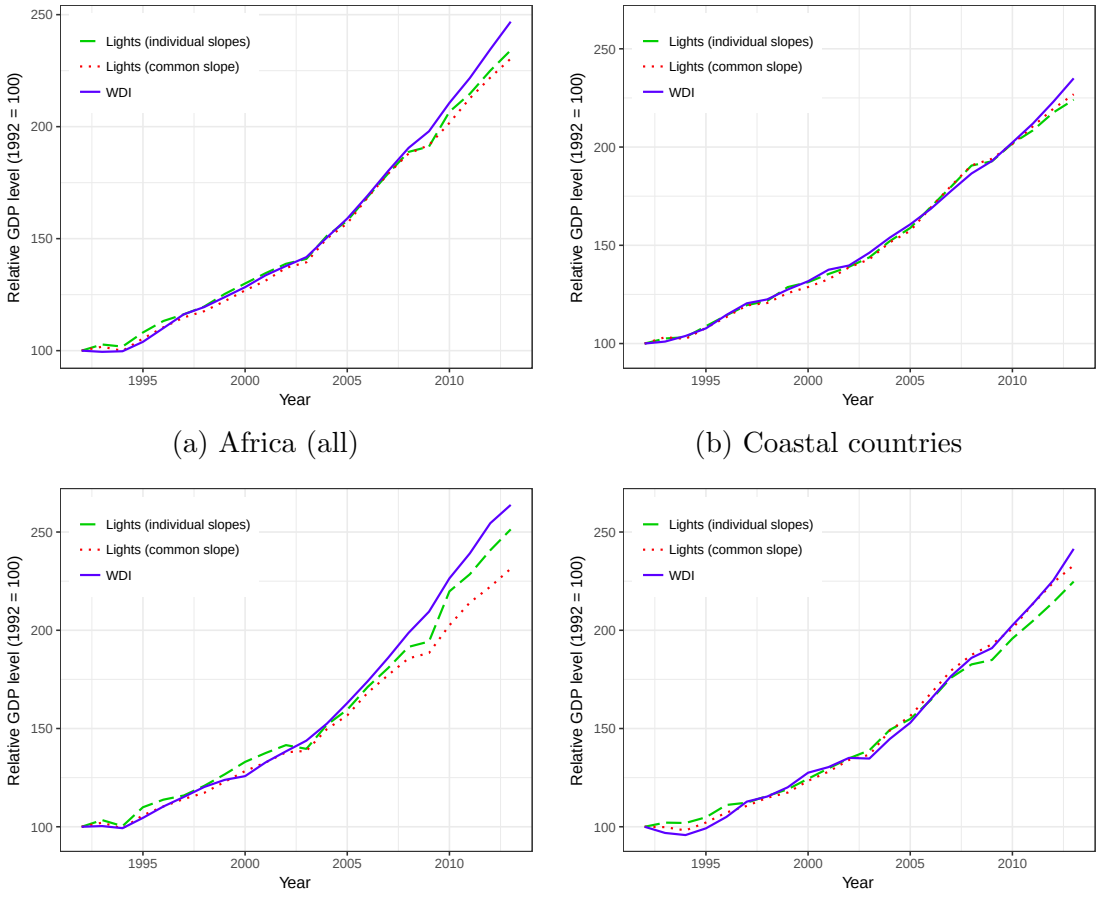
across the African countries in our sample. As suggested by the trajectory of the blue line, the WDI report consistently high growth rates starting from 1994. Across the entire period, the reported growth rates in the sample average at 4.28%. A noticeable acceleration in growth rates occurs around 2003: the average growth rate between 1993 and 2003 is 3.34%, compared to 5.32% between 2004 and 2013. Throughout the two decades we consider here, there appears to be, on average across countries, no substantial discrepancy between officially reported growth rates and those suggested by the luminosity series, whether we allow for heterogeneity in the γs or not. The series follow each other very closely, and only from 2007 onwards there is a slight discrepancy. However, as opposed to the narrative of an 'African Growth Miracle' that would have gone unrecorded in the official national accounts data, this discrepancy is indeed negative, with the luminosity measure suggesting slightly less economic growth than the official data. Note however that the discrepancy is moderate: Over the entire period, the average growth rate suggested by luminosity (with heterogeneous slopes) is 4.14%, 0.14 pp below what the official data suggests. For the period after 2007, these figures are 5.32% and 0.19 pp respectively. Overall, our aggregate results confirm Henderson et al. (2012)'s finding, whereby there does not appear to be a systematic directional bias in the mis-reporting or mis-measurement of GDP figures in SSA over the past decades.

Coastal, landlocked and resource-rich countries: Arguably the most influential economic typology of African countries is that by Collier and O'Connell (2009), who divide them into (resource-scarce) coastal, (resource-scarce) landlocked, and resource-rich countries (Collier and O'Connell, 2009, p. 126–127). For the period they consider, 1960–2000, they find that these are central defining features of the growth performance of sub-Saharan African countries and developing countries in general; while growth in SSA was disappointing during that period across all three categories, it was worst for the category of landlocked countries, a feature that is commonly considered a major impediment to the participation in international trade and economic growth. Implicitly or explicitly, much of the discourse about the economic performance of African countries has evolved along these lines; if growth rates had been systematically mis-reported across these categories over the past two decades, this may therefore likely have flawed considerable parts of the debate.

Panels (b)-(d) disaggregate the African sample into the categories suggested by Collier and O'Connell (2009).[18] All of the series we report agree on the stylised

[18]Note that Collier and O'Connell (2009, p. 77) report per capita figures and weight countries by their population numbers when computing growth rates, as they seek to best describe the

Figure 2.9: Growth performance by opportunity groups (Collier and O'Connell, 2009)



(a) Africa (all)

(b) Coastal countries

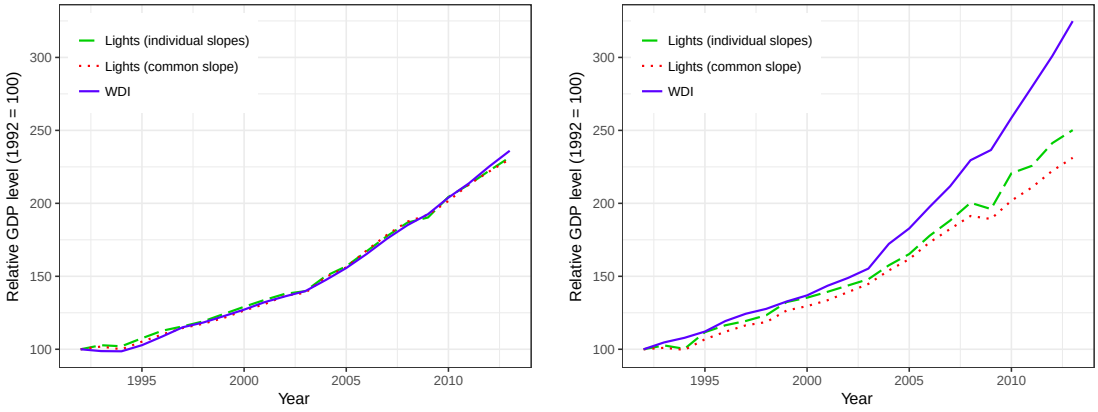
(c) Landlocked countries

(d) Resource-rich countries

fact that countries of all three categories have, on average, experienced substantial growth over the past two decades. The most substantial discrepancy between official growth rates and those suggested by changes in luminosity is for landlocked countries: The WDI suggest some convergence here, with the historically stagnating landlocked countries growing at 4.42% annually, as opposed to 3.89% for coastal and 4.29% for resource-rich countries respectively. Under the assumption of a common slope across all countries, the luminosity data is somewhat more pessimistic in that respect. For the later part of the sample period (2007 onwards), it indicates an average growth rate of 4.11% as opposed to the 4.42% suggested by WDI for landlocked economies. However, once we account for the characteristics that determine the relationship between GDP and lights, most of this discrepancy is bridged; in fact, the arithmetic average of annual growth rates suggested by the two-stage estimator is slightly higher than what WDI suggests, at 4.50%.

'experience of the typical African'. We discuss total GDP and focus on the accuracy of the reported figures, and countries are therefore given equal weight.

Figure 2.10: Growth performance of countries with recent GDP revisions



(a) No major rebase since 2000

(b) Major rebase since 2000

Countries that recently rebased GDP: Next, we seek to examine the hypothesis that those African economies that have recently performed major rebasings of their GDPs tended to report inflated recent growth rates (Jerven, 2014). As discussed in section 2.8.1, this could be the case for a number of resasons: First, the process of rebasing GDP estimates takes several years, during which the statistical capacity typically increases. Some previously unrecorded activity may therefore be recorded, showing up as economic growth in official figures ('statistical growth'). Second, a rebasing will increase the accuracy of current GDP levels, but it does not as such give any indication as to when the discrepancy between old and new estimates came about - that is, when economic growth occured (or did not occur, in the rare case of downward revisions). It can then simply be easier to attribute it to recent years. Third, it is politically advantageous for current leaders to be associated with good economic performance, and governments may opportunistically push for growth to be statistically attributed to the present.

Figure 2.10 is identical to figure 2.9 in its interpretation, but splits the African sample into those countries that recently had a major revision to their base year (panel (b)), and those who didn't (panel (a)). We define those as the countries where (i) GDP has been rebased after the year 2000, and (ii) the previous base year was at least 10 years before the new base year (the IMF recommends an update to the base year every 5 years). This is the case for 5 countries, namely Botswana, Ethiopia, Ghana, Niger and Nigeria. Indeed, the discrepancy is large, in particular after 2003: While the average growth rate according to WDI in these countries in this period is 7.49% (compared to 4.92% in the countries that didn't revise), their performance hardly exceeds the continent's average when proxied by lights. This is true both under the assumption of slope heterogeneity, suggesting 5.44% on average between 2003 and 2013 and when imposing a single slope for

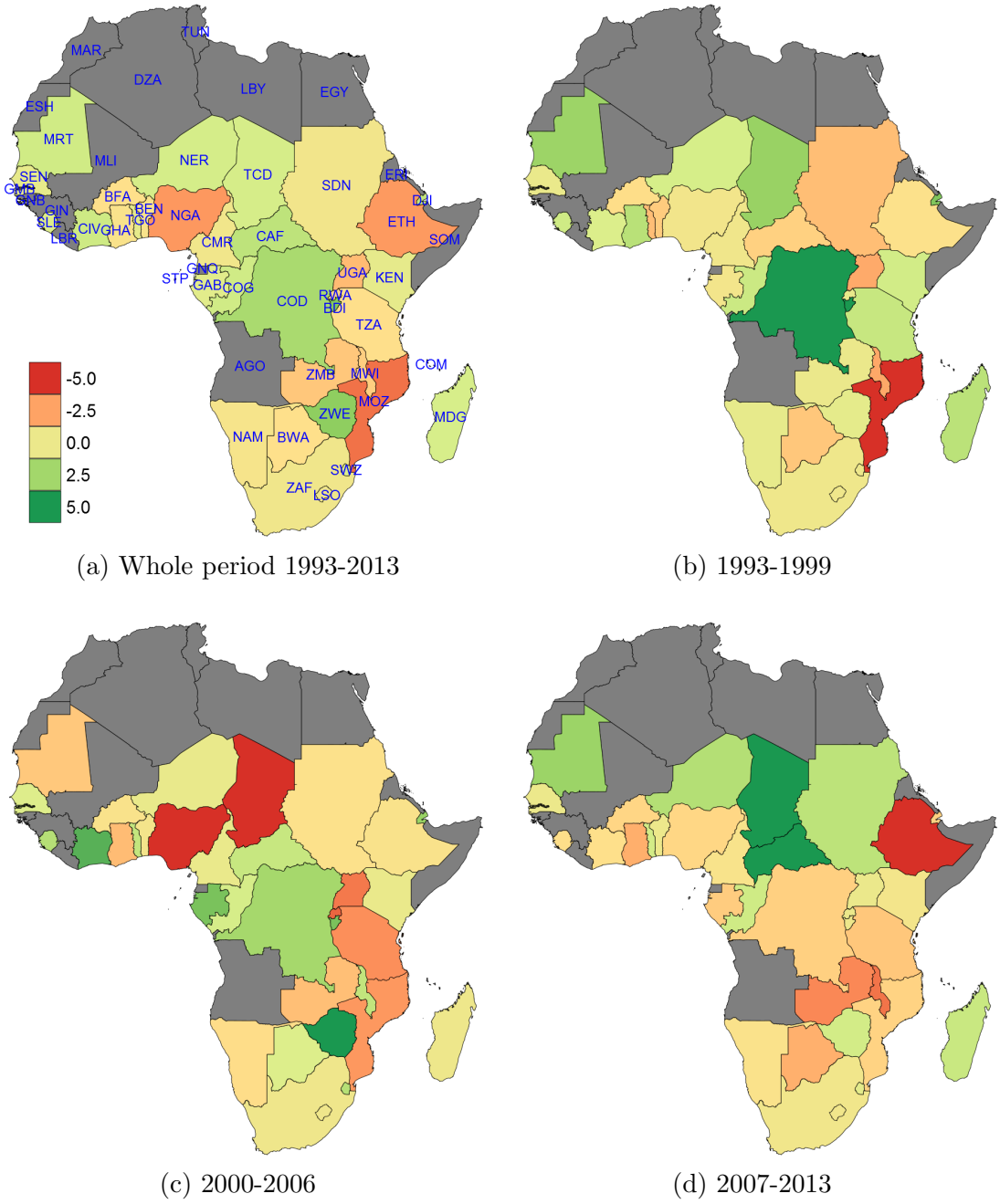
all countries (4.83%). The exception among these countries is Niger, where the two-stage estimates suggest a growth rate 1.10 pp higher than WDI (0.06 pp under the assumption of a common slope).

2.8.3 Country-wise estimates

While a comprehensive discussion of individual countries' growth records and their conformance with the luminosity proxy is beyond the scope of this paper, we will discuss some general patterns and particularly noteworthy cases. The maps in figure 2.11 provide an overview over the discrepancies between the growth rates reported in WDI and those predicted with the luminosity data, allowing for heterogeneity in γ. Equivalent figures under the assumption of a homogeneous γ as well as all underlying values in tabular form are provided in appendix A.8. Negative deviations (red) indicate that the lights measure suggests lower growth rates than WDI, positive deviations (green) mean that lights suggest higher growth rates than WDI. Note that the colour coding is capped at 5 pp and -5 pp deviation in average annual growth rates, which is exceeded in a few cases (see notes of the figure). The figure is divided into four panels: panel (a) depicts the average growth rates over the entire period, 1993-2013 (the first period now being 1993, as no growth rate can be computed for 1992 which is where our series in levels starts). Panels (b)-(d) divide the period into three sub-periods of 7 years each.

As a general pattern, the luminosity proxy tends to relativise some of the more extreme growth experiences, both high and low. For instance, a number of stellar growth records do not find their equivalent in terms of luminosity emissions. This is most pronounced for Mozambique, where growth rates average at 8.5% in WDI. Our luminosity estimates suggest a figure that is 3.6 pp lower – a substantial discrepancy, although 4.9% growth annually is still far above the average growth rate on the continent. Note also that the discrepancy is to the largest part concentrated on the first sub-period between 1993 and 1999. While this coincides with the country's transition to democracy and a substantial wave of repatriation (Sheldon and Penvenne, 2018), reported average growth rates of 10.6% (26.8% in 1996) are not corroborated by light emissions (suggesting instead only 3.8% of growth annually on average over that period). A similarly large discrepancy emerges in the case of Nigeria, one of the countries that recently revised its GDP estimates in a major way: WDI suggest an average growth rate of 6.0% per annum over the entire period, luminosity only 3.2%. Again, the disagreement between the luminosity series and official figures are highly concentrated in time. Between 2000 and 2006, WDI report an average annual growth rate of 9.9%, with a maximum of 33.7% in 2004. This spike is not reflected in the luminosity series, which

Figure 2.11: Growth according to official vs. lights data (individual slopes)



(a) Whole period 1993-2013

(b) 1993-1999

(c) 2000-2006

(d) 2007-2013

Notes: Colours indicate the discrepancy between average WDI growth rates and those obtained from the luminosity proxy, with positive numbers indicating that the luminosity proxy indicates higher growth. The colour coding is capped at -5 pp and 5 pp, meaning that individual discrepancies can be larger. This is the case for MOZ, COD and BDI in panel (b), NGA, TCD and ZWE in panel (c), and ETH, TCD and CAF in panel (d) (see appendix A.8). Borders are obtained from maplibrary.org for illustrative purposes only, and the authors do not imply the expression of any opinion concerning the legal status of any country, area or territory.

suggests a much more moderate 3.3%. In Ethiopia, the discrepancy between the luminosity based growth rates and the reported ones is concentrated in the latest period between 2007 and 2013. Where WDI report an average growth rate of 11.6%, our luminosity measure suggests 4.11%.

At the other end of the spectrum, a few notoriously under-performing countries have increased their levels of luminosity far beyond what their official growth record suggests. Interestingly, among the countries with the greatest positive divergence between the luminosity proxy and WDI growth rates, there appears to be a bunching of particularly conflict torn countries: The greatest positive discrepancies occur in Burundi (3.3 pp), Zimbabwe (3.0 pp), and the Democratic Republic of Congo (2.31 pp). In Burundi, the episode where the WDI and luminosity disagree most coincides with a prolonged period of extreme ethnic violence and civil war. While between 1993 and 1999, the WDI register a decline of, on average, 3.4% of GDP, luminosity increases to an extent that would suggest economic growth of about 3.1% annually – a difference of 6.5 pp. In the Democratic Republic of Congo, the worst official growth record is also in the first period from 1993 to 1999, a period encompassing the First Congo War (1996–1997). WDI report an average decline of 4.17% for this period (–13.5% in 1993 alone). The evolution of the luminosity data suggests a slow increase in economic activity of 1.72% per year. For Zimbabwe, WDI report an average decline of about 6.1% between 2000 and 2006, again a period marked by severe political tensions, including wide-spread government violence. Luminosity values, however, suggest that the economy would still have grown at 1.5% annually.

Evidently, many regularities collapse in periods of armed conflict, and the relationship between luminosity and GDP may well be one of them. For instance, the destruction of a power plant can abruptly stop entire cities from emitting artificial light, while much of the regular economic activity still takes place. On the other hand, warfare in itself is an activity that may generate substantial amounts of luminosity; at the extreme, a process of destruction may then be misinterpreted as constructive economic activity. Once more, we highlight that we consider our estimations as merely indicative of potential mis-measurement or mis-reporting. While our methodology aims at incorporating some of the factors that alter the relationship between lights and GDP, there are natural limits to this ambition. Especially where the interest lies in individual countries or episodes, interested scholars should carefully weigh the full body of evidence. Historical and institutional knowledge, as well as official statistics, should in our opinion form the main pillars of this evidence, and the luminosity can be a valuable complement where it is weak otherwise.

2.9 Conclusion

The close relationship between economic activity and light emissions can offer a unique perspective on historical growth records, especially where the data is weak. In this study, we raised the issue of heterogeneity, in the sense that economic activity may translate into luminosity at different rates in different countries. We offer a method that allows for the relationship to differ, to the extent where this can be explained by observable country characteristics. We then apply this methodology to economic growth in Africa over the past two decades.

At the core, our methodology seeks to disaggregate the elasticity between lights and GDP, allowing for it to differ across countries. To this end, we split the estimation of the relevant coefficients into two parts: First, we run a conventional fixed effects regression, but including a full set of $country \times lights$ interaction terms in order to allow for the slope to differ across countries. The resulting estimates reveal a strong variation in the relationship across countries, but the results have little value where the goal is to discern potential mis-measurement or mis-reporting: Any predictions from this model would merely tend to replicate each country's reported GDP series. This is why, in a second step, we apply the elastic net estimator to discern the part of the variation in the elasticity-coefficients that can be attributed to observable country characteristics. On this basis, we can then infer *expected* elasticities, conditional on countries' economic and geographic characteristics. The resulting coefficients are then used to derive GDP growth rates based on luminosity emissions.

We explore some basic properties of this approach in a simple simulation exercise. The results suggest that, compared to methods that assume a single coefficient across all countries, our method can offer substantial improvements in the accuracy of the inferred growth rates. This is mainly due to the the fact that our estimator suggests less false revisions where the data is indeed accurate. Our simulations show that the relative success of the method strongly hinges on the predictability of the elasticity coefficients based on observable characteristics in a way that is generalisable beyond the narrow sample. With about 55% of the variation in the parameter of interest ($\hat{\gamma}$) explained in our empirical exercise, we believe that our estimates constitute an improvement over those obtained from more conventional methods that assume a common coefficient for the relationship between GDP and lights across all countries. Indeed, we find that the revisions suggested by our estimator are substantially smaller for countries with good statistical capacity than for those with poor statistical capacity, a desirable pattern that does not occur when falsely assuming a common slope.

We apply our estimates to the debate about the recent growth performance of

African countries. Our results do not lend support to claims of an 'African Growth Miracle' (Young, 2012), according to which African countries would have grown at a rate several times higher than those recorded in the national accounts data. Instead, our estimates suggest that, overall, the luminosity emissions of African countries appear to be in line with what official estimates suggest. When looking at countries that recently rebased their GDPs after a long period of time (Botswana, Ethiopia, Ghana, Niger and Nigeria), we find that these countries' reported growth rates tend to exceed the growth rates suggested by the luminosity proxy quite substantially. This is in line with Jerven (2014)'s hypothesis of 'statistical growth', according to which already existing economic activity is discovered with increased statistical efforts, and in parts spuriously attributed to recent years.

When looking at individual countries' growth performances, the main pattern is that the most extreme growth records according to the national accounts data – both negative and positive – are typically relativised by the luminosity proxy. For instance, the stellar growth performances of Mozambique and Ethiopia over the past two decades do not find confirmation in the luminosity data. While the proxy still suggests that these countries grew substantially above average, it does not corroborate official series that in some years suggest GDP growth rates of more than 25%. On the other hand, some extremely negative growth records are corrected upwards, and the most substantial discrepancies on this end of the spectrum are for Burundi, Zimbabwe and the Central African Republic. Interestingly, the largest upwards revisions appear to be concentrated around periods of armed conflict.

This latter observations illustrates our case that growth estimates based on luminosity must be generally considered with care, and evaluated within the specific context: For instance, in periods of armed conflict, the mechanisms that are normally at work in a country cease to function. This is true for the collection of economic data by national statistical offices, but likely also for the interaction between GDP and lights. A damaged power plant may cause the remaining economic activity of an entire city to go unregistered by the proxy, while luminosity caused by purely destructive warfare activities may falsely be picked up as economic growth.

The study sought to incorporate in its methodological framework some of the country-specific context, by allowing the elasticity between GDP and lights to vary with a country's observable characteristics. This does not, of course, cover every aspect of how luminosity is linked to GDP. This relationship may vary over time, and be undermined by certain events, observable and unobservable ones. When assessing any country's historical growth performance, researchers have a plethora of evidence to consider. Besides historical and institutional knowledge, national

accounts data, and individual and household level surveys, night time luminosity should be treated as a valuable complement.

Chapter 3

Foreign Aid, Poor Data, and the Fragility of Macroeconomic Inference using Time Series Analysis

3.1 Introduction

In recent years, there have been growing concerns among scholars and policymakers about the reliability of macroeconomic data, especially for sub-Saharan African countries (see Jerven, 2013c). Partly as a consequence of the low quality of the National Accounts data reported by national statistical offices, there can be large differences between the numbers reported by different datasets such as the Penn World Table (PWT) or the World Development Indicators (WDI): First, data providers may base their series on different vintages of sometimes substantially revised National Accounts data, and second, they often account for shortcomings and gaps in the original data using different methodologies. A growing body of literature is concerned with the impact of such differences on macroeconomic inference (e.g., Ciccone and Jarociński, 2010; Ponomareva and Katayama, 2010; Johnson et al., 2013; Breton, 2015; Breton and García, 2016). Two core findings of this literature motivate the study at hand: First, the GDP series tend to diverge significantly more for countries with low incomes. Second, the results of studies that exploit annual variation in the data seem to be more sensitive to such discrepancies (Johnson et al., 2013). While this finding refers to a limited set of cross-country studies (Ramey and Ramey, 1995; Jones and Olken, 2005; Hausmann et al., 2005), it suggests that time series analyses, which rely entirely on the temporal variation in the data, may be particularly affected by this fragility.

On the other hand, they rely entirely on the variation *within* countries, and therefore remain largely unaffected by measurement issues related to purchasing power *across* countries. This may in turn increase their robustness compared to cross-country studies. To my knowledge, there has been no systematic investigation of this issue in the previous literature. The aim of this study is to fill this gap for the aid-effectiveness literature, which increasingly relies on country-specific time series analysis (e.g., Juselius et al., 2014; Gebregziabher, 2014; Juselius et al., 2017; Addison and Baliamoune-Lutz, 2017).

Before assessing the impact on the inference, this study examines the nature and extent of the discrepancies between the employed series, namely the Penn World Table versions 6.3, 7.1 and 8.0[1], and the World Development Indicators (PWT6, PWT7, PWT8 and WDI in what follows). Special attention is given to the growth rates of GDP and its components, which, from a theoretical perspective, are likely to matter most in the context of a time series analysis of economic growth. In the 36 sub-Saharan African countries included in the analysis, there is indeed striking disagreement between these sources. For instance, none of the correlations between the investment growth series in WDI and those in different vintages of the PWT are higher than 0.07. Even between different releases of the PWT, the series tend to change substantially: for instance, the GDP growth rates reported in PWT6 and PWT7 disagree by 3.3 percentage points on average in every year, which is almost as large as the average reported growth rate (3.7% in PWT6). A decomposition of the divergence between these two series reveals that the largest part of it is explained by changes in the underlying price estimates of the GDP expenditure shares.

For their far-reaching policy implications and dependence on often unreliable data, the results of the literature on foreign aid effectiveness arguably deserve special scrutiny. The conclusions of this literature – in the past mainly obtained from cross-country studies – have previously proven particularly susceptible to even modest changes in the dataset (Easterly et al., 2004; Roodman, 2007) or seemingly subtle nuances in the construction of variables (Van de Sijpe, 2013). In recent years however, an increasing number of studies made use of multiple equation time series models in order to identify the dynamics of foreign aid within countries over time (e.g., Osei et al., 2005; Gebregziabher, 2014; Juselius et al., 2014, 2017; Addison and Baliamoune-Lutz, 2017; Bwire et al., 2017; Mascagni and Timmis, 2017). There are substantial merits to this approach: These mostly country-specific studies do not rely on the restrictive assumption of parameter homogeneity (see Eberhardt and Teal, 2011), allowing for each country to have their own dynamics and long-run equilibrium. Furthermore, they can exploit the

[1]The most recent version of the PWT at the time the analysis was carried out.

full potential of the time series properties of the data, such as identifying long-run and short-run dynamics, while being robust to their pitfalls, like spurious regression due to non-stationary variables.

At the same time, however, these studies put particularly high requirements on the quality of the data, while the level of technicality and the intricate process of model specification can act as a barrier when it comes to performing robustness checks. It is therefore understandably not common practice in this branch of the literature to assess the robustness of one's conclusions to alternative datasets; the aim of this study is to assess the potential importance of this omission. As opposed to the cross-country literature, where robustness checks on the main results of a study often only require the implementation of a single regression model for the entire sample, time series methods generally involve careful modelling of each country under investigation. Indeed, most studies limit their attention to a single country. This substantially complicates any systematic investigation of the robustness of the results in this literature. In order to do so nonetheless, I will take as a point of departure the framework provided by Juselius, Møller and Tarp's (2014, henceforth JMT) recent study on the long-run effects of foreign aid on economic growth. It provides well-specified country-specific Cointegrated VAR (CVAR) models for as many as 36 sub-Saharan African countries, all derived within the same, clearly defined econometric framework. I will then take their models to the data from PWT6 (thus replicating JMT's study with the same data they used), PWT7, PWT8 and WDI. As it is at the heart of the philosophy of the CVAR framework (and the time series literature more generally) to '[allow] the data to speak freely' (Hoover et al., 2008, title), I will also explore the impact of changes in the data on the modelling process itself, and how this is reflected in the results.

The empirical analysis then comprises two parts: First, I apply the 36 country-specific models exactly as specified by JMT to the new datasets. If the data were consistent across datasets, one would expect the models fitted around the original data to also be accurate models of the new data, and the results to be the same or at least similar. If the results change however, this can indicate that either the new data tell a different story altogether, or that the model that accurately captures the old data does not apply to the new data (it is misspecified). In the latter case, it may still be the case that the data effectively tell the same story, but that modifications in the statistical model are needed in order to accurately describe the new data. To address this, in the second part, I re-specify the models for each dataset individually using the statistical criteria employed by JMT for the 12 countries where data are available from all datasets considered.

In the first exercise, using JMT's models, for approximately one third of the

countries in the sample the results change qualitatively, whereas two thirds remain stable. The second exercise, re-specifying the models, induces somewhat greater divergence in the results, but significantly less so for countries where the results remained stable in the first exercise. Interestingly, the stability of the results for any given country does not appear to be systematically linked to the divergence of the time series across datasets: The datasets disagree just as much over the levels and growth rates of GDP and the relative movement of its components in countries where the results remained stable throughout the robustness checks, as in those where the results turned out to be particularly unstable. Overall, my results confirm that the differences between standard datasets can often have very substantial effects on the conclusions drawn from time series analysis, and I suggest that robustness checks with respect to the data should be standard practice in this literature.

The remainder of this chapter is organised as follows: Section 3.2 provides a brief overview of the literature. Section 3.3 introduces the datasets and discusses the extent, nature and origins of their divergence. Special attention will be given to aspects that are particularly relevant in the context of time series analysis. Section 3.4 summarises the methodology employed by JMT, and in section 3.5 I specify my replication procedures and present the results of both exercises. Section 3.6 concludes.

3.2 Related literature

In his 2013 book, 'Poor Numbers', Morten Jerven laments the low quality of National Accounts statistics from developing countries, and warns that this may have serious consequences for researchers' conclusions and policy-makers' decisions (Jerven, 2013c). He argues that, due to a lack of statistical capacity, political will, and changing academic currents, GDP measures quantifying the economic performance of African economies tend to be severely flawed, and often incomplete (see also Jerven, 2011, 2013b,a, 2016). As data providers try and fill in gaps in the data, to restore consistency within series and to establish comparability across countries, large differences between datasets can accrue. A growing literature is concerned with quantifying and explaining these discrepancies, and how they affect the macroeconomic inference that is based on these datasets.

Ram and Ural (2013) show that, between the Penn World Table (version 7.1) and the World Development Indicators 2012, the estimated GDP per capita in 2005 diverges by more than 25% in as many as 33 countries. The differences range from -54% to +66% of GDP in WDI as compared to PWT. The largest relative differences (relative to GDP level) almost exclusively occur in low-income

countries, and no developed country exhibits differences of more than 25% of GDP. These differences do not appear to be following an obvious pattern across datasets, in the sense that neither WDI nor PWT systematically reports higher or lower incomes.

Perhaps more striking than the divergence *between* datasets is the divergence *within* different vintages of the same series, a finding that has often been confirmed for the Penn World Tables. Analysing the differences between four different versions of PWT, ranging from version 5.0 (1991) to 6.1 (2002), Ponomareva and Katayama (2010) find substantial divergence in the GDP growth rates. Using each of these datasets, they replicate the influential contribution by Ramey and Ramey (1995) on the link between business cycle volatility and economic growth. The results vary strongly from one dataset to another and the main result – that countries with higher volatility have lower growth – collapses using some versions of the PWT. In a similar exercise, Atherton et al. (2011) replicate Hanushek and Kimko (2000)'s analysis of the link between labour force quality and economic growth using PWT 6.1 and 6.2, also inducing significant differences in the results. Equally using PWT 6.1 and 6.2, Ciccone and Jarociński (2010) assess the robustness of Sala-i-Martin et al. (2004)'s 'agnostic' approach to growth empirics to differences in the data. Again, the differences between the PWT vintages have a strong impact on the results.

In a similar but more exhaustive exercise, Johnson et al. (2013) compare the growth rates reported in versions 6.1 and 6.2 of the PWT, and replicate 13 major empirical contributions to the growth literature using these datasets. Their analysis shows that the datasets tend to disagree more for countries with smaller size (as measured by GDP) and older benchmark years. In line with what is suggested by the studies discussed above, Johnson et al. (2013) find that some standard results in the growth literature are not robust across different versions of the PWT. But their results also suggest that this fragility is systematically linked to methodological properties of the studies. Crucially, studies relying on relatively long averages (over periods longer than 5 years) are comparatively stable to changes in the underlying data, while analyses exploiting annual variation in the data yield the least robust results. This situation is exacerbated when the analysis refers to poorer (non-OECD) countries. This is of particular relevance for the study at hand, as JMT's analysis, and therefore I, exploit the dynamics of annual data, and focus on some of the poorest countries in the world. Looking at more recent releases of the PWT, the inconsistency does not appear to have decreased, and Breton (2012) finds even larger differences than those observed by Johnson et al. (2013) between versions 6.3 (2009) and 7.0 (2011).

Any comparison between datasets is limited by the fact that there is no straight-

forward way of assessing the accuracy of one dataset against another. One way of addressing this question is to examine the assumptions and methodologies underlying the datasets, and to assess their validity. For instance, Breton (2015) and Breton and García (2016), show that the main difference between ICP 1996 and ICP 2005, reflected in the Penn World Tables' PPP estimates, are the estimated prices of investment in developing countries. They show that this is mainly driven by changes in the methodology used to estimate prices in the construction sector: ICP 2005 bases its estimates on the prices of inputs, while earlier versions of the ICP estimated the prices of construction projects as such – a more demanding, but arguably more precise procedure. Breton and García (2016) come to conclude that earlier versions of the PWT (versions before PWT7) were superior in this respect. The most recent round of the ICP, based on prices collected in 2011, again differs substantially from the 2005 vintage. Both Deaton and Tten (2017) and Inklaar and Prasada Rao (2017) show that this discrepancy is likely explained by methodological shortcomings of ICP 2005. These led to overstated prices in lower income countries, and consequently overstated differences in incomes across countries.

Pinkovskiy and Sala-i-Martin (2016) take a more data-driven approach at assessing the relative accuracy of datasets. Building on a recent literature that uses human nighttime light emissions as a proxy for economic growth (most notably, Henderson et al., 2012) they provide estimates of the relative accuracy of different vintages of PWT as well as WDI. In contrast to Breton and García (2016), their findings suggest that the 2005 ICP round provides an improvement over earlier rounds of the survey. Nevertheless, their results corroborate the frequent finding that 'newer need not be better', in the sense that PWT 7.1 consistently outperforms the more recent PWT 8.0 and 8.1 series. It further emerges from their analysis that PPP adjustments in general tend to decrease accuracy in terms of growth rates, and they recommend using WDI if the focus is on growth (a point also made by Deaton and Heston, 2010).

In summary, it is a well-established finding that the most commonly used datasets tend to disagree to a substantial degree, and this is more pronounced in poorer countries. There is no conclusive evidence as to which dataset is the most reliable, and a large number of studies have been shown to be fragile across datasets. The time series literature has so far been exempt from such an investigation, a gap that this study aims to fill.

3.3 Data

The datasets compared in this study are three versions of the Penn World Table (6.3, 7.1 and 8.0), as well as the World Bank's World Development Indicators 2015. From each, I use data on GDP and its components, namely Investment, Household Consumption and Government Consumption. All aid data are ODA net disbursements as reported by the Development Assistance Committee (DAC) and published by the OECD, which is the same source and measure JMT use.[2] This section will briefly discuss the key methodological differences and similarities between the GDP datasets, and then explore how this is reflected in the data.

Before discussing the individual GDP datasets, a general point concerning the choice of series needs to be clarified. As mentioned in the previous section, there are arguments against using PPP adjusted series where the main interest is in temporal variation (growth rates). Nevertheless, for various reasons (both conceptual and related to data availability), much of the empirical growth literature does rely on PPP adjusted datasets, most notably the PWT. Our baseline study, JMT, is no exception, and we choose our alternative series such that they are conceptually close to the series they originally employed and have sufficient coverage. Three of the four series we employ (PWT6, PWT7, PWT8) are then PPP adjusted, and one is not (WDI; see next section for discussion). Section 3.3.3 will provide some insights as to the influence of price estimates on growth rates, which is the main concern with PPP adjusted series in the context of growth studies.

3.3.1 Conceptual comparison of the datasets

Penn World Table 6.3 (Heston et al., 2009) is the dataset underlying JMT's original study, used here as the benchmark dataset for the purpose of replication. The series JMT employ is real[3] GDP at constant 2005 prices, computed using a Laspeyres index (labelled RGDPL). The prices underlying the PPP adjustments in PWT6 are based on ICP 1996 estimates.

Penn World Table 7.1 (Heston et al., 2012) provides the same variable, RGDPL, but conceptually differs from PWT6 in two ways: First, PWT7 exclusively relies on the prices from the newer ICP round 2005. As noted by Breton (2012), it discards all older price data, inducing major differences in the reported growth rates. Second, the underlying concept of consumption is different from that in earlier (and later) versions of the PWT. Instead of differentiating between household consumption expenditure (HCE) and government expenditure, PWT7 uses

[2]The aid data can sometimes be subject to revisions, which proved, however, not to have a substantial impact on the results as I will show in section 3.5.

[3]Note that *real* stands for *PPP adjusted* in the PWT, instead of *in constant prices*.

the concept of *actual individual consumption* (AIC) and *collective government consumption* (CGC). The difference between the two lies in the treatment of goods and services that are consumed by individuals but often paid for by the government, such as healthcare and education. AIC includes these expenditures, whereas HCE only includes such expenditures that are actually being paid for by the individual. While AIC arguably establishes better comparability across different economic systems, it has been discarded in later versions of the PWT as the required data are not readily available for most countries, increasing guesswork (Feenstra et al., 2013a).

The Penn World Table 8.0 (Feenstra et al., 2013b) introduces a wide range of methodological changes. Many of them specifically aim at increasing the consistency across (future) versions and reducing the amount of speculation in the reported data. As a consequence, 22 countries with particularly poor data coverage have been removed from the database, 3 of which were included in JMT's study, reducing our sample from 36 to 33 countries for PWT8. The new authors also discarded the RGDPL series JMT base their analysis on. I therefore use the conceptually most similar series included in the dataset, labelled RGDPna, which the authors confirm to be the measure most in line with previous versions of the PWT (Feenstra et al., 2015).[4] The series also corresponds to PPP adjusted GDP at 2005 prices, with the difference that the growth rates applied are GDP growth rates taken directly from the underlying national accounts data.[5] Note that the authors do not explicitly report the expenditure shares employed in the study at hand in the PWT release. They do, however, provide the underlying national accounts data which contains this information.

The World Bank's World Development Indicators (The World Bank, 2015) provide two different series of GDP at constant prices - in 2005 US dollars, and in PPP adjusted 2005 US Dollars. While the latter is conceptually closer to the PWT measures JMT and I employ, it only starts in 1990 for most countries. Note that, since all series are in constant prices, the temporal variation in prices is not directly reflected in the growth rates. Differences do, however, arise through the relative valuation of individual expenditure shares. For instance, a very low price for investment may substantially inflate the real GDP estimate in a period when investment is relatively high compared to the other measures (see section 3.3.3 below). However, there is no compelling reason to use PPP adjusted series in the

[4]RGDPna is widely unaffected by some of the more fundamental innovations in PWT8, which are reflected in the newly introduced series RGDPe and RGDPo, and aim at rendering GDP figures comparable across both time and space. The authors do however recommend RGDPna for single country analyses mainly concerned with growth rates (Feenstra et al., 2015).

[5]RGDPL used to be constructed from the growth rates of individual expenditure shares, with the result that changes in the underlying prices could induce major differences in overall GDP growth rates. Section 3.3.3 discusses this issue in some more detail.

present analysis other than data availability and conventions in the literature: By its nature, JMT's study (and therefore my analysis) does not rely on differences in income *levels* across countries, as it does not exploit the cross-sectional dimension. I therefore opt against the PPP adjusted series in WDI, in order to retain an already drastically reduced sample of 13 countries.[6]

3.3.2 Relative divergence of the datasets

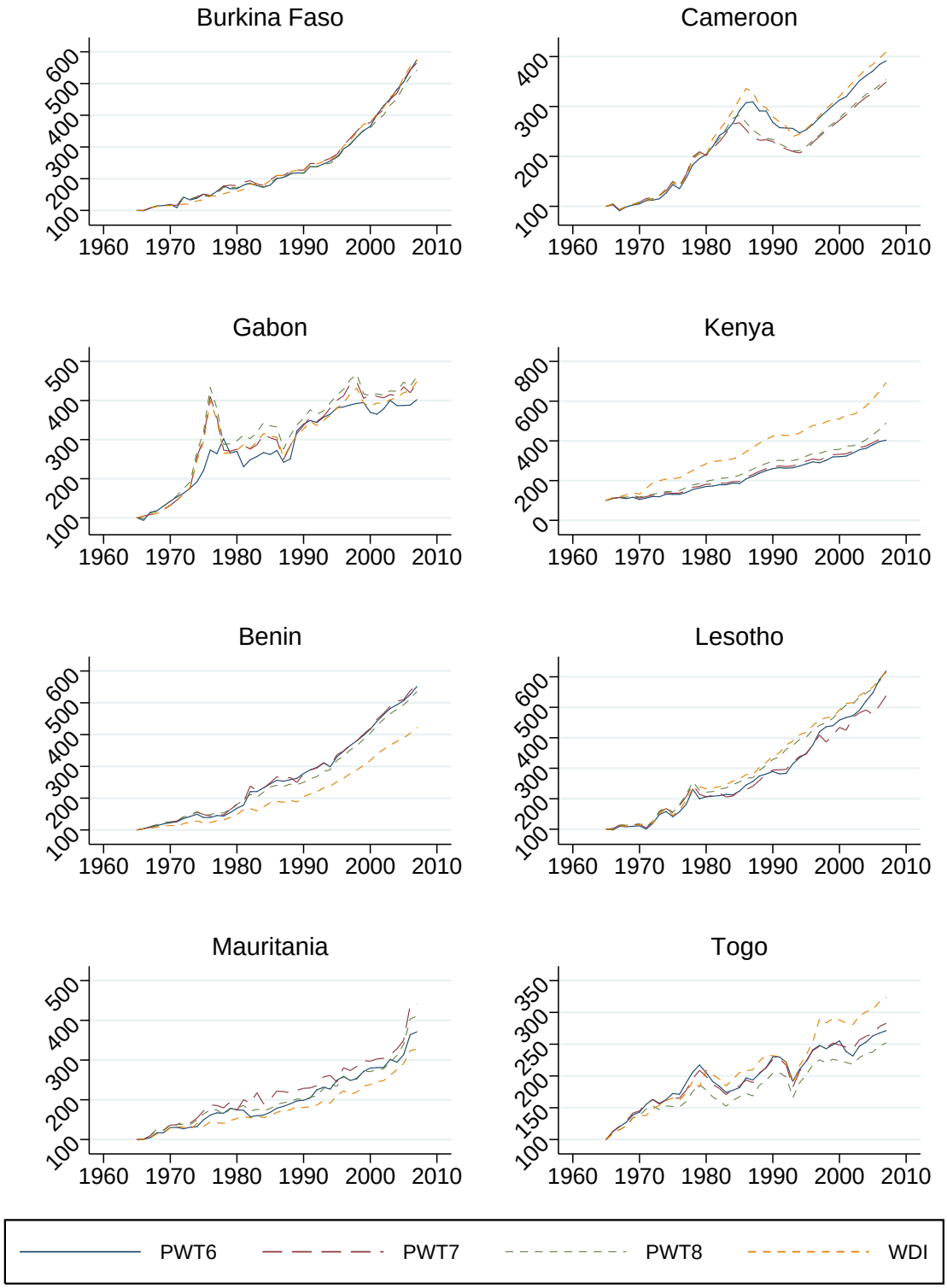
Figure 3.1 plots each of the GDP series for 8 selected countries; anticipating the findings of section 3.5.2, the four countries at the top of the panel are those where JMT's results remain the most stable in our replication exercise, and the countries at the bottom of the panel are those with the least consistent results. In order to illustrate differences in the reported composition of GDP, figure 3.2 plots the corresponding investment shares for those same countries.

All GDP series in figure 3.1 are normalised to their respective 1965 levels in order to abstract from persistent differences in levels, emphasising growth rates instead. Only in one country, Burkina Faso, GDP takes an almost identical trajectory in all four datasets. In other countries, the discrepancies look rather well behaved, that is, they occur only at specific points in time or concern only a single series. Kenya is such a case, where WDI indicates that GDP has increased by a factor of about 6.5 from 1965 to 2007, whereas the PWT measures agree on a factor of about 4. The opposite is the case in Benin, where WDI indicates a persistently lower growth rate subject to similar fluctuations as that reported by the PWT measures. In Cameroon, all measures follow an almost identical path until the mid-80's, but then split up: PWT8 and WDI register continued growth until the late 80's, followed by about a decade of recession, PWT6 and PWT7 start indicating a similarly severe recession earlier. The resulting differences are preserved in subsequent periods, where the datasets generally agree on the growth rates, but at different levels of (normalised) GDP.

In Togo, Mauritania, Gabon and Lesotho the differences are perhaps the most striking and least tractable. While the general patterns tend to be the same (they agree on major booms and recessions), differences arise throughout the period without following an apparent pattern, leading the graphs to intersect sometimes multiple times. That is, none of the datasets systematically under- or over-reports growth, but the sign of their relative bias varies over time. The most striking single discrepancy may be the one between PWT6 and the remaining series in the mid-70's in Gabon. This is almost entirely explained by different underlying

[6] While the coverage is larger for the GDP series as such, the corresponding expenditure shares are often missing.

Figure 3.1: GDP series from four sources, normalised to 1965 levels



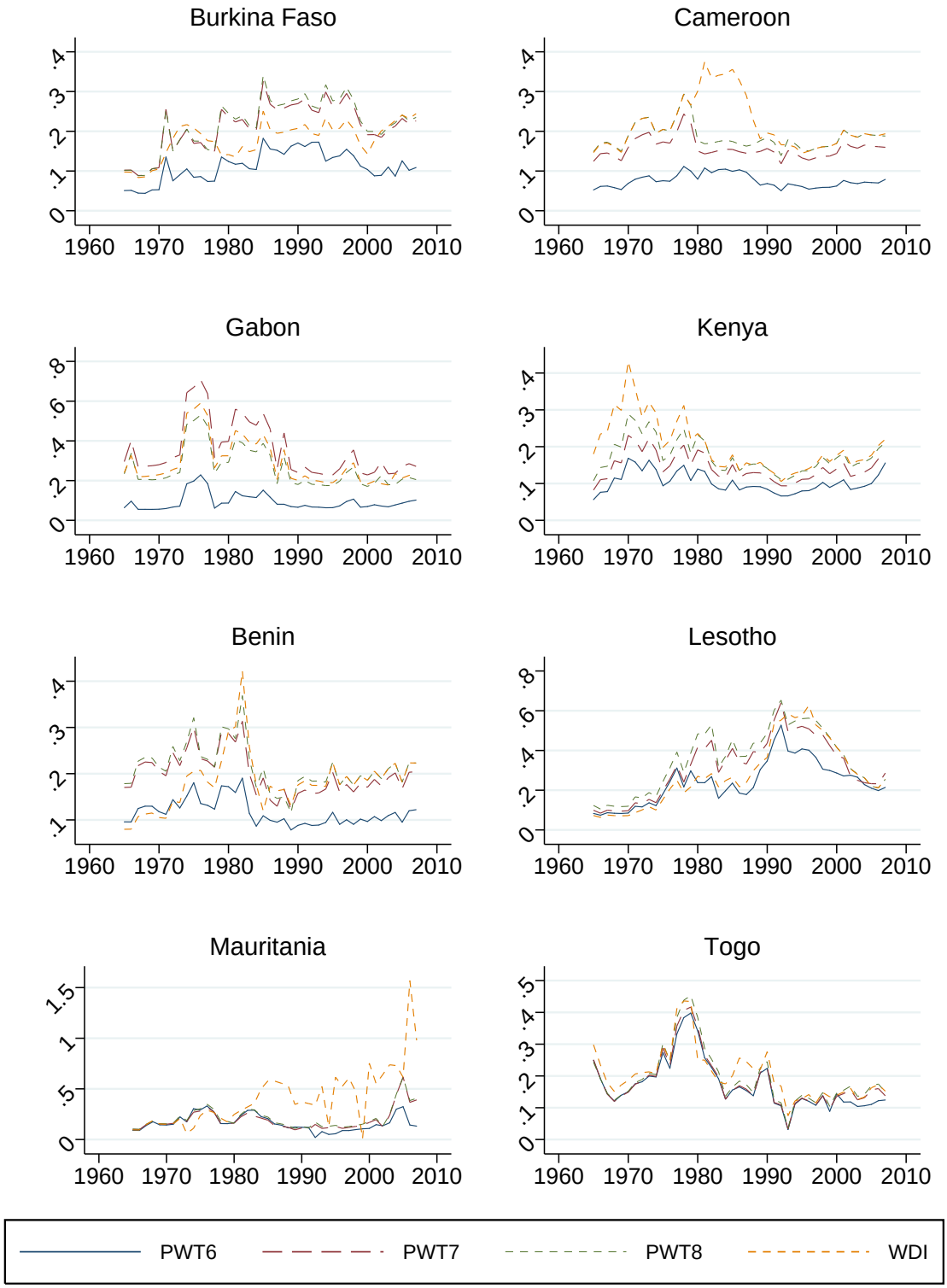
prices of investment, as will be discussed in more detail in section 3.3.3.

It is worth emphasising that there is no obvious pattern describing how the datasets behave relative to each other: In Gabon, PWT6 is the obvious outlier, in Cameroon (and in Lesotho, to some extent), PWT6 and PWT7 take one path, PWT8 and WDI the other, in Togo, PWT8 and WDI are precisely the ones the diverge most relative to each other. Only Kenya and Benin are consistent with the perhaps most intuitive expectation, that is, WDI diverging from otherwise consistent PWT measures. However, even in these two countries the divergence has opposite signs: in Kenya, WDI indicates higher growth, in Benin, it indicates lower growth than the remaining series.

In order to illustrate not only the divergence between GDP levels, but also in the relative movements of its components (which is the more influential aspect in the context of the study at hand), figure 3.2 plots the share of investment in GDP over the same period and for the same countries as those in figure 3.1. Note that the y-axis is scaled in order to depict a maximum of detail in the variation for each country. Its range is therefore informative in itself, and can vary with two factors: The temporal variation of the investment share within countries, and the discord between datasets regarding the investment share. For instance, investment in Lesotho varies from approximately 10% to more than 60% of GDP, but this is mainly due to temporal variation of similar amplitude in all datasets. In Cameroon, the scale is mainly stretched by upwards outliers in the 1980's in WDI (up to almost 40%), and consistently much lower estimates of less than 10% in PWT6. In most instances, the discrepancy is relatively constant over time, reflected in more or less parallel paths of the graphs; this applies in particular to Gabon, Kenya, and in a less pronounced manner to Burkina Faso, Togo, and Benin. The obvious outlier in the panel is Mauritania: While the PWT series already diverge substantially, this is dwarfed by the path suggested by WDI. Consistent with the other series until the early 1980's, the share of investment then skyrockets to levels of around 60% while the others agree on 10–20%, and reaches a peak of more than 150% of GDP (possible through a large trade deficit) where the PWT series report values between 20% and 40%.

Unlike in the GDP series depicted in figure 3.1, there seems to be a clear pattern across datasets in the investment share series: PWT6 almost consistently reports a lower share of GDP than the other datasets. This is mainly because PWT6 relies on price estimates from the ICP 1996, whereas the other series rely on those from 2005. The price estimates for investment are systematically lower when ICP 2005 prices are used, mechanically increasing the estimated real investment shares. Whether this reflects actual changes in prices, improved accuracy, or is a methodological artefact, is discussed in detail by Breton (2015) and Breton and

Figure 3.2: Shares of investment in GDP from four sources



García (2016), summarised in section 3.2.

Table 3.1: Correlations between growth rates of the core variables

		PWT6	PWT7	PWT8
Δ GDP	PWT7	0.76		
	PWT8	0.75	0.87	
	WDI	0.65	0.75	0.85
Δ Investment	PWT7	0.59		
	PWT8	0.72	0.76	
	WDI	0.04	0.05	0.07
Δ Consumption	PWT7	0.74		
	PWT8	0.50	0.61	
	WDI	0.34	0.44	0.20
Δ Government	PWT7	0.65		
	PWT8	0.65	0.96	
	WDI	0.51	0.75	0.75

Notes: The reported values are pairwise correlations between each of the variables in the four datasets employed. All correlations are significant at conventional levels, with the exception of those involving the WDI investment series. See main text for discussion.

Source: Author's calculations.

Table 3.1 reports the pair-wise correlations between the annual GDP growth rates, and the shares of investment, consumption and government expenditure across the four datasets employed in JMT's sample of countries in the analysis over the period from 1965 to 2007, subject to data availability. Given that they are conceptually in principle the same, it is a striking result that the *highest* correlation of the growth rates (first panel) between any of the datasets is 0.87 (PWT7 and PWT8), and goes as low as 0.65 (PWT6 and WDI). As much of the further analysis will focus on comparing results obtained from PWT6 (being JMT's original dataset) with those obtained from each of the other measures, it is also important to note that for PWT8 and WDI, the lowest correlation is precisely that with PWT6, and PWT7 is only marginally less correlated with WDI than with PWT6.

Looking at the second panel in table 3.1, the share of investment is only moderately correlated between PWT6 and PWT7 (0.59), and slightly more between PWT6 and PWT8 (0.72). Strikingly, seen from this perspective, the investment series in WDI shows almost no relationship with the PWT measures: all correlation coefficients are below 0.1, and none of them is statistically significant at the 5% level (the smallest p-value is 0.06 for WDI/PWT8). However, it needs to be pointed out that this is strongly driven by the case of Mauritania, where the

WDI investment series obviously bears almost no relationship with the remaining ones (figure 3.2). When excluding Mauritania, the correlations with investment in WDI become: 0.46 (PWT6), 0.51 (PWT7), 0.63 (PWT8).

The third and fourth panel of table 3.1 report the correlations for the remaining two measures, the growth in consumption and government expenditure. The magnitudes of these coefficients are more or less in line with those observed for investment. Note however that WDI is now less of an outlier, although it still features the poorest correlation with any series (at its worst, 0.2 with PWT8 in the consumption series).

A general and intuitive pattern (although not universal) in table 3.1 is that the correlations go down as we move away from the 'original' dataset – PWT6, by these measures, is more similar to PWT7 than to PWT8, and the least similar to WDI, which is both the most recent dataset and comes from a different data provider.

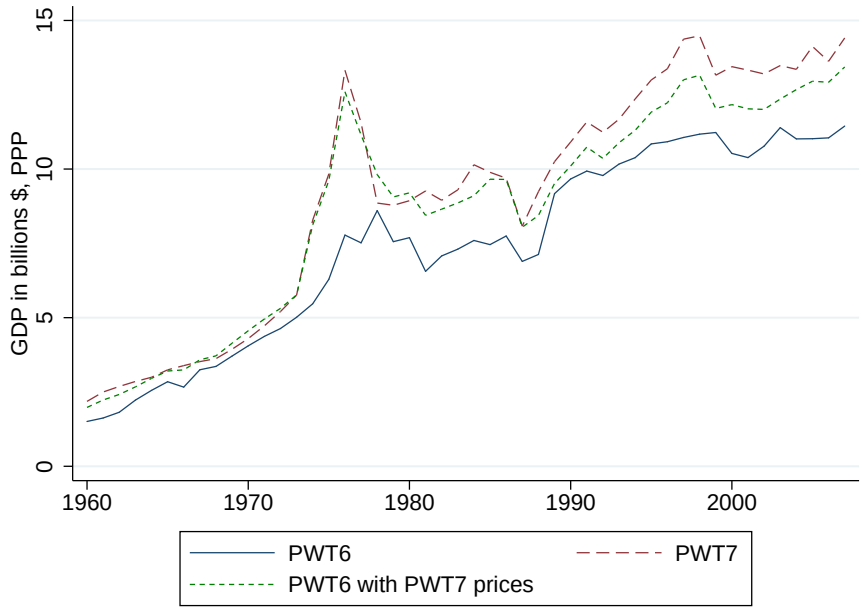
3.3.3 The relative importance of revisions to price and NA data

Some major differences between the datasets can be explained with differences in the underlying prices and PPP estimates. To illustrate this, figure 3.3 plots the constant price real GDP series (RGDPL) of Gabon as reported by PWT6 and PWT7. While the series generally seem to take similar paths, there are periods of striking divergence. The most obvious one is perhaps in the mid-1970's, where PWT7 reports an enormous increase in GDP that is almost completely missing in PWT6. At its peak in 1976, PWT7 reports a GDP figure 71% higher than PWT6. Although the series do converge slightly in the aftermath and then evolve almost in parallel, PWT7 consistently reports a higher GDP than PWT6, on average by 23%.

This changes as I recompute the GDP series using the methodology and the national accounts data underlying PWT6, but the prices underlying PWT7. As discussed above, the latter are derived from the ICP 2005, while PWT6 is based on ICP 1996 prices (see appendix B.1.1 for the details of the computation). The resulting series is plotted in green in figure 3.3: PWT6 and PWT7 are reconciled to a large extent, with a difference of less than 6% in 1976, compared to 71% before. The average difference between the series after 1976 drops from 23% to about 6%.

In fact, a closer look at the expenditure category prices in 2005 in Gabon reveals that the divergence between PWT6 and PWT7 almost entirely stems from the underlying price of investment (consistent with the findings of Breton and

Figure 3.3: GDP of Gabon according to PWT6 and PWT7, 1960-2007



Notes: The blue and the red line are the original RGDPL series from PWT6 and PWT7 respectively. The blue line is based on NA data and the methodology of PWT6, with the only difference that PPPs are constructed using PWT7 price data, reducing discrepancies dramatically. All series in 2005 $ (PPP).

García (2016) discussed in section 3.2). While all other prices are approximately the same, investment is estimated to cost 193% of the price in the US in 2005 (which is the reference point) in PWT6, but only 51% in PWT7. As a result, a nominally important increase in investment expenditures in the mid-1970's is highly *de*flated in real terms in PWT6, but *in*flated in PWT7. In PWT7, the increase in real investment that enters GDP is consequently estimated to be almost four times higher. When repeating this exercise for all 36 countries in the sample, PWT6 and PWT7 can be largely reconciled in almost all countries. The exceptions are Ghana, Liberia, Somalia and the Seychelles, where differences are mostly due to revisions to the NA data. The relevant plots and a more detailed discussion are provided in appendix B.1.2.

In order to assess the relative importance of changes in the national accounts data and prices more formally, I construct a counter-factual GDP series that is based on PWT7 NA data, but PWT6 price estimates for every country in the sample. Table 3.2 summarizes the extent to which each series diverges from the original PWT7 in terms of levels and growth rates. The divergence is quantified by their mean *absolute* deviation (MAD) and mean deviation (MD). MAD averages over the absolute differences between the series irrespective of their sign, and is therefore a good measure of general divergence between the series. In MD, positive

and negative differences can cancel each other out. It will therefore generally be smaller in magnitude, while its sign is indicative of the general direction of the divergence.

Table 3.2: Mean deviation and mean absolute deviation from PWT7

		Levels	Growth rates
PWT6	MAD	67.6%	3.3%
	MD	62.5%	0.1%
PWT6 NA, PWT7 prices	MAD	11.9%	2.3%
	MD	-6.3%	0.0%
PWT6 prices, PWT7 NA	MAD	59.3%	2.4%
	MD	57.0%	0.0%

As suggested by the visual inspection of the graphs, the prices play a much bigger role in explaining the divergence in levels between PWT6 and PWT7 than changes in the NA data. PWT6 differs, on average, by 67.6% from PWT7 in our sample - almost all deviations being positive, as indicated by the MD. When keeping the underlying NA data constant but applying PWT7 prices, this difference shrinks to 11.9%. The underlying NA data itself contributes much less to the divergence in terms of levels: When using PWT6 prices with PWT7 NA data, the difference to the PWT7 series remains high at 59.3% difference on average – a relatively small improvement over the original PWT6 series.

Since inference in times series studies is based on dynamics within a country over time, we are particularly interested in the growth rates. Consider the second column of table 3.2: The MAD between PWT6 and PWT7 is 3.3%. Compared to average growth rates reported in our sample of 36 countries between 1960 and 2007 of 3.7% (PWT6) or 3.8% (PWT7), this is a huge discrepancy of, on average, about 87% in any period. Note also that this divergence is not directed: The average growth rates are hardly different between PWT6 and PWT7: on average, PWT6 only indicates just about 0.1% higher growth rates. As opposed to the differences in levels, the underlying NA data and the prices have very similar impacts when it comes to growth rates: Reconstructing PWT6 using PWT7 prices reduces the MAD to 2.3%, using PWT6 prices but PWT7 NA data reduces it to 2.4%. Both changes in the underlying NA data and changes in the price series can therefore account for about one third if the differences in GDP growth rates between PWT6 and PWT7.

3.4 Baseline Study: Juselius, Møller and Tarp (2014)

Before proceeding to the robustness exercises, I will briefly outline the methodology of JMT, which I adopt. Section 3.4.1 describes the general CVAR framework, and section 3.4.2 describes the way in which JMT draw inference concerning the long-run effectiveness of aid.

3.4.1 The Cointegrated VAR Framework

JMT's paper analyses the long-run effect of foreign aid on key macroeconomic variables in 36 sub-Saharan African countries, specifying a Cointegrated VAR (CVAR) model for each country individually. While these models differ in terms of lag-length, deterministic components and rank restrictions, they share the basic structure that can be captured in the following moving average (MA) representation:

$$\mathbf{X}_t = \mathbf{C}\sum_{i=1}^{t}\varepsilon_i + \mathbf{C}\,\Phi\sum_{i=1}^{t}\mathbf{D}_i + \mathbf{C}^*(L)\varepsilon_t + \mathbf{C}^*\,\Phi\,\mathbf{D}_t + \mathbf{A}_0 \tag{3.1}$$

where $\mathbf{X}_t$ is a vector of $p = 5$ dependent variables, $\mathbf{X}_t = [aid_t,\ y_t,\ inv_t,\ c_t,\ g_t]'$ (inflow of foreign aid, total GDP, investment, private consumption and government expenditure in year t). $\mathbf{C}$ is a $p \times p$ matrix of rank $p - r$, r being the number of cointegrating relations between the variables. $\mathbf{D}_i$ ($\mathbf{D}_t$) are $m \times 1$ vectors containing the m deterministic components of the model at time i (t), such as trends in the variables or dummies accounting for extraordinary events. These enter the model weighted by the coefficients in the $p \times m$ matrix $\mathbf{\Phi}$. $\mathbf{C}^*(L)$ is a stationary lag polynomial and $\mathbf{A}_0$ contains the initial values of the variables in $\mathbf{X}_t$ and the initial values of the short-term dynamics. The interested reader may refer to JMT (p. 7-11), who lay out their methodology in more detail.

In the present analysis, the focus is on the long-run impact matrix $\mathbf{C}$, which has the following structure:

$$\mathbf{C} \quad = \quad \begin{array}{c} \\ aid_t \\ y_t \\ inv_t \\ c_t \\ g_t \end{array} \begin{array}{c} \begin{array}{ccccc} \hat{\varepsilon}_{aid} & \hat{\varepsilon}_y & \hat{\varepsilon}_{inv} & \hat{\varepsilon}_c & \hat{\varepsilon}_g \end{array} \\ \begin{pmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} \\ c_{21} & c_{22} & c_{23} & c_{24} & c_{25} \\ c_{31} & c_{32} & c_{33} & c_{34} & c_{35} \\ c_{41} & c_{42} & c_{43} & c_{44} & c_{45} \\ c_{51} & c_{52} & c_{53} & c_{54} & c_{55} \end{pmatrix} \end{array} = \begin{pmatrix} c_{11} & \mathbf{C_{12}} \\ \mathbf{C_{21}} & \mathbf{C_{22}} \end{pmatrix} \tag{3.2}$$

The element c_{ij} in row i and column j describes the impact that the cumulated exogenous shocks, measured as the residuals of the respective equation, on variable j have exerted on variable i in the long run. In this sense, for example, c_{21} can be interpreted as the long run effect of exogenous shocks to foreign aid on GDP.

In line with JMT, equation (3.2) provides a convenient decomposition of the matrix $\mathbf{C}$. The four coefficients describing the long-run impact of aid on the other variables are contained in $\mathbf{C}_{21}$. It follows from JMT's research question—the long-run effectiveness of foreign aid—that this will be at the centre of my discussion. $\mathbf{C}_{12}$ contains the long-run effects of the other macrovariables on aid, and is therefore informative when asking questions about the potential (and plausible) endogeneity of aid. This will however not be the subject of the study at hand. The $(p-1) \times (p-1) = 4 \times 4$ matrix $\mathbf{C}_{22}$ contains the remaining parameters, having analogous interpretations for the interaction between the macrovariables variables other than aid.

3.4.2 Inference in JMT

JMT's answer to the main question, the long-run impact of foreign aid on economic growth, is based on the sign and magnitude of the t-ratios associated with the coefficients of the vector $\mathbf{C}_{21}$. Aid is considered effective if either the coefficient of y_t, or inv_t, is positive and significant (absolute t-ratio larger than 2), or if both are. While the representation of the results also highlights marginally insignificant coefficients (absolute t-ratio between 1.6 and 2), these do not impact on the final inference. Aid harmfulness is defined analogously, with opposite signs. As their focus is on long-term growth, JMT do not consider the coefficients of c_t and g_t for their final inference. They do, however, discuss and report them, and I will consider them in the subsequent robustness checks.

In order to avoid a bias resulting from the researchers' economic prior, JMT consider the results from two different angles, adopting both the prior of effectiveness and that of harmfulness. The critical element here is the choice of the cointegration rank r, which especially in relatively short samples is not always clear cut and can have substantial impact on the inference. Instead of only determining one preferred choice of rank, JMT determine a best and second best choice (see also section 3.5.3). Under the prior of aid effectiveness, they report the preferred choice of rank if it indicates effectiveness. If the preferred choice of rank does not indicate effectiveness, but the second best choice does, they report the latter. The same procedure is then applied under the prior of harmfulness, but looking for negative coefficients.[7]

[7]JMT also report results using strictly the preferred rank, and first, second and third best

3.5 Assessing the impact of data discrepancies

In order to assess the impact of differences between the datasets, I proceed in two steps. In the first step, I apply precisely the statistical models derived by JMT to the alternative datasets. As these models have been tailored to the PWT6 data originally, it may however be the case that, when confronted to the new data, these models are technically misspecified, compromising the validity of any inference. In a second step, I therefore proceed to re-specifying the statistical models for each dataset individually, applying the same statistical criteria as JMT, in the 12 countries for which data is available from all four sources (PWT6, PWT7, PWT8, WDI). Before engaging in these exercises, I establish the criteria by which I will assess the robustness of the results.

3.5.1 Criteria to assess the stability of the results

Especially since JMT's mode of inference is rather unconventional, it is necessary to establish some sensible and objective criteria prior to assessing the stability of their results. My discussion will evolve around:

(i) The number of qualitative changes (that is, sign and/or significance) in *any* of the coefficients, i.e. those of y_t, inv_t, c_t and g_t. While JMT do not base their overall inference on c_t and g_t, they are still reported and discussed.

(ii) The number of qualitative changes in the *most relevant* coefficients, y_t and inv_t.

(iii) The number of qualitative changes in the *inference by country*. This is not equivalent to the previous point, as the inference is a joint product of the two coefficients of y_t and inv_t - even if both change sign/significance, we would still infer effectiveness as long as one of them is positive significant.

(iv) Any changes in the *overall conclusion*. JMT count 27 out of 36 cases of aid effectiveness. Irrespective of the results obtained for the previous criteria, this ratio may change or remain approximately constant, as reversals in one country may be compensated for by reversals in the opposite direction in other countries.

For criteria (i) and (ii) (coefficient specific), I consider changes in the coefficients a *reversal* if their absolute t-value changes at least from $|t| > 2$ to $|t| < 1.6$ (loses significance), or does the opposite (gains significance). The same is true if

choice of rank. Their results are largely robust to these alternative search algorithms.

the coefficient changes signs, unless it is insignificant in both instances (that is, in JMT's original study and in my replication).

Table 3.3 provides a full summary of this classification. The notation follows JMT: a plus (+) corresponds to a positive t-ratio larger than 2, meaning that the variable was positively affected by aid in the long run, while minus (−) indicates a t-ratio smaller than −2, indicating a negative relationship. The subscript 0 indicates absolute t-ratios between 2 and 1.6 (marginally significant), the subscript 00 an absolute t-ratio smaller than 1.6 (insignificant). I label coefficients *inconclusive* if the t-ratio remains of the same sign, and changes from fully insignificant ($|t| < 1.6$) to marginally significant ($2 > |t| > 1.6$), from marginally significant to significant ($|t| > 2$), or does any of these changes in the opposite direction.

Table 3.3: Classification of the replication results

		JMT					
		$-$	$-_0$	$-_{00}$	$+_{00}$	$+_0$	$+$
Replication	$-$	✓	·	✗	✗	✗	✗
	$-_0$	·	✓	·	·	✗	✗
	$-_{00}$	✗	·	✓	✓	·	✗
	$+_{00}$	✗	·	✓	✓	·	✗
	$+_0$	✗	✗	·	·	✓	·
	$+$	✗	✗	✗	✗	·	✓

Legend: ✓= Consistent; · = Inconclusive; ✗ = Reversal. + indicates a positive coefficient, − a negative one. The subscript '0' indicates an absolute t-ratio between 1.6 and 2, '00' indicates one lower than 1.6. The absence of a subscript indicates an absolute t-ratio > 2.

Regarding criterion (iii), the inference by country is well defined in the original paper and takes three possible values (effectiveness, harmfulness, or insignificant), and I classify any change from or to insignificance as inconclusive, and a change from harmfulness to effectiveness or vice versa as a reversal.

Overall inference, criterion (iv), boils down to a simple ratio between the number of cases of effectiveness and harmfulness respectively, and does not require further definition.

Note also that JMT's approach of adopting different economic priors (see section 3.4.2) adds a further complication: Cases can arise where, while the inference for a country (criterion (iii)) does not change, it will be based on a different choice of rank, with the originally reported rank now indicating a different result. I will count the rare cases where this occurs as consistent if the inference remains the same, no matter the rank this inference is based on.

3.5.2 Simple replication

The first replication uses the models developed in JMT, reported in table 2 of their paper. This involves a replication using the original data (PWT6, except for Sudan where JMT use WDI data) for all 36 countries, and the same exercise employing the alternative datasets. JMT's mode of inference of considering the first best and second best choice of rank requires two estimations per country and dataset (one for each choice of rank). Besides the original data for the 36 countries in JMT's sample, we have sufficient data for 36 (PWT7), 33 (PWT8), and 13 (WDI) of these countries in the alternative datasets, each of which has to be estimated twice, leaving us with 236 country-, dataset-, and rank-specific estimations. These have been carried out in an appropriately modified version of CATS in RATS version 2.6.[8]

Consistency across datasets

Table 3.4 summarises the key results from this first exercise. The four panels correspond to criteria (i)–(iv) as described above. The first two columns in table 3.4 report the results from the immediate replication of JMT, using the same data as them except for some possible revisions to the DAC aid data. Throughout the criteria, it is apparent that these revisions only have a very limited impact on the results: 127 out of 144 coefficients remain completely identical, and only 7 change in a significant way, that is, they are counted as reversals (panel (i)). Similar figures apply for the core coefficients, those of GDP and investment (panel (ii)). In all but one country (Mauritania), the final inference remains the same (panel (iii)). In Mauritania, the conclusion changes from effectiveness to harmfulness, leading us to conclude effectiveness in 26 countries under the prior of aid effectiveness, compared to 27 in JMT's original results. Under the prior of harmfulness, nothing changes and evidence for this hypothesis can be found in 10 countries.

Using alternative datasets has a sizeable impact on the results: The ratio of consistent coefficients (panel (i)) is just under two thirds in all alternative datasets (58%–63%), and about one quarter of them are reversed (23%–28%). Our main coefficients (panel (ii)) are no exception to this pattern. The last three rows of panel (ii) count whether either of the two main coefficients has changed sign or significance. In PWT7 and PWT8, these coefficients remained jointly stable in only 31% and 33% of all countries respectively, while at least one of the coefficients was subject to a reversal in 43% and 39% of the countries in the sample. WDI

[8]Replication programs and data can be obtained at `http://bit.ly/2sVH63e`. Parts of the relevant code are embedded in proprietary program files and can therefore not readily be shared, but detailed documentation on the necessary modifications can be provided.

Table 3.4: Summary of replication results

		Replication		Alternative Datasets					
		PWT6		PWT7		PWT8		WDI	
(i)	**Any coefficient**								
	Consistent	127	*88%*	89	*62%*	76	*58%*	33	*63%*
	Inconclusive	10	*7%*	15	*10%*	22	*17%*	7	*13%*
	Reversal	7	*5%*	40	*28%*	34	*26%*	12	*23%*
(ii)	**Main coefficients**								
	y_t								
	Consistent	33	*92%*	20	*56%*	21	*64%*	9	*69%*
	Inconclusive	2	*6%*	8	*22%*	5	*15%*	0	*0%*
	Reversal	1	*3%*	8	*22%*	7	*21%*	4	*31%*
	inv_t								
	Consistent	32	*89%*	20	*56%*	16	*48%*	9	*69%*
	Inconclusive	3	*8%*	3	*8%*	6	*18%*	3	*23%*
	Reversal	1	*3%*	13	*36%*	11	*33%*	1	*8%*
	y_t *or* inv_t								
	Consistent	29	*81%*	13	*31%*	12	*33%*	7	*54%*
	Inconclusive	5	*14%*	11	*26%*	10	*28%*	2	*15%*
	Reversal	2	*6%*	18	*43%*	14	*39%*	4	*31%*
(iii)	**Country-wise inference**								
	Consistent	35	*97%*	24	*67%*	20	*61%*	10	*77%*
	Inconclusive	0	*0%*	11	*31%*	11	*33%*	3	*23%*
	Reversal	1	*3%*	1	*3%*	2	*6%*	0	*0%*
(iv)	**Overall conclusion**								
	Prior: Effectiveness								
	Effective	26		18		13		6	
	Insignificant	*8*		*17*		*17*		*7*	
	Prior: Harmfulness								
	Harmful	10		9		7		3	
	Insignificant	*18*		*20*		*21*		*7*	
	Sample	**36**		**36**		**33**		**13**	

Notes: The table reports criteria (i)-(iv) as described in section 3.5.1 for all datasets when using the models as derived by JMT. The first column or each dataset reports the absolute number of coefficients (criteria (i) and (ii)) or countries (criteria (iii) and (iv)), the second column the respective share of the total number of coefficients or countries in the respective set-up.

Source: Author's calculations.

performed substantially better in this respect, with 54% of all countries having seen no changes in either of these coefficients.

The inference by country (panel (iii)) does not need to change in all of these cases, as a change in the coefficient of GDP can be compensated by a stable (or inconsistent in the opposite direction) coefficient of investment, and vice versa. In PWT7 and PWT8, the inference by country remains identical in 67% and 61% of cases, and gets reversed in 3% (1 case) and 6% (2 cases) respectively. Similar to what we saw for criterion 2, WDI yields results that are more consistent with JMT – inference is identical in 77% of the countries in the reduced sample (10/13) and is reversed nowhere.

Regardless of the relatively large variation in individual coefficients and country-wise inference, the overall conclusion of aid effectiveness (panel (iv)) remains unchallenged throughout all datasets; it is, however, slightly weakened. The count of cases of effectiveness versus cases of harmfulness (under the respective priors), goes down from 26:10 in my replication of JMT using PWT6 (27:10 in the original study) to 18:9 in PWT7, 13:7 in PWT8, and 6:3 in WDI. While for every case of harmfulness I (JMT) find 2.6 (2.7) cases of effectiveness using PWT6, there are about 2.0 cases of effectiveness for each case of harmfulness in the alternative datasets.

Note that while the results for WDI are proportionally speaking more in line with JMT than those obtained with PWT7 and PWT8, the significantly smaller sub-sample contained in WDI consists of countries that also exhibit more stable results in the other datasets. In the sub-sample of countries that are not covered in WDI, PWT7 and PWT8 yield consistent results only in 56% of the countries, compared to 77% in the sub-sample covered by WDI.[9] This could be reflecting a selection bias as a function of the quality of the original data, if for instance WDI were more conservative in 'filling the gaps'.

Consistency across countries

The share of consistent coefficients, excluding the PWT6 replication, ranges from 0% in Botswana and Lesotho, to 100% in Chad, Burkina Faso, Cameroon, Kenya and Gabon. The median proportion of consistent coefficients is 50%, the mean 57%. A table summarising the consistency by coefficients for each country can be found in appendix B.2.

In about half of the sample, 17 countries, the inference remains stable throughout all datasets. On the other hand, 9 countries do not yield the same inference

[9]The fact that this corresponds exactly to the ratio found in the replication using WDI is coincidental; only half of the reversed inferences are associated with the same countries as in WDI.

with any of the new data. Only in 3 out of a total of 82 cases (given sample sizes of 36, 33 and 13 in the alternative datasets), the inference is reversed. These are Lesotho in PWT7 and PWT8, and Liberia in PWT8; all three reversals correspond to a switch from effectiveness to harmfulness. All other changes in country-wise inference correspond to a loss of significance.

Note that it is difficult to predict whether the results for a country will be stable with respect to changes in the data or not. As documented in appendix B.3, a variety of measures of divergence (in levels, shares, relative shares) or data quality were only very weakly associated with the robustness of the results. In fact, the initial results seemed to be better predictors of their own consistency than any properties of the data: Initially insignificant coefficients were much more likely to remain insignificant in the replications than significant ones to keep their sign and significance. For the 8 countries that had no single significant coefficient in the original results, 80% of the coefficients were also insignificant in the replications. Within all other countries, 50% kept their sign and significance.

3.5.3 Re-specification for selected countries

The second exercise aims at exploring the full potential impact of the data on the results. It takes into account a fundamental aspect of the philosophy underlying the CVAR approach (and multivariate time series analysis more generally), namely that of 'allowing the data to speak freely' (Hoover et al., 2008, title): The priority lays in creating a statistically adequate representation of the data, avoiding strong theoretical assumptions *a priori*. In this spirit, each country-specific model developed by JMT, and thus applied in the previous section, has been specified as a function of the underlying data, that is PWT6 and the DAC ODA data. Given the sometimes striking divergence between the datasets, it seems reasonable to expect that differences may emerge not only in the results within the same models, but in the models themselves. I will therefore re-specify the models for the sub-sample on 12 countries where data is available for all datasets (the subset of countries covered by WDI bar Sudan, where JMT used WDI).

A priori, it is unclear what to expect from the results of this exercise. On the one hand, it might be the case that the changes observed in the previous exercise were partly a result of the models being effectively misspecified for the alternative datasets. In this case, it is possible that the data do indeed tell the same story, once analysed with adequately re-specified models. On the other hand, it may be the case that using the exact same models as JMT actually obscured the true impact of the changes in the data. Adjusting the models to the new data may then amplify the changes between datasets, and in fact induce even larger changes

in the results.

Model Specification Procedure

The models employed in JMT have the following variable elements: deterministic components such as trends and dummies, the lag-length, and the cointegration rank. While the econometric literature offers a plethora of formal criteria to specify either of these components, the sample sizes in this application ($T \approx 40$) undermine the power of some of the relevant tests. Furthermore, different criteria typically offer a slightly different angle at the data, and will sometimes indicate different choices. The resulting trade-offs eventually need to be resolved by the researcher's judgement. This section discusses the model specification procedure and reports the relevant criteria for each of the 48 models specified (4 datasets for 12 countries).

Deterministic components The choice of the deterministic components in this model consists of two elements, linear trends and dummy variables that account for extraordinary events or regime changes.

Linear trends Given the macroeconomic nature of our data, we would typically expect the presence of linear trends. These can be incorporated in the VAR in different ways, and the decision essentially boils down to the question whether the trends cancel in the cointegrating relations or not. In the case where they do cancel out, the most accurate specification would include an unrestricted constant term in the equation. If they do not cancel out, a trend that is restricted to the cointegrating relations should be included; this is the case when either some of the variables in $\mathbf{X}_t$ or any of the cointegrating relations are trend-stationary. For a detailed discussion, see Juselius (2006, chapter 6). One way of approaching the issue is to tentatively include a trend that is restricted to the cointegrating relations, which can then be tested for significance. In all of the present cases, this evidence unambiguously points towards the inclusion of a trend restricted to the cointegrating relations (see Juselius, 2006, p. 100, Case 4).

Dummy variables The more contentious, and less clearly defined choice, concerns the inclusion of dummy variables in order to account for extraordinary events such as droughts, floods, social unrest, or changes in equilibrium relations due to, e.g., regime changes. As it is difficult *a priori* to determine which historical events have an impact significant enough to enter the model, Juselius (2006, chapter 6.6) suggests to first scrutinise the data and the residuals from the baseline VAR, in order to determine where it is required to correct for outlier observations;

the modelling does thus first depend on the statistical evidence, which is then complemented by institutional and historical knowledge.

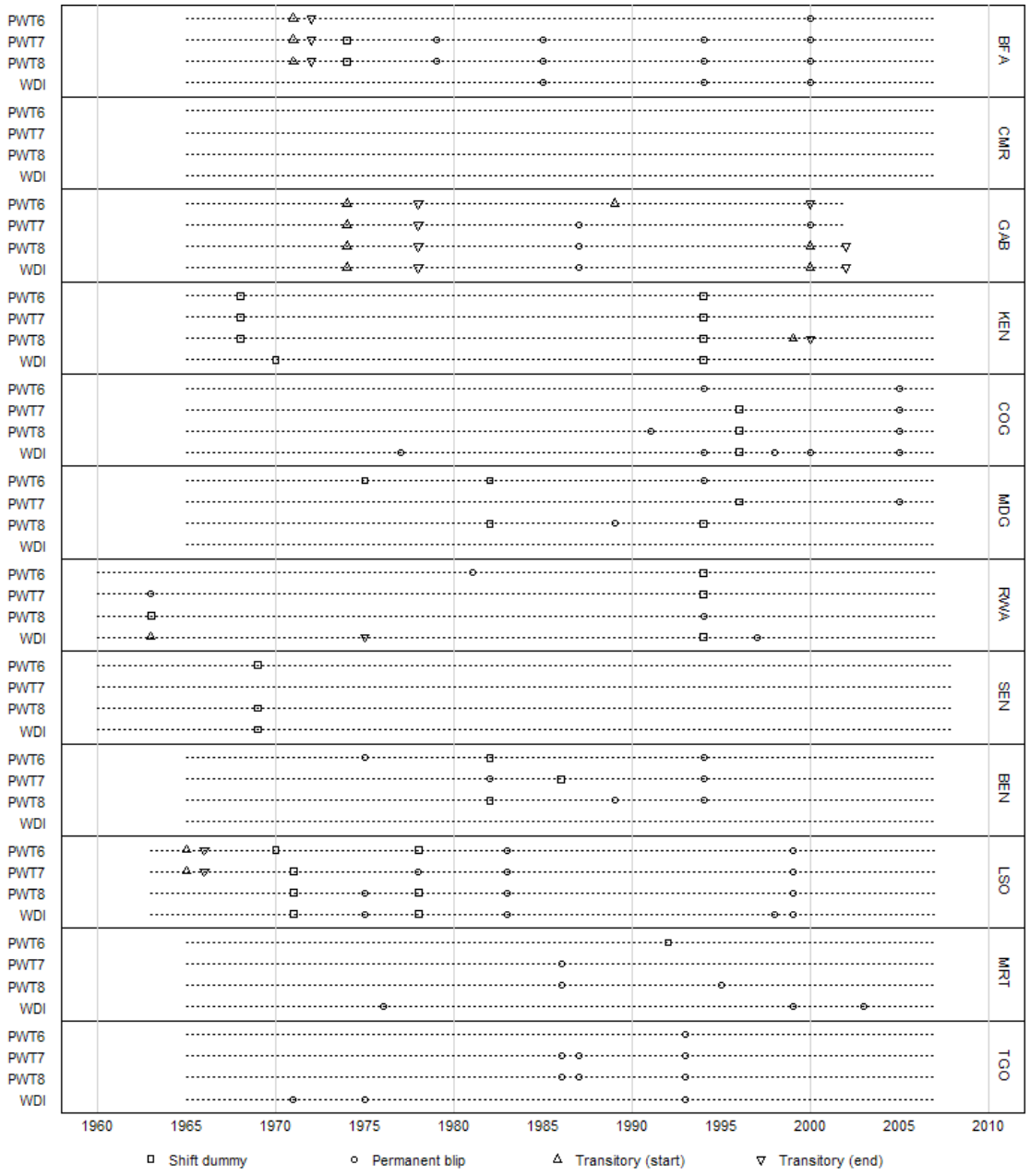
In line with JMT and Juselius (2006), I use three classes of dummies: Permanent blip dummies, labelled D_pZZ_t, having the structure $[0, 0, 1, 0, ..., 0]$, transitory blip dummies $D_{tr}ZZ_t$ with the structure $[0, 0, 1, -1, 0, ..., 0]$, and shift dummies D_sZZ_t, restricted to the cointegrating relations and taking the form $[0, 0, 1, 1, 1, ..., 1]$, indicating a shift in the equilibrium mean of the cointegrating relations. From a data perspective, the indications for the inclusion of a dummy are primarily derived from an inspection of the residuals. Large residuals, 'large' here being construed as corresponding to approximately 3 standard deviations, indicate that there may be an extraordinary event which, if not accounted for appropriately, would distort the analysis.[10] A unique blip in the error series, for instance, if it is not reverted by a shock in the opposite direction in one of the following periods, may then indicate a unique event (e.g., a drought) that permanently affected the economy. A temporary blip, for example, a large positive residual followed by a large negative residual in the following period, is an indication for a transitory intervention; typical cases would include a period of expansive monetary policy, compensated for by contractionary policy later on, or temporary fiscal stimuli. The determination of a shift dummy, D_sZZ, is in practice less straightforward, as it accounts for level changes in the long-run equilibrium, which are not readily observable from looking at the residuals. However, graphical inspection of the cointegrating relations can provide some indication. Furthermore, Juselius (2006, chapter III.9) proposes a battery of recursive and backward recursive tests that can provide indications for shifts in the equilibrium relations, which are taken into consideration.

For the final choice of the dummy variables, the statistical evidence is further complemented with historical data from the UCDP/PRIO Armed Conflict Dataset (Pettersson and Wallensteen, 2015) and the EM-DAT International Disaster Database (Guha-Sapir et al., 2014)[11], as well as knowledge about economically relevant historical episodes, most of which are documented in Ndulu et al. (2008). Figure 3.4 depicts the resulting dummy variables on timelines corresponding to the respective sample periods for each country and dataset, where the top line for each country corresponds to the model for the original PWT6 data, the second to PWT7, the third to PWT8 and the fourth to WDI. The presence of apparent bulks of dummy variables within each country across datasets illustrates the fact

[10]The outliers detected this way were generally consistent with those discerned through the dummy saturation procedure implemented in Autometrics (see Doornik, 2009). Due to the fundamental importance of institutional knowledge the present set-up, careful inspection of the residual series generally served as the primary source of information.

[11]The events documented in these two databases have been matched with the dummies included in all models in two tables available at `http://bit.ly/2tYKSu2`.

Figure 3.4: Timeline of dummy variables



that the datasets tend to agree on the impact of major events. Take for instance $Dp83_t$ in Lesotho, the year of a severe drought affecting most of the population (Guha-Sapir et al., 2014), reflected here in a dramatic drop in investment. Another consistent outlier is $Dtr7478_t$ in Gabon, corresponding to a unusually large temporary increase in investment, likely spurred by sharp increases in the price of oil after the 1973 oil crisis. Or the $Ds94_t$ shift dummy in Kenya, which coincides with a severe drought coupled with a dysentery epidemic costing about 1000 lives (*ibid.*). The latter also illustrates the trade-off one faces when weighing institutional knowledge against the statistical evidence. While the historical events are themselves not necessarily susceptible to induce a shift in long-run equilibrium relationships, the statistical evidence suggests precisely this, with the recursive test for fluctuations of the eigenvalue indicating non-constancy of our α and β parameters, with a sharp decline in the test statistic in 1994 for the third cointegrating relation. The inclusion of $Ds94_t$ establishes constancy over the entire sample period, justifying the inclusion both in JMT's and in my models.

On the other hand, there are about 25 dummy variables that are specific to the respective dataset, such as in Togo in 1971 (a decline in GDP and investment in the year of a drought affecting about 150.000 people (Guha-Sapir et al., 2014)) or 1975/76 (a spike in consumption and government spending coinciding with soaring phosphate prices, one of Togo's staple exports (Ndulu et al., 2008, Vol. 1, p. 150)). Particularly divergent, in terms of the modelling of extraordinary events, is Mauritania, where the 7 dummies included across the models occur in 6 different years, thus only agreeing once. Nevertheless, all of them can plausibly be attributed to historical events (1976 is the year of a severe drought effectively affecting the entire population; 1986 falls within a period of severe ethnic tensions and shortly follows a coup d'état by Ould Taya; 1992, a shift in equilibrium means, is the year of the adoption of a new constitution establishing the fourth republic; 1995, 1999 and 2004 have all seen major floods (Guha-Sapir et al., 2014; Seely, 2005)).

Note that some of the peculiarities in the data cannot that clearly be attributed to historical events, in which case the trade-off is between the statistical fit of the model, and its institutional/historical justifiability. The 1971/72 transitory dummy in Burkina Faso, indicated in all datasets except WDI, is such a case. While this falls within a period of some political instability, with a tightening of military control and a decline of the civilian role (Ndulu et al., 2008, Vol. 2, p.8 Appendix), there is no compelling historical evidence justifying the transitory nature of the shock. It is, however, by some margin the specification that best fits the pattern of the residuals, and my models therefore follow JMT in accounting for it with $Dtr71_t$. In other cases, the trade-off has been resolved in favour of

institutional adequacy, for instance in the case of an outlier in 1978 in the PWT7 Cameroon model: The model achieves a reasonable fit overall without a correction, and I could not identify a strong case for an extraordinary event in the country's history.

Overall, JMT's and my models (consistent in PWT6 except for Burkina Faso, the Republic of Congo and Madagascar, where the aid data appears to have significantly changed the requirements) follow the same patterns and account for the statistical and historical evidence in a coherent manner. To summarise, the data tends to tell a similar story – stylised in figure 3.4 – but puts varying emphasis on individual episodes.

Choice of lag-length Determining the optimal lag-length of the resulting VARs boils down to a trade-off between preserving a maximum of information, and retaining a reasonable number of parameters, especially given the relatively small number of ca. 40 observations for each country (certainly *small* in the context of a time series analysis). My choices are based on three groups of criteria, in line with Juselius (2006): (i) a likelihood test for lag reduction, (ii) information criteria, namely the Schwarz Criterion (SC) and the Hannan-Quinn Criterion (HQC), and (iii) a Lagrange-Multiplier test for no autocorrelation of the residuals. The p-values associated with (i) and (iii), as well as the values of the (ii) ICs are collected in table 3.5.

The (i) test for lag reduction tests for a significant decrease in the (log-) likelihood when adopting a more parsimonious model, which indicates a poorer fit of the latter. The first column of table 3.5 reports the outcome only of the test for the reduction to the final choice of lag-length k^* from $k^* + 1$ in order to save space. Note that this is rejected in almost all cases, including those exactly replicating JMT's models (all PWT6 except Burkina Faso, Congo, and Madagascar), indicating that higher lags may still contain valuable information. Although not explicitly reported here, this test rejects almost any reduction up to at least the fourth lag, which was the highest lag included in the procedure, therefore indicating prohibitively high autoregressive orders given the small sample sizes we are facing. This is because it does not account for the fact that each new lag increases the number of parameters by p^2, thereby rapidly consuming the already scarce degrees of freedom.

The (ii) information criteria, explicitly designed to take this trade-off into account, will therefore tend to indicate shorter lag-lengths. Columns 3 to 6 of table 3.5 report the values of the SC and the HQC for both k^* and $k^* + 1$; the lag length associated with the smaller value is considered more favourable by the respective criterion. In all cases but five, the ICs unambiguously indicate that

Table 3.5: Criteria for the choice of the lag-length

		Schwarz		Hannan-Quinn		Autocorrelation			
Data	*LR*	k^*	k^*+1	k^*	k^*+1	$k^*,1$	$k^*+1,1$	$k^*+1,2$	k^*
Burkina Faso									
PWT6	0.00	-22.44	-21.39	-23.66	-23.28	0.47	0.74	0.39	1
PWT7	0.00	-24.57	-23.65	-26.46	-26.21	0.98	0.91	0.14	1
PWT8	0.00	-24.42	-23.46	-26.31	-26.02	0.97	0.85	0.09	1
WDI	0.00	-22.72	-21.87	-24.07	-23.89	0.24	0.02	0.41	1
Cameroon									
PWT6	0.12	-23.57	-22.10	-24.51	-23.72	0.57	0.75	0.60	1
PWT7	0.00	-24.07	-22.94	-25.01	-24.56	0.31	0.17	0.02	1
PWT8	0.00	-23.67	-22.61	-24.62	-24.23	0.17	0.26	0.03	1
WDI	0.90	-22.10	-20.21	-23.04	-21.82	0.97	0.08	0.75	1
Gabon									
PWT6	0.00	-16.83	-15.45	-18.43	-18.07	0.40	0.42	0.17	1
PWT7	0.00	-17.72	-16.20	-19.61	-19.25	0.56	0.72	0.21	1
PWT8	0.00	-17.92	-16.45	-19.81	-19.51	0.56	0.53	0.13	1
WDI	0.00	-17.88	-16.50	-19.77	-19.55	0.43	0.54	0.34	1
Kenya									
PWT6	0.00	-23.88	-23.21	-25.36	-25.36	0.15	0.61	0.76	1
PWT7	0.00	-24.82	-24.21	-26.30	-26.37	0.09	0.52	0.32	1
PWT8	0.00	-24.95	-24.18	-26.57	-26.47	0.20	0.04	0.85	1
WDI	0.00	-24.88	-24.33	-26.36	-26.48	0.01	0.12	0.37	1
Congo, Rep.									
PWT6	0.00	-16.21	-15.32	-17.17	-16.96	0.46	0.11	0.85	1
PWT7	0.03	-16.91	-15.60	-18.28	-17.65	0.50	0.41	0.41	1
PWT8	0.00	-17.55	-16.44	-19.06	-18.63	0.10	0.03	0.27	1
WDI	0.00	-16.17	-15.63	-18.08	-18.23	0.02	0.97	0.05	1
Madagascar									
PWT6	0.00	-26.10	-24.96	-27.74	-27.29	0.19	0.00	0.01	1
PWT7	0.02	-25.00	-23.73	-26.64	-26.06	0.72	0.05	0.18	1
PWT8	0.01	-24.67	-23.47	-26.32	-25.79	0.83	0.52	0.07	1
WDI	0.02	-24.14	-22.86	-25.10	-24.50	0.15	0.02	0.01	1
Rwanda									
PWT6	0.00	-20.53	-19.89	-21.90	-21.94	0.12	0.29	0.23	1
PWT7	0.00	-18.30	-17.44	-20.36	-20.17	0.81	0.45	0.31	2
PWT8	0.00	-20.24	-19.04	-22.29	-21.78	0.26	0.02	0.11	2
WDI	0.00	-21.15	-20.41	-22.65	-22.60	0.74	0.20	0.20	1
Senegal									
PWT6	0.26	-26.71	-25.09	-27.96	-27.03	0.80	0.92	0.25	1
PWT7	0.17	-24.96	-23.40	-26.62	-25.76	0.15	0.29	0.86	2
PWT8	0.35	-26.62	-27.87	-27.87	-26.89	0.81	0.90	0.39	1
WDI	0.01	-24.71	-23.54	-25.96	-25.48	0.07	0.39	0.75	1
Benin									
PWT6	0.07	-25.78	-24.39	-27.26	-26.54	0.36	0.41	0.71	1
PWT7	0.00	-25.02	-23.75	-26.50	-25.91	0.34	0.34	0.14	1
PWT8	0.01	-24.58	-23.35	-26.06	-25.51	0.85	0.75	0.37	1
WDI	0.00	-24.12	-22.88	-25.06	-24.50	0.15	0.03	0.03	1
Lesotho									
PWT6	0.00	-20.49	-19.80	-22.32	-22.29	0.52	0.16	0.48	1
PWT7	0.00	-21.50	-20.58	-23.20	-22.93	0.84	0.77	0.83	1
PWT8	0.00	-22.85	-22.13	-24.69	-24.62	0.42	0.06	0.08	1
WDI	0.02	-20.79	-19.57	-22.23	-21.66	0.82	0.23	0.94	1

Continued on next page

Table 3.5 (continued)

		Schwarz		Hannan-Quinn		Autocorrelation			
Data	*LR*	k^*	k^*+1	k^*	k^*+1	$k^*,1$	$k^*+1,1$	$k^*+1,2$	k^*
Mauritania									
PWT6	0.00	-17.80	-16.84	-19.01	-18.73	0.04	0.37	0.28	1
PWT7	0.11	-18.05	-16.60	-19.13	-18.35	0.74	0.21	0.57	1
PWT8	0.01	-17.76	-16.53	-18.98	-18.42	0.71	0.48	0.99	1
WDI	0.07	-16.96	-15.55	-18.31	-17.57	0.30	0.49	0.70	1
Togo									
PWT6	0.00	-21.42	-20.47	-22.50	-22.23	0.06	0.47	0.34	1
PWT7	0.00	-21.24	-20.53	-22.59	-22.55	0.13	0.91	0.35	1
PWT8	0.00	-21.02	-20.33	-22.37	-22.35	0.13	0.80	0.31	1
WDI	0.02	-19.21	-17.97	-20.56	-19.99	0.08	0.83	0.43	1

Notes: The reported values are p-values with the exception of the Schwarz Criterion and the Hannan-Quinn Criterion. k^* is the eventually inferred optimal lag length.
Source: Author's calculations.

k^* is preferable over k^*+1; and $k^*=1$ in all cases but Rwanda PWT7 and PWT8, and Senegal PWT7, where $k^*=2$. In Kenya, HQC favours a lag-length of $k=2$ in PWT6, PWT7 and WDI. Similarly, the ICs disagree in Congo WDI, Rwanda PWT6, and Senegal PWT8. In these ambiguous cases, I opt for the more parsimonious specification. The fact that the ICs almost unambiguously indicate a short lag length is not surprising; both SC and HQC penalise the inclusion of new parameters more harshly, the shorter the T, and will therefore apply particularly high penalties in the short samples at hand.

Finally, columns 7-9 report the p-values from the (iii) LM test for no residual auto-correlation in the first lag of the VAR(k^*) models, and in the first and second lag of the VAR(k^*+1) models. The null hypothesis of no autocorrelation is rejected at the 5% level only three times in the first lag in the VAR(k^*) models (Kenya WDI, Mauritania PWT6, and Congo WDI). In most cases, it does not oppose k^*+1 either, although the test indicates residual autocorrelation seven times in the first lag and four times in the second lag.

Overall, all datasets appear to be best described with a lag-length of $k=1$, with the exception of Rwanda PWT7 and PWT8, and Senegal PWT7. The most contentious choices are arguably Kenya and Congo in WDI, where HQC favours a higher lag-length, and there is some evidence for residual autocorrelation in the first lag. In the light of otherwise reasonable properties of the model (discussed in the next section), I stick to $k=1$ as indicated by SC in these cases, but note it as a possible caveat when interpreting the results.

Tests for Misspecification Before proceeding to the choice of the cointegration rank, it is worth checking the models for misspecification, as not only the

trace test, but the VAR model as such, rely on a number of assumptions - most crucially, that of normal and independent residuals. Table 3.6 reports the p-values for tests for autocorrelation in the second lag in the final model (the first lag having been considered and reported as a criterion for the choice of the lag length before, see table 3.5), multivariate normality, and autoregressive conditional heteroskedasticity (ARCH) in the first and second lag. The last column reports the trace correlation, roughly interpretable as an average R^2 over the five equations in the model, thus summarising its overall fit.

Note that there is no indication for residual autocorrelation in the second lag in any of the models when applying a 5% significance level, and only in three cases (Gabon PWT7, Benin PWT8, Madagascar PWT6) the test would reject at the 10% level (first column, table 3.6). Coupled with the overall good results with respect to the first lag reported in table 3.5, residual autocorrelation does not appear to be a typical problem in the models.

The second column reports the p-values resulting for the test of multivariate normality suggested by Doornik and Hansen (2008), based on a transformation of the relevant moments (kurtosis and skewness) proposed by Shenton and Bowman (1977). In the final models, multivariate normality is almost never rejected at the 5% level. The exceptions are Kenya PWT8, and Lesotho in all datasets but PWT8, which also includes the PWT6 model, identical in its specification to JMT's and yielding output almost identical to the second decimal place (meaning that the impact of differences in the aid data is negligible, and the original specification almost certainly relied on the same test statistic). In these cases, no sensible variations in the specification could rectify the issue. This illustrates the limits of heterogeneity even in this comparatively highly flexible framework: A VAR including the present set of variables has limited validity in these cases, and more fundamental changes in the framework may be required to adequately represent the macroeconomic dynamics described by the data. As the focus of the study at hand is to assess the robustness of the results within the framework provided by JMT, we refrain from more fundamental changes to the specification.

Similarly, in many models there is some evidence for the presence of ARCH effects (column 3 and 4), both in the PWT6 models consistent with JMT, and in the re-specified ones. This could be another factor undermining the performance of the trace test, which has however been shown to be relatively robust in this respect (Rahbek et al., 2002).

The trace correlation, reported in the last column, is quite persistent within each country and across datasets, indicating that the re-specified models typically reach a similar fit as the ones derived and employed by JMT.

Table 3.6: Misspecification tests

Country / Data	*Autocorr.*	*Norm.*	ARCH(1)	ARCH(2)	*Trace Corr.*
Burkina Faso					
PWT6	0.35	0.42	0.05	0.02	0.40
PWT7	0.37	0.11	0.32	0.07	0.63
PWT8	0.42	0.11	0.21	0.05	0.62
WDI	0.72	0.59	0.10	0.08	0.51
Cameroon					
PWT6	0.59	0.32	0.20	0.01	0.32
PWT7	0.47	0.28	0.03	0.04	0.28
PWT8	0.73	0.36	0.01	0.02	0.30
WDI	0.48	0.38	0.01	0.02	0.36
Gabon					
PWT6	0.46	0.09	0.00	0.21	0.55
PWT7	0.07	0.31	0.00	0.05	0.68
PWT8	0.23	0.18	0.00	0.04	0.64
WDI	0.11	0.30	0.00	0.03	0.65
Kenya					
PWT6	0.51	0.28	0.65	0.29	0.52
PWT7	0.45	0.25	0.54	0.50	0.19
PWT8	0.91	0.01	0.36	0.01	0.59
WDI	0.90	0.31	0.17	0.08	0.56
Congo, Rep.					
PWT6	0.32	0.05	0.52	0.31	0.41
PWT7	0.43	0.32	0.06	0.02	0.44
PWT8	0.64	0.33	0.04	0.02	0.51
WDI	0.25	0.06	0.00	0.01	0.68
Madagascar					
PWT6	0.07	0.23	0.00	0.01	0.70
PWT7	0.34	0.08	0.01	0.08	0.65
PWT8	0.05	0.15	0.00	0.00	0.64
WDI	0.11	0.42	0.02	0.02	0.35
Rwanda					
PWT6	0.23	0.25	0.04	0.05	0.52
PWT7	0.11	0.25	0.34	0.06	0.63
PWT8	0.27	0.45	0.46	0.11	0.66
WDI	0.04	0.18	0.63	0.15	0.59
Senegal					
PWT6	0.88	0.17	0.55	0.30	0.47
PWT7	0.28	0.60	0.06	0.13	0.41
PWT8	0.95	0.11	0.54	0.35	0.46
WDI	0.98	0.37	0.06	0.05	0.39

Continued on next page

Table 3.6 (continued)

Country / Data	*Autocorr.*	*Norm.*	ARCH(1)	ARCH(2)	*Trace Corr.*
Benin					
PWT6	0.29	0.21	0.01	0.02	0.65
PWT7	0.32	0.23	0.00	0.01	0.58
PWT8	0.07	0.14	0.01	0.01	0.58
WDI	0.11	0.42	0.02	0.02	0.35
Lesotho					
PWT6	0.10	0.00	0.67	0.03	0.63
PWT7	0.57	0.01	0.53	0.66	0.62
PWT8	0.21	0.06	0.02	0.05	0.67
WDI	0.14	0.00	0.05	0.03	0.57
Mauritania					
PWT6	0.10	0.37	0.23	0.06	0.38
PWT7	0.99	0.21	0.63	0.02	0.37
PWT8	0.91	0.41	0.26	0.03	0.38
WDI	0.62	0.18	0.45	0.30	0.55
Togo					
PWT6	0.34	0.55	0.06	0.05	0.44
PWT7	1.00	0.92	0.21	0.17	0.53
PWT8	0.99	0.91	0.16	0.17	0.53
WDI	0.42	0.06	0.24	0.06	0.48

Notes: All reported values are p-values, with the exception of *Trace Corr.*, which is the trace correlation. *Autocorr., Norm.*, and ARCH are tests for no residual autocorrelation, multivariate normality and no ARCH effects respectively, further discussed in the main text.

Source: Author's calculations.

Cointegration rank r Perhaps the most influential and often contentious choice is that of the cointegration rank, that is, the rank r of the long-run coefficient matrix $\mathbf{\Pi}$. Given the short sample period, the standard procedure for the determination, the trace test (Johansen, 1988) has very low power in the current set-up (JMT, p. 14). It therefore fails to reject unit roots at ranks that are both economically and statistically implausible, meaning that it will tend to indicate unreasonably low ranks.

In line with JMT, a number of criteria are employed in order to assure a well-grounded choice of r. Apart from the aforementioned (i) trace test, I will base my choice of rank on (ii) the largest unrestricted roots of the companion matrix, (iii) the t-ratios of the α-coefficients, and (iv) a visual inspection of the graphs of the cointegrating relations. These are the same criteria employed by JMT (p. 14); for a more comprehensive discussion, refer to Juselius (2006, chapter 8.5).

As reported in the first column of table 3.7, and in line with the above mentioned low power of the trace test for cointegration (i), it systematically suggests

Table 3.7: Criteria for the choice of rank

	Trace test		Largest root		max $\|t\|$ in $\boldsymbol{\alpha}_{r^*}$			
Country / Data	*Inf.*	*p-value*	r^*	r^*+1	r^*	r^*+1	*Graph*	$r^*(r')$
Burkina Faso								
PWT6	0	0.52	0.32	0.86	2.98	3.12	1(2)	2(1)
PWT7	2	0.21	0.59	0.63	-4.1	-4.54	2(1)	2(1)
PWT8	2	0.17	0.54	0.61	-4.17	4.42	2(1)	2(1)
WDI	1	0.18	0.34	0.63	4.51	3.21	2(3)	3(2)
Cameroon								
PWT6	0	0.08	0.63	0.74	-4.66	1.34	3(2)	3(2)
PWT7	0	0.55	0.64	0.68	-2.44	-1.18	2(3)	3(2)
PWT8	0	0.25	0.66	0.69	2.55	1.44	2(3)	3(2)
WDI	1	0.11	0.6	0.73	3.84	-2.59	3(2)	3(2)
Gabon								
PWT6	1	0.21	0.92	0.92	4.16	-2.89	3(4)	3(4)
PWT7	3	0.19	0.38	0.44	7.66	-3.46	3(4)	3(4)
PWT8	2	0.16	0.87	0.99	-3.67	-4.57	3(4)	3(2)
WDI	2	0.08	0.93	0.99	5.92	-2.98	3(4)	3(2)
Kenya								
PWT6	1	0.20	0.61	0.83	4.63	4.31	2(3)	3(2)
PWT7	1	0.06	0.67	0.85	-2.86	-3.71	2(3)	3(2)
PWT8	3	0.22	0.55	0.85	3.53	-4.49	2(3)	3(2)
WDI	3	0.08	0.64	0.78	4.41	5.02	4(3)	3(4)
Congo, Rep.								
PWT6	1	0.52	0.64	0.77	3.65	3.71	2(3)	2(3)
PWT7	1	0.43	0.66	0.81	2.83	-3.83	3(2)	3(2)
PWT8	1	0.26	0.69	0.80	3.18	3.75	2(3)	3(2)
WDI	4	0.28	0.53	0.91	3.58	-2.54	3(4)	4(3)
Madagascar								
PWT6	5	0.00	0.52	0.63	-2.71	-2.96	4(3)	4(3)
PWT7	3	0.31	0.51	0.64	2.87	-2.50	3(4)	4(3)
PWT8	3	0.38	0.21	0.66	-6.58	-3.83	3(2)	3(2)
WDI	1	0.14	0.35	0.79	-3.61	-2.00	2(3)	2(3)
Rwanda								
PWT6	2	0.27	0.48	0.50	-3.64	3.13	2(3)	3(2)
PWT7	1	0.72	0.55	0.89	-3.2	-2.52	2(1)	2(1)
PWT8	1	0.35	0.58	0.87	5.73	-2.67	2(1)	2(1)
WDI	2	0.24	0.67	0.70	-4.23	3.02	3(2)	3(2)
Senegal								
PWT6	2	0.14	0.28	0.86	-3.51	2.9	2(3)	3(2)
PWT7	0	0.51	0.21	0.87	3.45	2.22	3(2)	3(2)
PWT8	2	0.17	0.32	0.84	3.28	2.84	2(3)	3(2)
WDI	1	0.18	0.59	0.8	-2.73	-2.27	2(3)	2(3)
Benin								
PWT6	3	0.23	0.36	0.65	-4.83	3.18	3(4)	3(4)
PWT7	2	0.33	0.31	0.82	4.23	-1.45	3(2)	3(2)
PWT8	2	0.10	0.24	0.63	3.97	-3.12	3(2)	3(2)
WDI	0	0.16	0.25	0.35	-4.34	-3.61	2(1)	2(1)

Continued on next page

Table 3.7 (continued)

	Trace test		Largest root		max $\|t\|$ in $\boldsymbol{\alpha}_{r^*}$			
Country / Data	*Inf.*	*p-value*	r^*	r^*+1	r^*	r^*+1	*Graph*	$r^*(r')$
Lesotho								
PWT6	5	0.00	0.85	0.89	4.6	4.46	1(2)	3(2)
PWT7	4	0.07	0.55	0.83	5.02	-3.53	2(1)	2(1)
PWT8	5	0.00	0.55	0.6	5.76	5.01	2(3)	3(2)
WDI	2	0.37	0.58	0.82	5.34	-3.15	1(2)	2(1)
Mauritania								
PWT6	0	0.23	0.53	0.63	-3.17	-2.68	2(3)	3(2)
PWT7	0	0.20	0.41	0.77	-4.54	-2.94	3(2)	3(2)
PWT8	0	0.33	0.45	0.62	-4.27	-4.08	2(3)	2(3)
WDI	2	0.41	0.00	0.61	4.09	3.28	2(1)	2(1)
Togo								
PWT6	2	0.41	0.34	0.68	-4.96	-3.01	2(1)	2(3)
PWT7	3	0.26	0.59	0.66	4.53	-1.85	3(2)	3(2)
PWT8	3	0.17	0.63	0.68	-4.93	1.95	3(2)	3(2)
WDI	2	0.62	0.54	0.78	3.43	1.62	3(2)	3(2)

Notes: *Trace test* reports the rank suggested by Johansen (1988)'s trace test with the corresponding p-value of acceptance. *Largest root* and *max* $|t|$ *in* $\boldsymbol{\alpha}_{r^*}$ report the respective values for the preferred rank and the one above, *Graph* indicates the rank most confidently suggested by the graph and the best alternative in parentheses, $r^*(r')$ the inferred preferred and second best choice of rank.
Source: Author's calculations.

ranks that are equal to, or lower, than the preferred rank that I eventually determine after consideration of criteria (ii)-(iv). The only exceptions to this occur for Lesotho PWT6, PWT7 and PWT8, as well as Madagascar in PWT6. This is not surprising in the light of the results discussed in the previous section, where the residuals in these models have been found likely to violate the assumption of normality, on which the test fundamentally relies.[12] I emphasise once more that this is pervasive throughout both the re-specified models, as well as those employed by JMT.[13]

The (ii) largest unrestricted roots are reported in columns 4 and 5 of table 3.7, for both the eventually preferred rank r^*, and for r^*+1. The idea here is that the largest unrestricted root in the system should be significantly smaller than one, in order to ensure a stationary process $\Delta\mathbf{X}_t$. Practically speaking, the largest unrestricted root at r^* should be small, while the one at r^*+1 should be close to one, as otherwise the rank could be confidently increased, preserving the information concerning another equilibrium relationship. It is, however, only indicative again, especially because the confidence intervals of the roots are unknown, and there is

[12]The table does not report a p-value in the cases where the trace test concludes full rank $r = p = 5$, as this inference emerges from the rejection of all lower ranks versus the alternative hypothesis of $r = p$.

[13]The relevant output underlying the original choices of rank, kindly provided by the authors, confirms this.

therefore no hard criterion in order to determine whether a root is significantly smaller than one (Juselius, 2006, p. 143). Nevertheless, the largest roots at r^*+1 are in the vast majority of models substantially larger than those at r^*, providing some justification for this choice. Most importantly in the context of a robustness check, they tend to be of a similar amplitude as those obtained in the PWT6 (JMT) models. For illustration, appendix B.4 plots the roots of the companion matrix for all ranks for Burkina Faso in PWT6.

The (iii) largest t-ratio in the α-vector associated with the r^*'th CI relation (column 5) gives an indication about the relevance of the last potential equilibrium relationship included in the model. The same figure for the r^*+1'th α-vector (column 6) provides such an indication for the first vector dismissed from the analysis. Juselius (2006) proposes a threshold of about $|t| > 2.6$, which is surpassed in most specifications, as r^*+1 is typically found to be susceptible of non-stationarity. Where this is not the case, r^*+1 is generally determined as the second best choice of rank, reported in parentheses in the last column.

In practice, the (iv) visual inspection of the graphs of the CI relationships turns out to be the most common tie-breaker. While in principle this is quite a subjective criterion, in most cases there tends to be a rather sharp difference in the appearance of the graphs of the r^*'th CI (the CIs being ordered by the corresponding eigenvalues) and the r^*+1'th (or, if this is the second best choice, the r^*+2'th). The highest included CI typically resembles a white noise process, while larger order CIs have a distinctly persistent, that is, non-stationary appearance. The values reported in the second last column of table 3.7 report the rank that can be best justified based on the graphs, followed by the best alternative rank choice resulting from them in parentheses; In cases where the difference is quite clear-cut, the alternative rank will be below the preferred choice. Where the next-highest CI also appears to be acceptable, e.g., has a short period of persistence but otherwise looks stationary, it may be reported as the suggested second best rank. Appendix B.5 plots the cointegrating relations for all ranks for Burkina Faso in PWT6.

The last column reports the final choice of ranks after weighting of criteria (i)-(iv).[14] It is apparent from the discussion in this section that the choice of the cointegration rank is everything but straightforward, and the researcher faces significant trade-offs in the process. This provides justification for JMT's procedure of assessing the results from two different economic angles, establishing trans-

[14]The choice of 3(2) for Togo PWT6, a model repeatedly found to be problematic in the previous sections, may deserve some discussion as it seems at odds with the indications provided by the criteria. The main rationale behind the choice of rank is to preserve consistency with JMT, even though their choice may appear rather surprising in this particular instance. As noted earlier, JMT's output is very much in line with the results I obtain; this can be verified for the trace test and the roots of the companion matrix.

parency by essentially picking the rank that yields the results most consistent with the respective prior of effectiveness or harmfulness, and reporting both.

Table 3.8 reports the final model choices, stating their respective lag-length, dummy variables, and first and second best choices of cointegration ranks.

Results of the re-specified models

The results emerging from these specifications are summarised in table 3.9, the underlying t-ratios are reported in appendix B.6. As in table 3.3, + and − stand for positive and negative coefficients, while the subscripts 0 and 00 denote coefficients with absolute t-ratios below 2.0 and 1.6 respectively. The table is organised in two columns for each dataset, reporting the results under the prior of effectiveness and the prior of harmfulness respectively, and four rows per country, one for each of the macrovariables under consideration. The first, second and third panel comprise the four countries where the results were most stable, relatively stable and least stable in the first replication exercise (see table 3.8).

The last two columns of table 3.9 count the number of coefficients that are consistent with the ones obtained from PWT6[15], where 'consistency' is still defined as in table 3.3. An apparent pattern is that the countries that yielded particularly consistent results under the original JMT models also remain substantially more consistent under the re-specified models, with 43 out of 48 consistent coefficients under the prior of aid effectiveness. This compares to 16 stable coefficients in both the countries with the least stable results previously, and the intermediate ones. Overall, 87 out of 144 possible coefficients (or 60%) of the coefficients remained stable. This pattern is repeated under the prior of harmfulness in a slightly less pronounced manner, and at a generally lower level of consistency. The datasets agree for 22 coefficients in the stable countries, 26 in the intermediate, and for 12 in the least stable ones. Overall, this sums up to only 60 out of 144 coefficients (42%) that remain stable under the prior of harmfulness.

One possible driver of the difference between the stability of the results is that the original results seem to be systematically different between stable and unstable countries: most of the coefficients of the stable countries are in fact insignificant. As noted in the previous replication exercise, insignificant coefficients had a stronger tendency to remain stable than significant ones. Note however that even Kenya, where 3 out of 4 coefficients are significant with PWT6, is almost

[15]Note that while these are overall consistent with the results obtained by JMT, but may differ in some cases; where this is the case, I use my own results as a benchmark for the sake of internal consistency. This affects Burkina Faso, the Republic of Congo and Madagascar, where differences in the aid data also lead me to a different specification than JMT.

Table 3.8: Final model specifications

Country	*Lags*	*Dummy variables*	r^*	r'
		Stable countries		
Burkina Faso	*Sample: 1965-2007*			
PWT6	1	$Dtr71_t, Dp00_t$	2	1
PWT7	1	$Dtr71_t, Ds74_t, Dp79_t, Dp85_t, Dp94_t, Dp00_t$	2	1
PWT8	1	$Dtr71_t, Ds74_t, Dp79_t, Dp85_t, Dp94_t, Dp00_t$	2	1
WDI	1	$Dp85_t, Dp94_t, Dp00_t$	2	3
Cameroon	*Sample: 1965-2007*			
PWT6	1	None	3	2
PWT7	1	None	3	2
PWT8	1	None	3	2
WDI	1	None	3	2
Gabon	*Sample: 1965-2002*			
PWT6	1	$Dtr7478_t, Dtr8900_t$	3	4
PWT7	1	$Dtr7478_t, Dp87_t, Dp00_t$	3	2
PWT8	1	$Dtr7478_t, Dp87_t, Dtr0002_t$	3	2
WDI	1	$Dtr7478_t, Dp87_t, Dtr0002_t$	3	2
Kenya	*Sample: 1965-2007*			
PWT6	1	$Ds68_t, Ds94_t$	3	2
PWT7	1	$Ds68_t, Ds94_t$	3	2
PWT8	1	$Ds68_t, Dtr9900_t, Ds94_t$	3	2
WDI	1	$Ds70_t, Ds94_t$	3	4
		Intermediate countries		
Congo, Rep.	*Sample: 1965-2007*			
PWT6	1	$Dp94_t, Dp05_t$	2	3
PWT7	1	$Ds96_t, Dp05_t$	3	2
PWT8	1	$Dp91_t, Ds96_t, Dp05_t$	3	2
WDI	1	$Dp77_t, Dp94_t, Ds96_t, Dp98_t, Dp00_t, Dp05_t$	4	3
Madagascar	*Sample: 1965-2007*			
PWT6	1	$Ds75_t, Ds82_t, Dp94_t$	4	3
PWT7	1	$Ds82_t, Dtr85_t, Ds94_t$	4	3
PWT8	1	$Ds82_t, Dp89_t, Ds94_t$	3	2
WDI	1	None	2	3
Rwanda	*Sample: 1960-2007*			
PWT6	1	$Ds94_t, Dp81_t$	3	2
PWT7	2	$Ds94_t, Dp63_t$	2	1
PWT8	2	$Ds94_t, Dp63_t$	2	1
WDI	1	$Ds94_t, Dtr6375_t.Dp97_t$	3	2
Senegal	*Sample: 1960-2008*			
PWT6	1	$Ds69_t$	3	2
PWT7	2	none	3	2
PWT8	1	$Ds69_t$	3	2
WDI	1	$Ds69_t$	2	3

Continued on next page

Table 3.8 (continued)

Country	*Lags*	*Dummy variables*	r^*	r'
		Unstable countries		
Benin	*Sample: 1965-2007*			
PWT6	1	$Dp75_t, Ds82_t, Dp94_t$	3	4
PWT7	1	$Dp82_t, Ds86_t, Dp94_t$	3	2
PWT8	1	$Ds82_t, Dp89_t, Dp94_t$	3	2
WDI	1	None	2	1
Lesotho	*Sample: 1963-2007*			
PWT6	1	$Dtr65_t, Ds70_t, Ds78_t, Dp83_t, Dp99_t$	3	2
PWT7	1	$Dtr65_t, Ds71_t, Dp78_t, Dp83_t, Dp99_t$	2	1
PWT8	1	$Ds71_t, Dp75_t, Ds78_t, Dp83_t, Dp99_t$	3	2
WDI	1	$Ds71_t, Dp75_t, Ds78_t, Dp83_t, Dp98_t, Dp99_t$	2	1
Mauritania	*Sample: 1965-2007*			
PWT6	1	$Ds92_t$	3	2
PWT7	1	$Dp86_t$	3	2
PWT8	1	$Dp86_t, Dp95_t$	2	3
WDI	1	$Dp76_t, Dp99_t, Dp03_t$	2	1
Togo	*Sample: 1965-2007*			
PWT6	1	$Dp93_t$	2	3
PWT7	1	$Dp86_t, Dp87_t, Dp93_t$	3	2
PWT8	1	$Dp86_t, Dp87_t, Dp93_t$	3	2
WDI	1	$Dp71_t, Dp75_t, Dp93_t$	3	2

Notes: The table summarises all models derived in the re-specification exercise. Dp, Dtr, Ds are permanent, transitory and shift dummies respectively, as defined in the main text. r^* and r' are the first and second best choices of rank respectively.

Table 3.9: Results of the re-specified models

		PWT6		PWT7		PWT8		WDI		Stable	
		Eff.	Harm.	Eff.	Harm.	Eff.	Harm.	Eff.	Harm.	Eff.	Harm.
Burkina Faso	y_t	$-_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$+_{00}$	3	2
	inv_t	$+_{00}$	$+_{00}$	$+_{00}$	$-$	$+_{00}$	$+_{00}$	$-_{00}$	$-$	2	1
	c_t	$-_{00}$	$-_{00}$	$-_{00}$	$+_{0}$	$-_{00}$	$-_{00}$	$+_{00}$	$+$	3	1
	g_t	$+_{00}$	$+_{00}$	$-_{00}$	$+_{0}$	$+_{00}$	$+_{00}$	$+_{0}$	$-$	2	1
Cameroon	y_t	$-_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$-$	3	2
	inv_t	$+_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$+_{00}$	$+_{00}$	3	3
	c_t	$-_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$-$	3	2
	g_t	$-_{00}$	$-$	$-_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	3	0
Gabon	y_t	$+_{00}$	$+_{00}$	$+_{00}$	$-_{00}$	$+_{0}$	$+_{0}$	$+_{00}$	$-_{00}$	2	2
	inv_t	$+_{00}$	$-_{00}$	$+_{00}$	$-_{00}$	$+_{0}$	$+_{0}$	$+_{00}$	$-_{00}$	2	2
	c_t	$+_{00}$	$-$	$-_{00}$	$-_{00}$	$+_{00}$	$+_{00}$	$-_{00}$	$-_{00}$	3	0
	g_t	$+_{00}$	$+_{00}$	$+_{00}$	$-_{00}$	$+_{00}$	$+_{00}$	$+_{00}$	$-_{00}$	3	3
Kenya	y_t	$+$	$+_{00}$	$+$	$+$	$+$	$+$	$+$	$+$	3	0
	inv_t	$+$	$+$	$+$	$+_{0}$	$+$	$+_{0}$	$+$	$+$	3	1
	c_t	$+$	$+_{0}$	$+$	$+$	$+$	$+$	$+$	$+$	3	0
	g_t	$+_{00}$	$+_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$+_{0}$	$-$	2	2
Congo, Rep.	y_t	$-_{00}$	$-_{00}$	$+_{00}$	$+_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	3	3
	inv_t	$+_{00}$	$+_{00}$	$+_{00}$	$+_{00}$	$+_{00}$	$+_{00}$	$-_{00}$	$-_{00}$	3	3
	c_t	$-_{00}$	$-_{00}$	$+_{00}$	$+_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	3	3
	g_t	$+_{00}$	$+_{00}$	$+_{00}$	$+_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	3	3
Mada-gascar	y_t	$+$	$+_{00}$	$-_{00}$	$+_{00}$	$+_{0}$	$+_{00}$	$-_{00}$	$-_{00}$	1	3
	inv_t	$+$	$-_{00}$	$-_{00}$	$-$	$+$	$-_{00}$	$+_{00}$	$+_{00}$	1	2
	c_t	$+$	$+$	$+_{00}$	$+$	$+$	$+$	$-_{00}$	$-_{00}$	1	2
	g_t	$-$	$-$	$-_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$+_{00}$	$+_{00}$	0	0
Rwanda	y_t	$+_{00}$	$-$	$+$	$+_{0}$	$+$	$+$	$+$	$+$	0	0
	inv_t	$+$	$-$	$+$	$+_{0}$	$+_{0}$	$+_{0}$	$+$	$+$	2	0
	c_t	$+$	$+_{00}$	$+_{00}$	$+$	$+$	$+$	$+$	$+$	2	0
	g_t	$-_{0}$	$-_{00}$	$+_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	$+_{00}$	$+_{00}$	0	3
Senegal	y_t	$+$	$+$	$+$	$+_{00}$	$+$	$+_{00}$	$+_{00}$	$+_{00}$	2	0
	inv_t	$+_{00}$	$+_{00}$	$+_{00}$	$+_{00}$	$+$	$+_{00}$	$+_{00}$	$+_{00}$	2	3
	c_t	$+$	$+$	$+$	$+_{0}$	$+$	$+_{0}$	$+$	$+$	3	1
	g_t	$+$	$+$	$+$	$+_{0}$	$+$	$+_{0}$	$+_{00}$	$+_{00}$	2	0
Benin	y_t	$+_{00}$	$-$	$+$	$-_{00}$	$+$	$-_{00}$	$-_{00}$	$-_{0}$	1	0
	inv_t	$+$	$-$	$+$	$-$	$+$	$+_{00}$	$+_{00}$	$-$	2	2
	c_t	$+$	$+$	$+$	$+$	$+$	$+$	$-_{00}$	$+_{00}$	2	2
	g_t	$-$	$-$	$+$	$-_{00}$	$-$	$-$	$+_{00}$	$+$	1	1
Lesotho	y_t	$+$	$+$	$-$	$-$	$+_{00}$	$-_{00}$	$+_{00}$	$-$	0	0
	inv_t	$+$	$+$	$+_{00}$	$+_{00}$	$+$	$+_{00}$	$+_{00}$	$-$	1	0
	c_t	$+$	$+$	$-$	$-$	$-$	$-$	$+_{00}$	$-$	0	0
	g_t	$+$	$+_{00}$	$-_{00}$	$-_{00}$	$+_{00}$	$-$	$-_{0}$	$-_{00}$	0	2
Mauri-tania	y_t	$-_{00}$	$+_{00}$	$+_{0}$	$-$	$-_{00}$	$-$	$-$	$-$	1	0
	inv_t	$+$	$-_{00}$	$+_{00}$	$-$	$+_{00}$	$-_{00}$	$-_{00}$	$-_{00}$	0	2
	c_t	$+_{00}$	$+_{00}$	$+$	$-_{00}$	$+$	$+$	$+_{00}$	$+_{00}$	1	2
	g_t	$+$	$+_{00}$	$+_{00}$	$+$	$+_{00}$	$+_{00}$	$+$	$+$	1	1

Table 3.9 (continued)

		PWT6		PWT7		PWT8		WDI		Stable	
		Eff.	Harm.	Eff.	Harm.	Eff.	Harm.	Eff.	Harm.	Eff.	Harm.
Togo	y_t	+	$+_0$	+	+	+	+	+	+	3	0
	inv_t	$-_{00}$	$+_0$	−	−	−	−	+	$+_{00}$	0	0
	c_t	+	$-_{00}$	+	+	+	+	+	+	3	0
	g_t	+	+	$+_{00}$	$+_{00}$	$+_{00}$	$+_{00}$	$+_{00}$	$+_{00}$	0	0
Σ	y_t	5	2	5	2	5	1	3	3		
	inv_t	6	2	3	5	5	1	3	3		
	Σ	8	2	5	6	7	2	3	5		

Notes: + indicates a positive coefficient, − a negative one. The subscript '0' indicates an absolute t-ratio between 1.6 and 2, '00' indicates one lower than 1.6. The absence of a subscript indicates an absolute t-ratio > 2. The last three rows count the number of cases of effectiveness and harmfulness on y_t, inv_t, and overall. The last two columns count the cases of consistency under each prior.

perfectly consistent under the prior of effectiveness, and far better than any of the inconsistent countries.

The overall *conclusions* suggested by these results are summarised in the last three rows of table 3.9 for each of the four datasets, across the 12 countries in the sub-sample. The first two of these rows report the number of positive (negative) significant coefficients for GDP (y_t) and investment (inv_t) respectively. The last row (Σ) counts the number of countries where at least one of the two is significant and positive (negative), and where therefore aid is considered to have been effective (harmful) in the long run, overall. In the sub-sample at hand, the overall conclusion changes twice once we account for the impact of the data on the modelling process, and remains constant only once, but with less strong support. The conclusion of aid effectiveness, clearly supported by PWT6 with 8 out of 12 countries providing evidence for it, compared to only 1 country (Benin) providing evidence for harmfulness, also finds support in PWT8, where 7 countries indicate effectiveness, and 2 harmfulness. PWT7 and WDI now lend some support to the hypothesis of harmfulness: In PWT7, there is evidence of effectiveness in 5 countries, and evidence of harmfulness in 6. In WDI, these figures are 3 and 5, respectively.

Adjusting the models to the new data did therefore not help to re-establish consistency where the results turned out to be shaky when applying JMTs original models to different datasets. Instead, it seems like the divergence of the results has been even exacerbated by letting the new data shape the statistical models.

3.6 Conclusions

In this chapter, I used the econometric framework provided by Juselius, Møller and Tarp (2014) in order to explore the stability of macroeconomic inference drawn from methods of time series analysis with respect to inconsistencies across datasets.

This included, in a first exercise, the application of the statistical models as specified by JMT to data from the Penn World Table versions 6.3, 7.1, 8.0, and the World Development Indicators 2015. About one third of the relevant coefficients changed qualitatively in each of the datasets applied, that is, in a way that affects inference (change of sign or significance level). Similarly, the country-wise conclusion regarding whether aid was effective, harmful or neutral, remained stable in about two thirds of the country-dataset combinations.

The second exercise allowed for a more fundamental impact of the data as I re-specified the country-specific models for the subset of 12 countries for which sufficient data from all considered sources were available. The results show that the modelling process can be influenced in a significant manner by the inconsistencies between the datasets, especially when it comes to the detection of outliers and the resulting dummy variables, and to the choice of the rank of cointegration. As a general pattern, the countries that proved to be particularly consistent in the first exercise also yielded more consistent models in the second exercise, and consequently more similar results. In the countries that were particularly inconsistent in the first exercise, the alterations to the models frequently restored significance in the results where it had been lost. These results did not necessarily correspond to the original results obtained by JMT, and overall, the re-specification exacerbated the divergence.

My results suggest that the choice of dataset can have substantial impact on the results obtained from time series analysis, and future research needs to address this. One obvious recommendation is that robustness checks with alternative data should be standard in the literature, wherever feasible. This is not always straightforward to do, especially in the context of time series analysis, as the specification may have to be fundamentally rethought when the data changes. At the same time, this only increases the potential ramifications of changes in the data and renders the exercise even more urgent. My findings do, however, suggest that it may be a useful and practical heuristic to confront one's final model specification to alternative datasets, as the results that remained robust in this exercise were also much more likely to persist after the entire model had been re-specified.

Beyond routine robustness checks, future research should also seek to establish whether alternative estimation techniques are more robust to data revisions and measurement error than the Cointegrated VAR framework employed in the study

at hand. Candidates could include likelihood analysis based on error distributions with broader tails or bootstrap based inference.

It is worth noting that for most of the countries included in the analysis (one half to two thirds, depending on the criterion), the results remain unaffected by any of the differences across datasets, even though those differences often appear substantial. In the light of the recent debate on 'Poor Numbers' (Jerven, 2013c), the 'Anarchy of Numbers' (Roodman, 2007), and 'Africa's Statistical Tragedy' (Devarajan, 2013), this result indeed seems more optimistic than the tone of the debate suggests. Categorically dismissing the only available evidence on the past economic performances of African countries, as advocated by some, would represent a considerable waste of information; it is what the best estimates tell us, and in most cases the reported sums seem to be sufficiently in agreement in order to provide valuable insights about the mechanisms at work.

Chapter 4

Exchange Rate and Inflation Dynamics in a Resource Rich Setting: the Case of Zambia[1]

4.1 Introduction

As economies become more open to the rest of the world, understanding the effects of global shocks has become an important issue in macroeconomic policy management (Berg et al., 2011; Kenourgios and Padhi, 2012; Essers, 2013), and the literature has devoted particular attention to commodity price shocks (Charnavoki and Dolado, 2014; Drechsel and Tenreyro, 2017). For countries in sub-Saharan Africa (SSA), an important concern is the macroeconomic effects of shocks to the commodity they export because many are highly dependent on a specific commodity (Venables, 2016). Often this is oil, which typically accounts for at least 90% of export earnings for Angola, Nigeria, the Republic of Congo, Guinea, Equatorial Guinea, and Sudan, but there are exceptions: gold accounts for 75% of Malis exports and copper typically accounts for more than 70% of Zambian export earnings.

Commodity price shocks can have direct effects on the real economy, as the commodity is typically a major production sector (e.g., Bergholt, 2014; Drechsel and Tenreyro, 2017), but as far as prices are concerned, the more immediate macroeconomic transmission is through the effect on the exchange rate. The strength of this channel is dominated by the exchange rate pass-though to prices (ERPT), that is, the extent to which changes in the exchange rate translate into lo-

[1]This essay is a result of joint work with Oliver Morrissey and Gregory Smith. An earlier version of it is published in the World Bank Policy Research Working Paper series (number 8128) and can be downloaded at `http://documents.worldbank.org/curated/en/393871498656454681/Exchange-rate-and-inflation-dynamics-in-Zambia`.

cal price levels. The ERPT tends to be higher in low- and middle-income countries than it is in advanced economies (Ca'Zorzi et al., 2007; Aron et al., 2014), leaving resource-rich low-income countries like those in SSA particularly susceptible to price instability induced by global fluctuations in commodity prices.

Macroeconomic management is difficult under these conditions, and for many of the challenges faced, the recommendations emerging from research do not offer clear guidance. This starts with the very fundamental challenge of choosing the right exchange rate regime. International institutions have long been advocating floating exchange rates and inflation targeting as key elements of sound monetary policy. Especially for emerging and developing economies, the universal appropriateness of this recommendation has been prominently challenged by Calvo and Mishkin (2003), and high ERPT is considered one of the main obstacles (Eichengreen, 2006). At the same time, much is at stake, as recent research shows that the implications of the choice of exchange rate regime for macroeconomic dynamics prove to be substantial (Morón and Winkelried, 2005; Hoffmann, 2007; Obstfeld et al., 2017; Siklos, 2018). Irrespective of whether monetary policy targets inflation, the exchange rate, or if it adopts a hybrid policy, a choice needs to be made as to whether or not the central bank should actively anticipate the effects of commodity price fluctuations on the economy, for instance when setting the interest rate. While this seems desirable at first, any such intervention requires a deep understanding of the transmission of shocks in the specific context; besides, some research points at the disastrous potential of such policies altogether (Bergholt, 2014). To complicate things even more, policy makers face non-trivial trade-offs between long-term and short-term goals. For instance, there is growing evidence that the exchange rate pass-through itself is affected by the inflation environment, with low inflation being associated with a low pass-through (Choudhri and Hakura, 2006; Devereux and Yetman, 2010; Lopez-Villavicencio et al., 2016; Jasova et al., 2016). Avoiding inflation targeting on the grounds of a high ERPT in the short run may then be an obstacle to reducing that same ERPT in the long run. Similarly, the degree to which commodity price shocks affect the exchange rate in the first place may be endogenous to institutional arrangements (Bodart et al., 2015).

The absence of simple *one size fits all* policy recommendations creates a need for in-depth analysis of the macroeconomic relationships within these countries, that is, a need for case studies. The aim of this study is then to shed light on the interrelationship of export commodity (copper) prices, the exchange rate, and consumer price inflation in the case of Zambia. Zambia is a suitable country to examine in this context for two reasons. First, a recent tumultuous episode in the country has once more illustrated the importance and complexity of the nexus

between exchange rates, prices, and commodity prices, and sparked the demand for a thorough analysis of the mechanics at work: In late 2015, the country's currency – the kwacha – experienced a dramatic devaluation of the exchange rate of 60% in just ten weeks, and a rise in consumer prices by about 20%. Prior to this, the decline in copper prices that had begun in 2012 had once more accelerated. Second, Zambia shares key characteristics in this respect with many SSA countries, and we believe that – with the usual caveats of external validity – conclusions drawn from this case study can be informative beyond the narrow country-specific context. Importantly, the country has high dependence on one export commodity, copper, the price of which has experienced large fluctuations over the past decades. And, in 1994, the country was one of the first of the thirteen SSA countries that had adopted floating exchange rate regimes[2] by 2016 (IMF, 2016), with more countries intending to follow.

In concrete terms, we seek to establish:

1. The relative importance of factors affecting the value of the kwacha.

2. The extent to which changes in the value of the kwacha are reflected in changes in consumer prices (inflation), that is, to measure the ERPT.

3. The extent to which the ERPT (2) depends on the shock that initially caused the exchange rate (1) to move. That is, we seek to disaggregate the ERPT according to the shock that initially hit the economy.

Regarding (1), the existing country-specific literature almost unambiguously finds that the price of copper is a key driver of the exchange rate of the kwacha (Bova, 2009; Muhanga et al., 2014; Chipili, 2016). Similarly, Cashin et al. (2004) test for a long-run relationship between main export commodities and the exchange rate in a sample of 58 countries, and conclude that the Zambian kwacha is indeed a 'commodity currency'. Our results corroborate this finding, and we estimate that a 10% rise in the price of copper leads to an appreciation of the kwacha of about 4% in the long run.

Regarding (2), the evidence is more scarce. To the best of our knowledge, the only country-specific study explicitly aimed at quantifying the ERPT in Zambia is Zgambo (2015), who estimates it to be relatively high (between 41% and 49%). In their seminal study concerned with estimating the determinants of ERPT in a global sample, Choudhri and Hakura (2006) report a similar value for Zambia (41% after 4 quarters). Our estimates are substantially lower, ranging from 7% to

[2] *De jure*; as in most of these countries, the *de facto* classification has changed over time, see section 4.2.

about 25%, depending on the shock that caused the exchange rate to move in the first place.

This leads us to (3), a question that has only recently been raised in the literature (Shambaugh, 2008; Forbes et al., 2015) and is, to the best of our knowledge, unexplored not only for Zambia but for small commodity exporters altogether. The argument here is that the ERPT need not be the same for all the shocks that drive the exchange rate. For instance, a depreciation caused by a drop in commodity prices may translate into local consumer prices to a lesser extent than the same shift in the exchange rate that has been triggered by expansionary monetary policy by the domestic central bank: while the commodity price shock simultaneously hits one of the key sectors of the economy and may therefore lead to a downturn in aggregate demand (counteracting inflationary pressure), the expansionary monetary policy operation simultaneously increases money supply (further accelerating inflation). Forbes et al. (2015) have recently underlined the importance of disaggregating the ERPT for the United Kingdom, and in order to analyse the question for Zambia, we borrow heavily from their empirical framework. Our results, in turn, underline the importance of this endeavour; in particular, we find that genuine (exogenous) exchange rate shocks, and those emanating from the copper price, have a moderate ERPT (around 7%), while those caused by money supply translate into price much more sharply (around 25%).

In our empirical analysis, we identify six shocks to key variables of the Zambian economy (namely shocks to the copper price, external interest rate, output, domestic money supply, exchange rate and inflation) using a combination of sign- and zero-restrictions in the framework of a structural vector autoregressive (SVAR) model, estimated on quarterly data ranging from 1995 to 2014. Since Sims (1980), Blanchard and Watson (1986) and Bernanke (1986), the SVAR method has had numerous applications in empirical macroeconomics. The most common identification strategy is the triangular Cholesky decomposition, which reduces the number of free parameters exactly to the point where the system is just-identified. This approach often implies assumptions that are hard to justify, and the results may be susceptible to the ordering of the variables. Uhlig (2005) suggests the remedy of using economically plausible but less restrictive sign restrictions on the impulse response functions, which can be estimated in a Bayesian VAR. However, Baumeister and Hamilton (2015) show that statistical models that appear to be 'agnostic' at first can in fact be heavily influenced by the priors imposed by the researcher, even if they were originally designed to be uninformative, that is, designed to have little impact on the results. Indeed, they show that there can be situations where results may be driven solely by the priors unintentionally imposed by the researcher. Baumeister and Hamilton (2015) then recommend

imposing informative priors on economically interpretable parameters, where the prior values can, for instance, be derived from earlier empirical literature or from theory. We acknowledge their argument but note that, especially in larger VARs and for less studied economies and questions, there may simply not exist conceptually similar estimates for the relevant parameters, and that the parameters of theoretical models can rarely be directly mapped to the empirical model. In order to obtain sensible priors nonetheless, we first estimate short-run elasticites using a conventional Cholesky decomposition, and then feed these into our Bayesian VAR as priors.

Specifically, adopting this more flexible framework allows us to lift a particularly controversial restriction concerning the interaction between inflation and the exchange rate, our key variables of interest. Earlier literature, relying on strictly identified systems of equations, invariably had to assume that either the exchange rate would only react to changes in inflation with a delay, or the opposite, that inflation would only react to exchange rates with a lag. However, either of these assumptions is hard to defend in practice, and especially where the focus is specifically on the interaction between these two variables, imposing such strong *a priori* assumptions is highly undesirable.

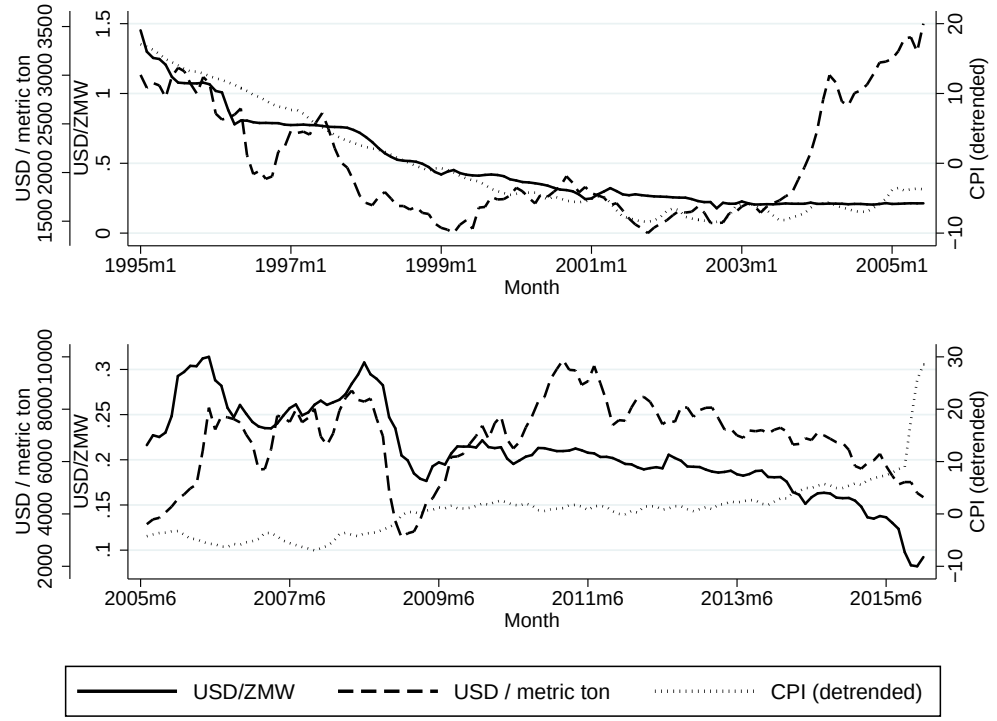
The remainder of this paper proceeds as follows: Section 4.2 discusses relevant aspects of the institutional and historical background and related literature for Zambia. Section 4.3 presents the methodology and underlying assumptions. Section 4.4 describes the data, the basic time series properties, and transformations implemented. Section 4.5 presents the results, that is, the relevant impulse response functions, historical variance decompositions, and estimates of differential ERPTs. Section 4.6 discusses the turbulence of late 2015, and draws on qualitative evidence in order to explain what led the dynamics discerned earlier in this study to break down. Section 4.7 concludes.

4.2 Background

Zambia is a small, open commodity-dependent economy and regularly faces challenges from supply shocks. These include shifts in the global copper price, rain-fed agricultural outputs, hydro-electric generation output, and the global price of fuel. It has been a copper producer since prior to its independence in 1964, but from the early 1970s production fell considerably. Nationalisation in 1973 and low prices hampered investment in the sector until the 2000s. Fresh investment followed privatisation and higher copper prices in the mid-2000s supported a doubling of copper production between 2004 and 2014. The boost to the economy from the copper industry was also complemented by better macroeconomic fundamentals in

the 2000s. For example, Zambia maintained a fairly low level of inflation, judged by its historical standards, and substantial debt relief improved investors' perception of the country after it qualified for the Heavily Indebted Poor Countries (HIPC) initiative in 2005. These factors supported the expansion of the economy and annual GDP growth, which averaged 7.4% between 2004 and 2014. As the production of copper increased and the economy prospered, much larger foreign exchange inflows followed from the mining sector. Added to this in the late 2000s were inflows from large Chinese infrastructure lending and from 2012 in the form of large commercial Eurobonds placements (2012, 2014 and 2015).

Figure 4.1: Key variables from 1995 to 2015 (monthly)



The Bank of Zambia states as the primary objective of its monetary policy to maintain price stability and promote financial system stability. Since April 2012, it uses a target policy interest rate as its main instrument, which is adjusted quarterly in regular times; prior to 2012, monetary policy focussed on monetary aggregates. In 1994, Zambia was one of the first countries in SSA to introduce a floating exchange rate regime *de jure*. It has since adhered to this commitment to varying extents, but compared to other low-income countries, there have been relatively few and limited interventions to stabilise the exchange rate. The most comprehensive and recent survey of *de facto* exchange rate regimes, Ilzetzki et al. (2017), however, makes an important distinction. Compared to, e.g., the IMF's annual report on exchange rate arrangements (IMF, 2016), the authors introduce a

category of 'freely falling' currency regimes to designate episodes of rampant inflation along with dramatic exchange rate devaluations; other classifications subsume these episodes under the category of 'freely floating'. Between 1994 and 2001, Ilzetzki et al. (2017) classify the Zambian exchange rate regime as a combination of 'freely falling' and 'freely floating', indicating that while the exchange rate has indeed been moving relatively freely, this has sometimes been accompanied by substantial instability. Between 2001 and October 2015, the authors classify Zambia's exchange rate regime as 'de facto band' or 'managed floating', indicating that the country had effectively shifted to a soft peg (to the US Dollar). Following the currency crash in late 2015, the regime is again classified as 'freely falling', and the authors note the presence of a parallel currency market in the aftermath of the crisis. Generally, an important property of Zambia's foreign exchange markets is that they are 'thin', in the sense that there is not much trading, and that they are dominated by few players (notably the large mining firms, the commercial banks, and the Bank of Zambia). There are often times of low trading volume, when large players are not trading and the exchange rate is fairly stable. The participants that are large relative to the level of transactions can exert influence over the exchange rate, sometimes by the timing of their transactions, even if they do not intend to.

In 2014 and 2015 the economy experienced several shocks including a further decline in copper prices and after nearly a decade of running a trade surplus, the external account fell into deficit in December 2014. The initial consequence was a steady depreciation of the kwacha during the first half of 2015. In August of that same year the confidence of the economy was eroded further as a power crisis began to impact all sectors of the economy, and the trade deficit accelerated. Between August and November, the currency depreciated by 60% in what appears to have been an instance of market panic. Some gains were recovered in 2016 following much tighter monetary policy, but at the end of 2016 the kwacha remained substantially weaker than prior to the shock. The depreciation of the currency abruptly fed through to consumer prices between September and November 2015, lifting annual inflation from less than 8% in the preceding years to more than 20% in 2015. Section 4.6 will contrast this episode to the 'usual' dynamics described in this study, and discuss some avenues for future research on the topic.

Figure 1 traces the key macroeconomic variables relevant to our narrative (the exchange rate, the copper price and consumer prices) since 1995, after the introduction of the floating exchange rate regime in 1994 and at the start of our sample period (see section 4.4). Before proceeding to the econometric analysis, it is worth relating some of the key changes in the data to the institutional context outlined above. First, notice the changing patterns in the development of the exchange

rate (expressed throughout this article in USD/ZMW, such that a decrease corresponds to a depreciation). In the earlier periods (top panel) it seems to be moving rather independently of the copper price (expressed in USD per metric ton). It is only after 2005, when investment and copper production increased, that the two variables began to noticeably move together.

Second, this coupling coincides with a strong appreciation following scaled-up debt relief under the HIPC initiative in 2005. To be sure, this appreciation precedes the increase in the copper price around the same time; later it appears the sequence is reversed (changes in the copper price precede those in the exchange rate), consistent with the interpretation that copper became a stronger determinant of the exchange rate after the sector had attracted investment and gained importance. In general, the relationship between the value of the kwacha and the price of copper is now a relatively well established finding. It is supported by case studies such as Muhanga et al. (2014), who focus on the short run in an SVAR, or Bova (2009) and Chipili (2016), who both focus on long-run cointegrating relationships and infer an equilibrium relationship between the price of copper and the exchange rate. Similarly, Cashin et al. (2004) test for a long-run relationship between main export commodities and the exchange rate in a sample of 58 countries, and conclude that the Zambian kwacha is indeed a 'commodity currency'. To the best of our knowledge, the only study that does not support this view is Pamu (2011), who cannot discern a relationship between the price of copper and the exchange rate.

Lastly, while consumer prices (CPI) and the exchange rate do generally appear to be trending in the same direction in most periods, prices do not react to exchange rate fluctuations in any obvious manner. This is surprising, as *a priori*, one would expect a small country with historically high inflation like Zambia to have a large ERPT (Choudhri and Hakura, 2006; Devereux and Yetman, 2010; Lopez-Villavicencio et al., 2016; Jasova et al., 2016), hence a pronounced co-movement of the two variables. Additionally, the few existing econometric estimates of the ERPT in Zambia would suggest this: both Zgambo (2015) in a case study, and Choudhri and Hakura (2006) in their seminal cross-country study, find Zambia's ERPT to exceed 40%. Interestingly, the observation that prices and the exchange rate appear to be largely discoupled does not hold in late 2015. In the period of crisis described above, the sharp depreciation of the currency comes with an equally salient increase in inflation. This observation will be further discussed in Section 4.6.

4.3 Methodology

Subsection 4.3.1 outlines the general form of the structural VAR employed and discusses the central problem of identifying the relevant shocks. Drawing from recent methodological advancements, especially Baumeister and Hamilton (2015) and building on Uhlig (2005), we avoid enforcing a fully identified structure on our system, and instead rely on a combination of sign- and zero-restrictions to achieve set identification in an under-identified system. This allows us to relax a common, but in our view overly strong (and as our results will show, influential), assumption about the contemporaneous interaction of the exchange rate and inflation. Unlike earlier studies, our specification will allow these variables to contemporaneously affect each other in either direction. As our system is under-identified, estimation will be carried out in a Bayesian framework, which raises the critical issue of selecting sensible priors. As Baumeister and Hamilton (2015) point out, these can be highly influential even where they are designed for opposite (supposedly flat priors). Priors therefore need to be chosen with corresponding care, and reported in a transparent manner. Subsection 4.3.2 will discuss and make explicit our choices in this respect.

4.3.1 Identification

We estimate the macroeconomic dynamics of interest in the framework of a vector autoregression of the following form (using the notation of Ramey (2016) and Stock and Watson (2016)):

$$A(L)Y_t = \eta_t, \tag{4.1}$$

where η_t is a vector of reduced form VAR innovations, that is, the observed residuals as obtained when forecasting Y_t based on its lagged values. We make the usual assumptions that these innovations have a zero mean, constant variance, and are uncorrelated with each other. Furthermore, $A(L)$ is a lag polynomial of the form $A(L) = I - \sum_{k=1}^{p} A_k L^k$, where L^k is the lag operator, shifting values by k lags. It effectively applies weights to each lag and element of Y_t in order to explain its current values, short of the innovations η_t.

However, the idea of a structural VAR is specifically to represent Y_t in terms of structural shocks ε that are unexpected, exogenous disturbances to the system, in order to identify the effect of unexpected changes in each of the variables, and to subsequently compute parameters of interest such as impulse response functions and variance decompositions. The SVAR assumes that these structural shocks ε are related to the VAR innovations η in the following way:

$$\eta_t = H\varepsilon_t, \tag{4.2}$$

Where the matrix H describes the contemporaneous relationships between the structural shocks ε_t and the reduced form VAR residuals η_t, or how an exogenous shock to any variable will affect each of the variables in Y *within* the period. We will normalize it such that the structural shocks have unit effect, that is, its diagonal elements are unity (an exogenous shock to any given variable will translate one to one into a change of that same variable). Without any further restrictions however, the system is *unidentified*, and there exists an infinite number of parameter combinations in H that are equally consistent with the observed data. The challenge is then to impose some structure that will allow us to uncover the parameters of H, and therefore the impact of each of the individual shocks onto the system.

This leads us to the critical issue of identification, which sets our study apart from most of the previous literature. Much of this literature relies on a Cholesky decomposition first suggested by Sims (1980) and still common in the exchange rate and monetary literature (with respect to Zambia, recent examples include Muhanga et al. (2014) and Zgambo (2015)). For all its benefits (most importantly, pinpointing parameters by just-identifying the system), this comes at the cost of substantial rigidity and requires the researcher to impose strong assumptions on the economic mechanisms that are being investigated. Our approach will therefore build on more recent developments in the literature, which has increasingly moved away from strict just-identified systems towards more flexible specifications; specifically, we will rely on a combination of sign- and zero-restrictions in a fashion similar to Uhlig (2005). The general idea here is to restrict parameter combinations to the set of those that are consistent with fundamental macroeconomic principles, instead of imposing so many (but potentially unwarranted) restrictions that only one combination remains possible. We also acknowledge recent improvements on the Uhlig (2005) approach, and in particular we use Baumeister and Hamilton (2015)'s methodology in order to avoid the undue impact of supposedly uninformative priors (see section 4.3.2).

Consider now the following general set of equations that describe the relationship between the variables in our system, which are chosen in line with the literature on exchange rates and inflation. The model is limited to six variables: the *US Federal Funds rate* (to represent the external interest rate), the *international price of copper*, domestic *money supply*, the domestic *output gap*, the *exchange rate* and *consumer price inflation* (the choice and details of the variables will be discussed in more detail in section 4.4, and the assumptions underlying this structure later

in this section):

$$\begin{aligned}
Fed_t &= \varepsilon_{F,t} \\
Copper_t &= \phi_C Fed_t + \varepsilon_{C,t} \\
Money_t &= \phi_M Fed_t + \kappa_M Copper_t + \varepsilon_{M,t} \\
Output_t &= \phi_Y Fed_t + \kappa_Y Copper_t + \mu_Y Money_t + \varepsilon_{Y,t} \\
ER_t &= \phi_E Fed_t + \kappa_E Copper_t + \mu_E Money_t + \nu_E Output_t + \pi_E Infl_t + \varepsilon_{E,t} \\
Infl_t &= \phi_P Fed_t + \kappa_P Copper_t + \mu_P Money_t + \nu_P Output_t + \xi_P ER_t + \varepsilon_{P,t}.
\end{aligned} \tag{4.3}$$

More concisely, we can represent the dynamic behaviour of this system in matrix notation as:

$$Y_t = \Phi(L) Y_t + \Omega \varepsilon_t \tag{4.4}$$

where $\Phi(L) = \Phi_0 + \sum_{k=1}^{p} \Phi_k L^k$: The structural coefficients in system (4.3) are captured in the contemporaneous matrix Φ_0, while the dynamics of additional lags are captured in the matrices Φ_k. The ε's are now the structural shocks we will aim to identify, entering the system as determined by matrix Ω. We follow common practice in assuming that each structural shock only enters one equation, or $\Omega = I$.

As outlined above, our identification strategy will rely on imposing structure on H in order to identify the contemporaneous relationship defined in equation 4.1. The combination of sign- and zero-restrictions we employ takes the form:

$$H = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
\phi_F & 1 & 0 & 0 & 0 & 0 \\
\phi_M & \kappa_M & 1 & 0 & 0 & 0 \\
\phi_Y^+ & \kappa_Y & \mu_Y & 1 & 0 & 0 \\
\phi_E & \kappa_E^- & \mu_E^- & \nu_E & 1 & \pi_E^- \\
\phi_P & \kappa_P & \mu_P^+ & \nu_P & \xi_P^- & 1
\end{bmatrix} \tag{4.5}$$

This specification is similar to a Cholesky decomposition, but it relaxes the restrictions on the contemporaneous interactions on our main variables of interest, inflation and the exchange rate. Instead, we impose a number of sign restrictions on the parameters, indicated by the superscripts '$-$' and '$+$'. More specifically, this specification reflects three types of assumptions:

Small economy: The domestic macroeconomic variables in Zambia are assumed to have no contemporaneous impact on the two global variables in the system, the

US Federal Funds rate and copper prices. Given the importance of copper as an export in Zambia, the latter may seem restrictive; however, Zambia only accounts for about 2% of global copper production.

Delayed responses: Monetary policy is assumed to react only with a delay to developments in the economy. This assumption is commonly made and rooted in the observation that information on macroeconomic variables typically becomes available only with a lag. In Zambia, the Monetary Policy Committee meets every quarter to review monetary policy, which is also the frequency of our data. Similarly, output is typically observed to react to changes in other variables only with a lag. The ordering above implies that output may respond to monetary policy contemporaneously, but not vice versa. Note that a few studies have made the opposite assumption in the past and constrained GDP to only react with a lag based on the argument that it is a notoriously slow variable; in a robustness check, we changed the order and the results were robust to this choice.

Sign restrictions: We impose sign restrictions on contemporaneous parameters where this choice is consistent with a broad range of macroeconomic theories. Specifically, the exchange rate is constrained to depreciate (if anything) as the fed funds rate increases, as broad money increases, and as domestic inflation increases. Similarly, we constrain the contemporaneous impact of an exchange rate appreciation on prices to be negative or zero, thus excluding parameter combinations that would imply a negative ERPT at impact. Further, the contemporaneous reaction of output to an increase in copper prices is assumed to be positive or zero, as well as that of inflation to an expansionary monetary shock.

4.3.2 Priors

As our system is not fully identified as in the pure Cholesky case, we follow the literature in taking a Bayesian approach in order to estimate the VAR. Crucially, this requires us to select priors. As shown by Baumeister and Hamilton (2015), seemingly uninformative priors (e.g., the commonly used Haar prior) can effectively be very influential, in a manner that is neither intended, understood, or related in any intuitive way to the underlying economic meaning. This can be the case when the outcome of interest (e.g., the impulse response function) is a non-linear transformation of the parameters the priors are imposed on. Instead, they suggest to formulate informative priors directly on parameters with an intuitive economic interpretation, in our case on the contemporaneous coefficients in matrix H. We will follow their approach, and use their numerical algorithm to

obtain our results.

This poses the fundamental challenge of finding sensible prior values, that is, expectations about the values of the contemporaneous elasticities between the variables included in our system. In fact, we found it impossible in the present study to base all our priors on results of the earlier literature, which would be the obvious way to proceed. This is the case for a number of reasons: First, *reporting conventions* make it difficult to draw from earlier studies. It is not common for the empirical literature to report the contemporaneous coefficients making up the matrix H. For instance, Muhanga et al. (2014), even though they study the same country with a similar research question and a comparable set of variables, only report their results in the form of impulse response functions, and thus do not offer a starting point for our priors.

Second, the *size of our VAR* complicates the search for priors. With an increasing number of parameters (we estimate 16 contemporaneous parameters), it becomes increasingly difficult to find sensible priors for each of them that would be conceptually similar and could reasonably be expected to apply to Zambia.

Third, there is *little guidance by theory* in the choice of priors. The theoretical literature is generally not parametrised in a way that suggests any particular amplitudes for the coefficients in our empirical model. To illustrate this, consider Baldini et al. (2012), who develop a DSGE model to analyse monetary policy in Zambia. While conceptually and contextually close to our study, and very transparent about the parameters applied and derived, none of the macro-parameters of interest here correspond to those employed in their model, which by construction roots the dynamics in deep parameters at the micro level.

In order to obtain sensible priors, we take the following approach. Instead of collecting or deriving coefficients from earlier literature, we generate our own priors from a more traditional approach (*as if* there was previous literature sufficiently consistent with our study). We start by estimating the model using a conventional Cholesky decomposition, thereby obtaining results based on more restrictive assumptions as typically made in the earlier literature. We then feed the resulting estimates as the prior modes into the system described earlier. Thus, we take traditional estimates with all their shortcomings as a first guess, and then use state-of-the-art algorithms to improve on them.

Note that – given our structure of choice described above – it is not possible to estimate priors for π_E and ξ_P simultaneously. In order to obtain prior modes for both, we estimate the system twice, once as described above, and then reversing the order of the last two variables (exchange rate and inflation). All other parameters are those obtained from the estimation with the order as described throughout the article (exchange rate before inflation). Beyond the mode, in or-

der to complete the specification of our priors, we need to specify the nature of our prior distribution (as the parameters are inherently treated as random variables). In line with Baumeister and Hamilton (2015), we formulate the priors on the parameters in H as student t distributions with three degrees of freedom, and a scale parameter of 0.6, reflecting moderate confidence in our prior modes. Appendix A explicitly reports the prior modes. The posteriors are then obtained from running a Metropolis-Hastings algorithm with 2×10^6 repetitions, 5×10^5 of which are stored (that is, 1.5×10^6 iterations serve as a burn-in period).[3] Appendix C.1 reports the prior modes, appendix C.2 reports the posterior modes.

4.4 Data

We use quarterly data from 1995:2 to 2014:4. This corresponds to a period of relative stability in the Zambian economy, and a regime of flexible exchange rates since 1994. Furthermore, we only have data on the output gap starting from 1995, the first quarter of which drops out as we take first differences. We deliberately exclude the tumultuous year 2015 from the main analysis as it exhibits extreme developments that are likely to dominate and distort our baseline estimations. Instead, section 4.6 will quantify the discrepancy between the 'usual' dynamics and the episode in 2015 and discuss explanations, pointing at possible avenues for future quantitative research on the topic.

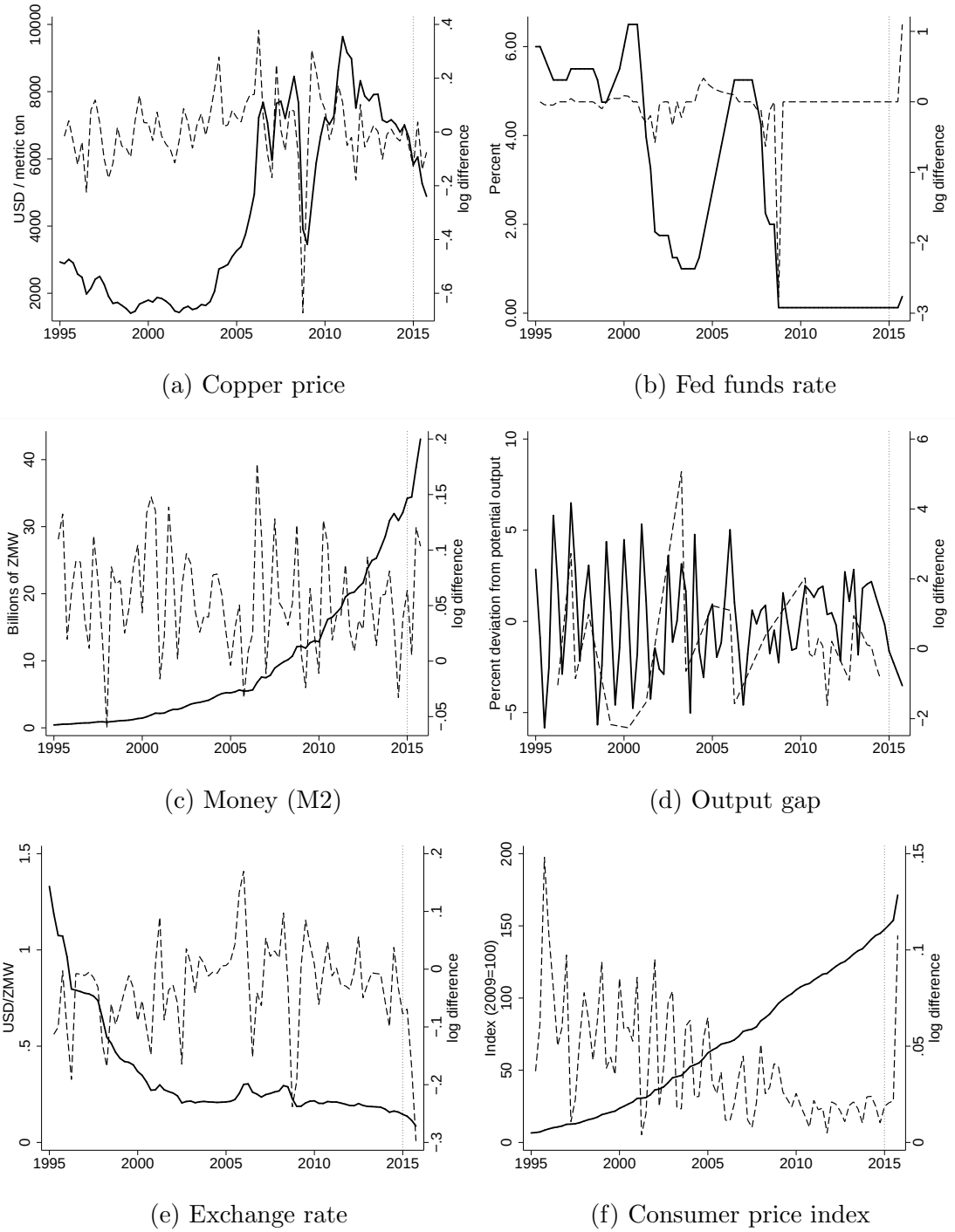
We obtain the price of copper from the IMF's Primary Commodity Price database; the Federal funds rate, the nominal USD/ZMW exchange rate (expressed such that a decrease corresponds to devaluation) and the Zambian CPI from the IMF's International Financial Database; Broad Money (M2) from the Bank of Zambia's fortnightly statistics[4]; and an internal estimate of the output gap kindly provided to us by the Bank of Zambia. All series are originally obtained at a monthly frequency and averaged over quarters, except for the output gap which was provided to us at quarterly frequency.

Figure 4.2 plots all variables in levels (solid lines) and in log differences (dashed lines). All series, except for the output gap (especially in early periods) resemble non-stationary series; the differenced series however appear stationary. This is confirmed by augmented Dickey-Fuller tests reported in appendix C.3. As our study focusses on relatively short time horizons, we will restrict our attention to the differenced series.

[3] Baumeister and Hamilton (2015) keep 1×10^6 repetitions for their posterior distribution; for a lack of computational capacity (working memory), we constrained ourselves to half that number.

[4] For most of our sample period (until April 2012), the Bank of Zambia used a money supply target, which is why we include M2 rather than the interest rate.

Figure 4.2: Variables in levels and log differences



(a) Copper price

(b) Fed funds rate

(c) Money (M2)

(d) Output gap

(e) Exchange rate

(f) Consumer price index

Notes: The solid lines describe the variables in levels and native units (see left y-axis), the dashed lines indicate the first differences of the logarithm of these series. The dotted vertical line at the beginning of 2015 indicates the end of the sample period used for estimation. See main text for further transformations applied to the series.

Specifically, we use the log-difference (multiplied by 100) of the copper price, money, the exchange rate and the consumer price index, so as to measure percentage changes between quarters. The Federal Funds rate, already expressed in percentages, will be included in first differences, therefore reflecting changes in percentage points. The output gap is expressed in percentage deviations from potential output and will be included as obtained. Furthermore, we de-trend the consumer price and money series in the usual manner, that is, by regressing a trend term on each of them and retaining the residuals from that equation.

4.5 Results

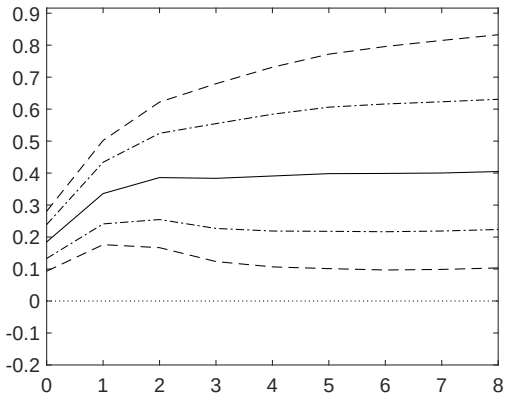
We discuss our results in three steps. First, we present the basic impulse response functions of our main variables of interest, the exchange rate and inflation. These indicate a strong dependence of the exchange rate on copper, as well as a moderate effect of the exchange rate on price levels (low ERPT). Second, we report variance decompositions over the sample periods in order to isolate the most important contributing factors on each of the variables, and the evolution of this pattern over time. Third, we analyse whether there is a differential ERPT depending on the shock that drives the exchange rate fluctuation, following Forbes et al. (2015). We further investigate whether and to what extent food and non-food inflation are affected differently, and also report differential ERPTs obtained using the respective disaggregated CPIs.

4.5.1 Impulse response functions

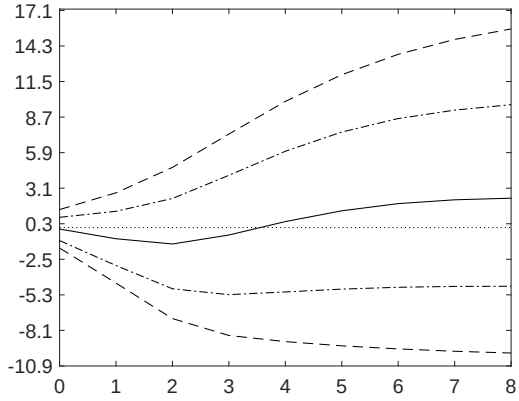
Figure 4.3 depicts the cumulative impulse response functions (IRF) for the exchange rate as a reaction to a 1% positive shock in the other variables. The solid line reports the pointwise mean IRF, which is the recommended point estimate to be reported given a linear loss function (Baumeister and Hamilton, 2015). The dashed lines depict the 68^{th} and 90^{th} percentiles of the estimated IRFs (following the Bayesian convention of *credible* intervals, in lieu of frequentist confidence intervals).

Panel (a) summarizes one of the key results of this exercise: the exchange rate reacts strongly and quickly to an increase in copper prices. The mean response suggests that, at impact (within the quarter), an unanticipated 1% rise in copper prices leads to an exchange rate appreciation of 0.2%. The exchange rate continues to markedly appreciate for the next two quarters, up until 0.4%, and then stays at roughly that level. This is slightly lower than the 4.8% appreciation following a 10% increase in copper prices in the long run reported by Chipili (2016), although

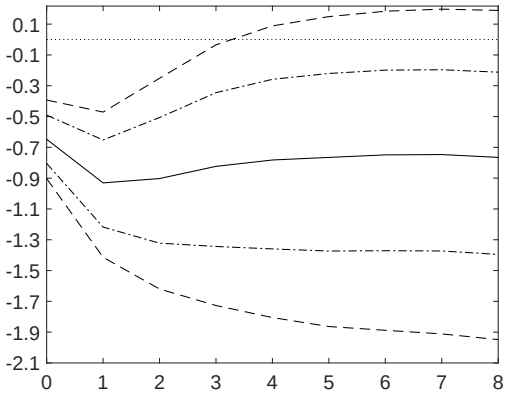
Figure 4.3: Responses of the exchange rate to a 1% shock to key variables



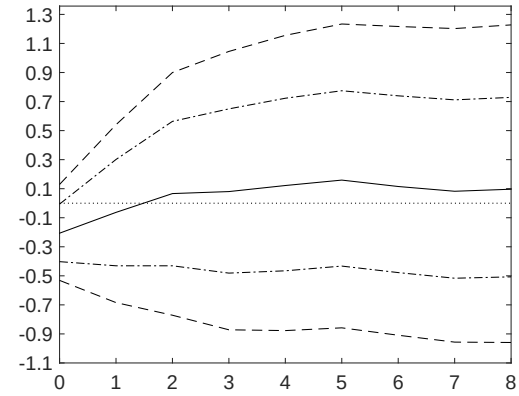
(a) Copper price



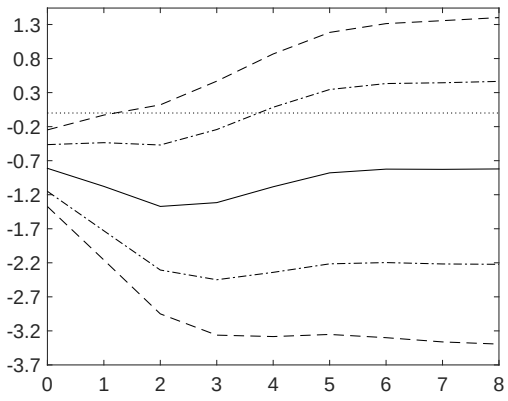
(b) Fed funds rate



(c) Money (expansionary)



(d) Output gap



(e) CPI

Notes: The solid line reports the pointwise mean IRF, the dashed lines depict the 68^{th} and 90^{th} percentiles of the estimated IRFs (credible intervals).

this estimate is within the bounds of the 68% interval.

An expansionary monetary shock of 1% (panel (c)) is estimated to induce a marked depreciation of about 0.7% within the quarter, reaching its peak in the next quarter at 0.9%. The point estimates suggest a slight overshooting here, and the persistent effect is of about 0.8%. Note that after three quarters, the estimates become less confident to the point where a zero effect cannot be excluded based on the 90% credible interval.

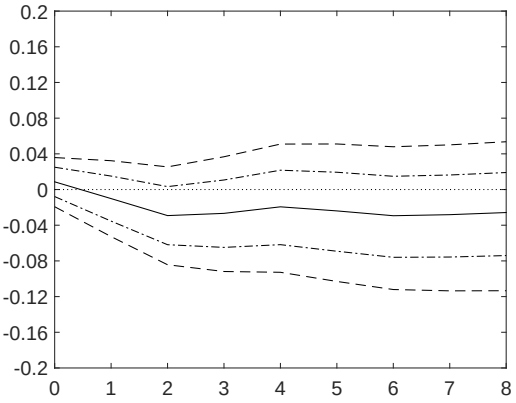
The third most salient driver is domestic prices, reported in panel (e). Although the estimates are too imprecise for any stark conclusions, the point estimates suggest a contemporaneous depreciation of about 0.8% following a 1% inflationary shock to CPI. The median response suggests a peak depreciation at about 1.4% after one quarter, which then levels off at about 0.9% after five quarters. This is evidence that our decision to leave the contemporaneous impact of inflation on the exchange rate unrestricted was well-warranted; restricting it to zero would have forced the intercept of this IRF to be zero, and exerted great but possibly undue effect on our estimates of the exchange rate pass-through, discussed below. The effect of the output gap and the fed funds rate, panels (b) and (d), are ambiguous and estimated with little precision.

Figure 4.4 is identical in notation and plots the responses of consumer prices to 1% shocks to any of the other variables. The general picture here is that consumer prices do not react very strongly to the other macro variables, at least based on the short run dynamics analysed in our framework. The most pronounced effect comes from monetary shocks (panel (c)): While a monetary expansion does not have much of an impact contemporaneously, after two quarters, a 1% monetary shock is estimated to induce a 0.2% increase in prices.

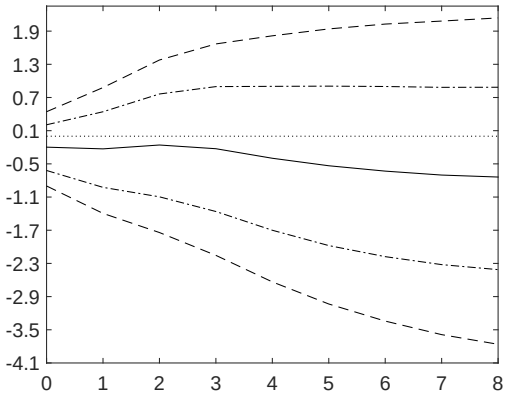
Panel (e), reporting the response of CPI to a shock in the exchange rate – the ERPT to consumer prices – reveals a very moderate response of consumer prices to changes in the exchange rate based on the short run dynamics. At impact, there is no evidence for an increase in consumer prices. While the precision of the estimate is very low given the small amplitude of the effect, the median IRF suggests a steady decrease (increase) in prices following an unanticipated appreciation (depreciation) of the exchange rate. The dynamics settle down after four quarters, after consumer prices have decreased (increased) by about 0.07%.

Our estimates therefore imply an ERPT to consumer prices of 7% after four quarters. This is very low compared to other estimates in the literature. The corresponding figures obtained by Zgambo (2015) range between 41% and 49%, although this is reduced to 20% when focusing on a more recent period (2004-2014) characterized by lower inflation. Choudhri and Hakura (2006) estimate an ERPT to domestic prices in Zambia of 15% at impact, and 41% after four quarters.

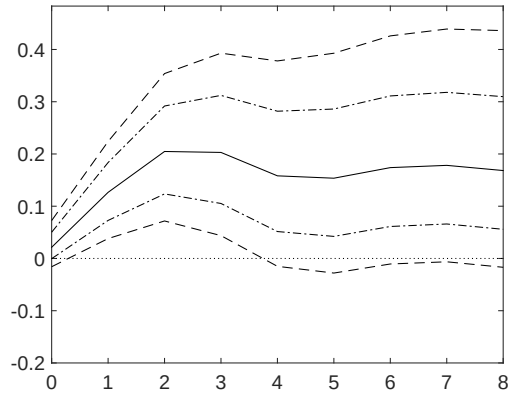
Figure 4.4: Responses of consumer prices to a 1% shock to key variables



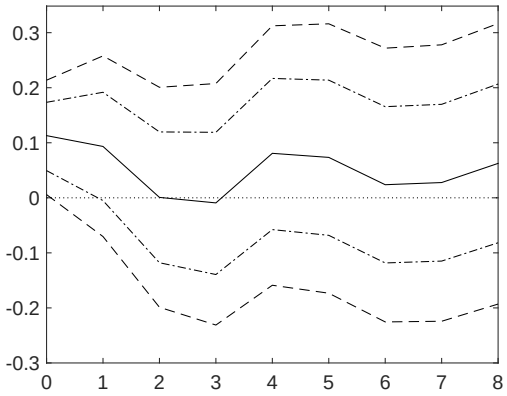
(a) Copper price



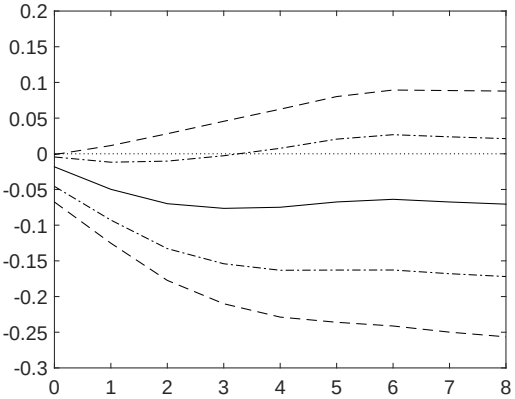
(b) Fed funds rate



(c) Money (expansionary)



(d) Output gap



(e) Exchange rate

Notes: The solid line reports the pointwise mean IRF, the dashed lines depict the 68^{th} and 90^{th} percentiles of the estimated IRFs (credible intervals).

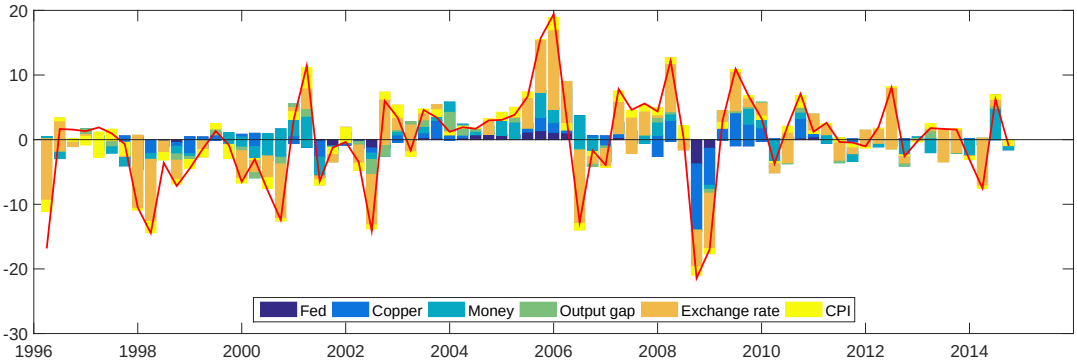
The average pass-through after four quarters they obtain for developing countries is 24% (although with some notable exceptions, like 9% in Burundi, and -1% in Colombia). It is worth emphasizing that, besides other important methodological differences in their study, Choudhri and Hakura (2006) study the period 1979–2000 which overlaps with our sample for only five years. The discrepancy is somewhat consistent with the main point of their paper: low inflation is associated with a low pass-through (a point that has since frequently been confirmed, see Devereux and Yetman, 2010; Lopez-Villavicencio et al., 2016; Jasova et al., 2016). In our sample period, Zambian inflation is much lower than in earlier decades, averaging 18% per annum over 1995-2014 compared to 71% per annum during 1986-2000 (WDI 2016). Our results may therefore reflect a genuine increase in the resilience of the Zambian economy until 2014. A second factor is our decision to leave the initial impact of inflationary shocks on the exchange rate, π_E, unrestricted. Estimating the same set of equations with full identification through a classical Cholesky decomposition ($\pi_E = 0$, no sign restrictions) leads to a higher ERPT estimate of 9% at impact and about 15% after four quarters. This is indeed what much of the previous ERPT literature has been doing, suggesting that some of the results are inflated as causality was constrained to run from exchange rates to prices.

4.5.2 Historical decomposition

Figures 4.5 and 4.6 describe the historical contributions of each of the shocks to the changes in the exchange rate and consumer prices. The red line is the observed value of the respective (de-trended) series, expressed in quarter to quarter changes. The coloured bars indicate the contribution of each of the shocks to this value over the eight periods (two years) prior to the indicated date, computed from the median IRFs discussed in the previous section. These illustrate the decompositions reported in table 4.1, which shows that money supply (M2) and the copper price account for most of the variation in exchange rates, and M2 is usually the more important factor (except for 1996–99 and 2008–11). The lower panel shows that the output gap, money supply and the copper price account for most of the variation in exchange rates (their relative importance varies, with M2 generally becoming more important over time). The role of domestic inflation as a driver of the exchange rate appears to have steadily declined, from 19% in 1996-99 to 6% in 2012–14.

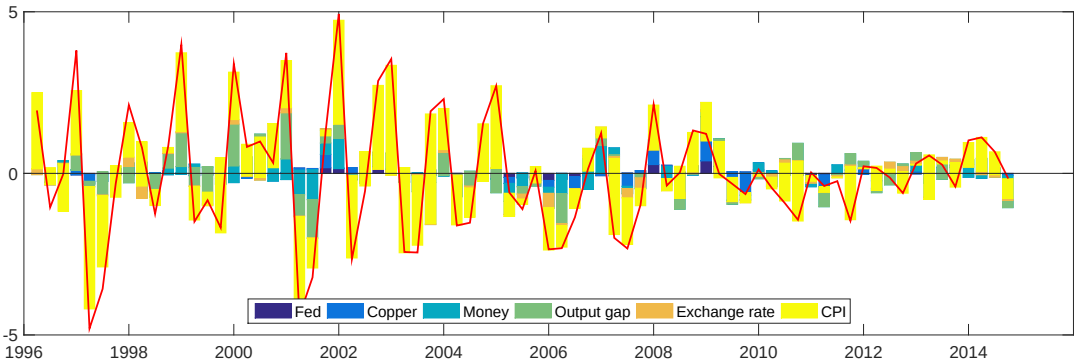
Figure 4.5 illustrates these results for the exchange rate and depicts the year by year variation. It is apparent that the factors contributing to exchange rate fluctuations have varied quite substantially over time, but also that the amplitude

Figure 4.5: Historical decomposition of changes in the exchange rate (USD/ZWA)



Notes: The red line indicates the values of the exchange rate in log-differences, the bars indicate the contributions of the respective shocks to the respective value based on our estimated parameters.

Figure 4.6: Historical decomposition of changes in Zambian consumer prices



Notes: The red line indicates the values of the CPI in log-differences (de-trended series), the bars indicate the contributions of the respective shocks to the respective value based on our estimated parameters.

of the shocks – and therefore that of the exchange rate fluctuations – has declined since about 2009. The graph also offers insights about the nature of individual fluctuations. For instance, the early 2000s saw a number of stark movements in the exchange rate, with money appearing to be the main driver. The strong appreciation in 2005, corresponding to the debt relief under the HIPC initiative, enters mainly as a genuine exchange rate shock. While the proportional contribution of the price of copper appears rather moderate over much of the sample period, the abrupt depreciation in 2009 illustrates its potential importance, as it accounts for about two-thirds of the variation. The variety of the drivers of these shocks further underlines the importance of our research question, whether the ERPT will vary with the underlying economic shock that originally caused the exchange rate to move.

Figure 4.6 decomposes changes in consumer prices in the same way. As expected from the low amplitude of the IRFs presented in section 5.1, the contribu-

tion of the other variables in our system to changes in CPI appears to be relatively limited, and much of the variance is attributed to price shocks that are exogenous to our system. However, as fluctuations decline over time, it is apparent from table 4.1 that the role of our macro-variables is increasing towards the end of the period: the role of genuine price shocks steadily decreases from 55% in 1996-99 to 34% in 2012-14. A few factors have a notable impact in certain periods. This is the case for the fluctuations in output, especially in periods before 2002. Contractionary monetary shocks contributed to slowdowns in inflation in 2001 and in 2006. While copper prices do not play a prominent role in determining inflation in the first half of the sample period, it becomes more important in the late 2000s. This is plausible, as both copper production as well as prices were significantly higher in later periods (see section 4.2).

Table 4.1: Historical variance decomposition of exchange rate and price fluctuations

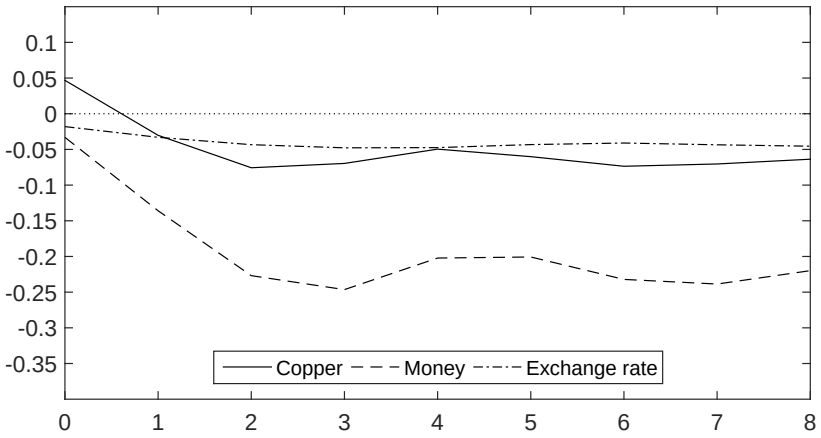
		1996q2–1999q4	2000q1–2003q4	2004q1–2007q4	2008q1–2011q4	2012q1–2014q4
Exchange rate	Fed	3%	8%	9%	6%	3%
	Copper	22%	13%	16%	29%	15%
	M2	15%	24%	24%	20%	31%
	Output Gap	9%	10%	9%	4%	6%
	Exchange Rate	30%	32%	33%	32%	40%
	CPI	19%	13%	8%	8%	6%
CPI	Fed	2%	4%	5%	6%	1%
	Copper	10%	6%	13%	23%	14%
	M2	7%	15%	18%	20%	24%
	Output Gap	21%	21%	15%	10%	21%
	Exchange Rate	5%	3%	8%	6%	6%
	CPI	55%	51%	41%	35%	34%

Notes: The first panel reports the contribution of each variable to the fluctuations of the exchange rate averaged across quarters within five sub-periods. The second panel reports the same information for fluctuations in CPI.

4.5.3 Differential exchange rate pass-through

We will now examine the extent to which different shocks that affect the exchange rate entail different reactions of consumer prices. Our approach is similar to Forbes et al. (2015), in that we use the ratio of the IRF of CPI over the IRF of the exchange rate for a shock to any given variable. These estimates therefore indicate the change in consumer prices induced by a shock that leads to a 1% appreciation in the exchange rate in any period. The interpretation is then slightly different to the IRFs reported in section 4.5.1. These traced the effect of a shock that causes

Figure 4.7: Pass through to consumer prices for different shocks



a 1% appreciation at impact. The measures reported here are computed for a shock that causes a 1% appreciation in the period itself. For instance, in order for a shock to the copper price to lead to an exchange rate appreciation of 1% at impact, it needs to have an amplitude of $1\%/0.2 = 5\%$ (see figure 4.3). But the shock takes some time to fully deploy its effect on the exchange rate; a copper price shock that leads to an appreciation of 1% after four quarters therefore only needs to have an amplitude of $1\%/0.4 = 2.5\%$.

Note also that the direction of the original shocks we consider in this section need not be identical to that discussed in section 4.5.1 when describing the IRFs. Specifically, figure 4.7 reports the result of a contractionary monetary shock (reducing inflation), while our previous discussion as well as figures 3 and 4 were based on expansionary shocks (increasing inflation). This is because we consider shocks that lead to an exchange rate appreciation, which for money needs to be a contractionary one. Since in section 4.5.1 we established that both the fed funds rate and the output gap had little discernible effect on the exchange rate, we focus on shocks to copper prices, money and the exchange rate itself.[5]

Table 4.2: Pass through for different shocks and time horizons

Horizon	**0**	**1**	**4**	**8**
Copper	5%	-2%	-5%	-6%
Money	-3%	-14%	-21%	-22%
Exchange rate	-2%	-3%	-5%	-5%

Figure 4.7 plots the results. First, note that for a shock to the exchange rate,

[5]The latter may seem redundant, as conceptually it is very similar to the exchange rate pass-through already quantified above. Because of calculation in each period, and because an exchange rate shock that raises it by 1% in the first period may (and does) have a different effect in the second period, it is however not identical.

the effect is even slightly lower than that reported in section 4.5.1. This is because a shock that raises the exchange rate by 1% at impact still raises it in the subsequent periods; effectively, Figure 4.7 refers to smaller exchange rate shocks in all periods after impact. A 1% increase in the exchange rate caused by a shock to copper prices comes with an increase in consumer prices of 0.05%, implying a negative pass-through at impact of about -5%. This may reflect aggregate demand effects related to the stimulus to the copper industry (increased employment, rents). The effect is reversed in later periods, with a pass through between 5% and 8% in the following quarters. Finally, monetary shocks are associated with by far the strongest pass-though: exchange rate fluctuations that are caused by a monetary shock are estimated to translate into consumer prices with an ERPT of about 25% after three quarters, with a lasting effect of about 22%. This is particularly relevant in the view of the results from the historical variance decomposition in section 4.5.2, which indicates that, while the variation of the exchange rate has markedly declined overall, it has recently increasingly been driven by monetary shocks. The remaining exchange rate fluctuations may therefore be disproportionately associated with consumer prices. Note however that this is more than offset by the smaller amplitude of the fluctuations in general.

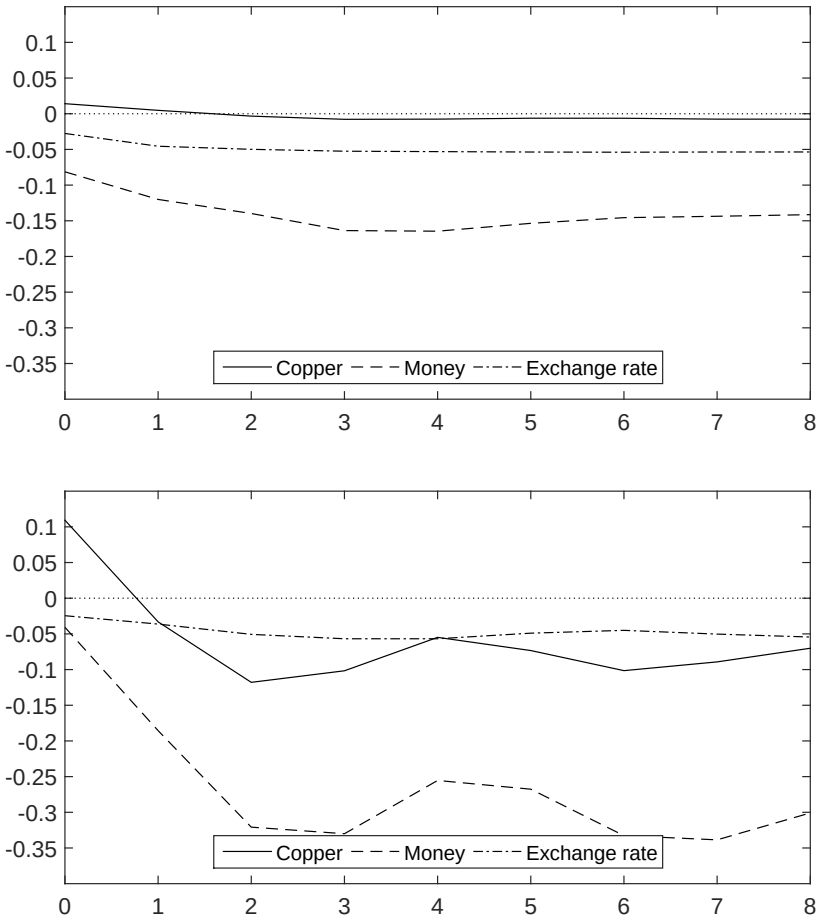
4.5.4 Food versus non-food inflation

We further repeat the previous exercise with disaggregated measures of inflation, that is, we repeat the entire analysis as outlined previously, replacing only the overall CPI series with the food and non-food CPI respectively. Appendix C.1 also reports the priors employed in this additional exercise, which we derived for each CPI category separately in the same manner as the one employed previously and described in section 3.1. For the sake of conciseness, we restrict our discussion to the differential exchange rate pass-through.[6]

Figure 4.8 then summarizes the differential pass-through, with an interpretation identical to that of figure 4.7. For genuine exchange rate shocks, the difference appears to be only marginal with an ERPT of about 5%. As for monetary and copper price shocks, there seems to be a marked difference between food and non-food inflation. Non-food inflation (top panel) appears almost indifferent to the exchange rate fluctuations induced by copper price shocks, and exhibits a pass-through of about 15% for fluctuations driven by monetary shocks, as opposed to nearly 25% in the case of overall inflation (previous section). In fact, the exercise suggests that the overall pass-through is overwhelmingly driven by food prices

[6]A document containing all the graphs from the results section for the alternative CPI measures can be downloaded at `https://tinyurl.com/y7zyhs5b`.

Figure 4.8: Differential ERPTs for non-food (top) and food (bottom) inflation



(bottom panel). Exchange rate fluctuations induced by copper prices come with a pass through of up to 10%, those resulting from monetary shocks with up to 33%. This corroborates the finding of a higher ERPT for food items in Zgambo (2015).

To some extent, this may appear puzzling as food items are typically sourced domestically (with most of the labor force working in agriculture), and often produced on small scale farms that do not tend to employ large amounts of imported inputs. Possible explanations could be lower menu costs in the often informal market for food items, or the fact that processed food items are almost exclusively imported. The latter would additionally increase in relevance if there existed an urban bias (or a bias towards formal food vendors such as supermarkets) in the CPI figures, as this is where imported food items prevail. We leave this as an open question for future research.

4.6 Exchange rate and price shocks 2015

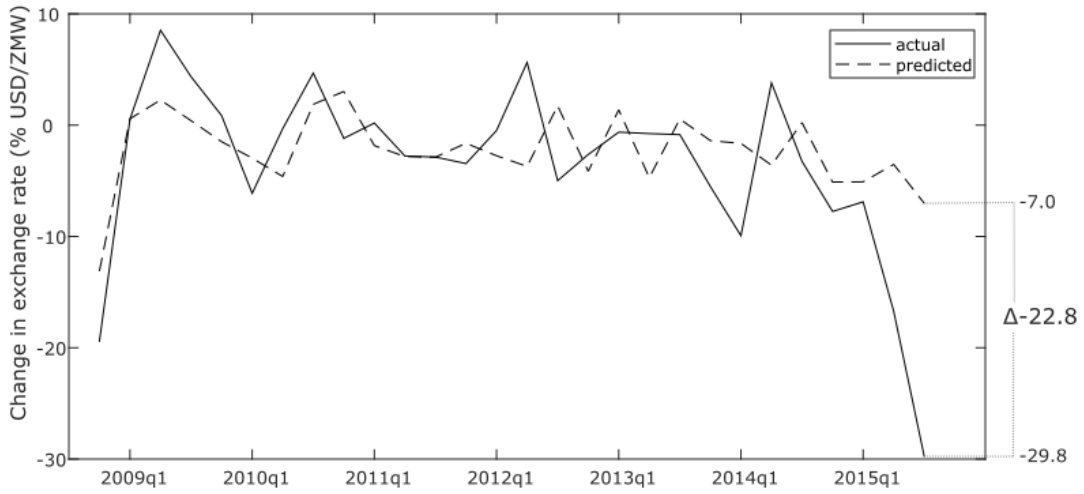
In order to investigate the dynamics of the kwacha exchange rate and its interaction with consumer prices, we deliberately excluded the periods after 2015 from our estimations, as the amplitude of these events means that they would have exerted undue influence on our parameter estimates. Having established the dynamics in arguably normal times, we will conclude our analysis by quantifying the extent to which the variables deviated from their econometrically expected behaviour in late 2015. To this end, we compute the predicted values for each period up until 2015, therefore extending beyond our original sample period but using the parameters derived from it.

Figure 4.9 plots the predicted series against the observed values since 2009. In the first six years, the difference between the observed exchange rate fluctuations and the predicted series is on average below 4 percentage points (top panel). However, the fundamentals included in our model fail to predict or explain the massive depreciation in late 2015. For the last quarter of 2015, it would have predicted a depreciation of 7%, while the kwacha actually depreciated by almost 30%, a discrepancy of 23 percentage points. An even more striking discrepancy arises between the observed and predicted CPI series (figure 4.9, bottom panel). Mainly due to a general economic downturn, as captured in the Output gap, that counteracts the inflationary pressures arising from the depreciation, consumer prices (adjusted for the trend) were expected to stay stable (-0.1%) according to the model in the last quarter of 2015. Instead, monthly inflation shot up by an unexpected 9.6%.

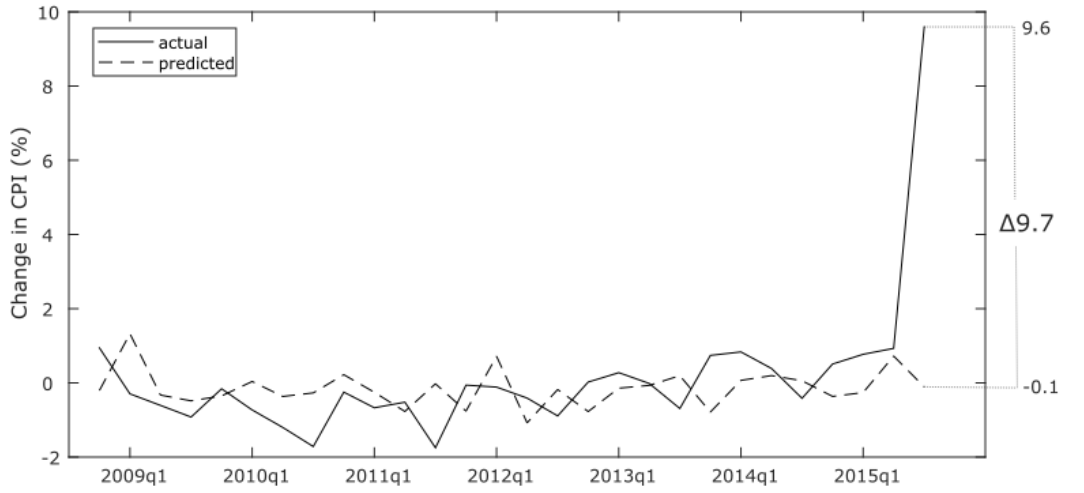
These discrepancies point towards dynamics that are substantially out of the ordinary, taking the preceding two decades as a benchmark. In view of the abruptness and peculiarity of the events, it should be emphasized that a narrative case study is needed in order to understand the mechanisms that led up to the sweeping currency depreciation and price inflation. While this is not the main focus of our paper, we wish to highlight some key aspects that future research should seek to explore. Our discussions with policy makers and private stakeholders suggested the following aspects were key in understanding the abrupt currency depreciation:

Thin market: The currency market in Zambia is very thin and dominated by a few large players, most notably mining companies, the central bank, and some private banks. Individual transactions can have a great weight in the determination of the market exchange rate, something that is perpetuated as such transactions work as a signal for other market participants.

Figure 4.9: Predicted and observed changes in exchange rate and CPI



(a) Predicted exchange rate



(b) Predicted CPI

Trade deficit: In December 2014, Zambia's balance of trade turned negative after nearly a decade of surpluses. In the first half of 2015, there was a widespread apprehension among currency traders that the negative trade balance would persist or worsen. This was confirmed by the government in late August, inducing a sharp and sudden depreciation.

Breakdown of price revealing mechanisms: In an attempt to limit speculative and panic behavior, authorities closed the conventional foreign exchange market, where offers could openly be published and observed by participants. Contrary to the original intention, this effectively led to a breakdown of price revealing mechanisms and nurtured suspicions among market participants that authorities were artificially keeping the exchange rate from dropping to its 'true' value. This

turned into a self-fulfilling prophecy, and the actual rate consequently dropped to excessively low levels, reaching a low of 14 ZMW/USD on November 10.

Delayed policy response: The central bank remained passive in terms of monetary policy at first; at least in part because of the factors described above, the developments only took up more pace instead of settling. When authorities substantially tightened monetary policy in mid-November, the exchange rate rapidly stabilized, as did inflation.

The fact that consumer price inflation so markedly increased in line with the exchange rate depreciation during this episode may seem at odds with the finding of a relatively low overall exchange rate pass-through. It is also not the case that the exchange rate fluctuations were driven by monetary shocks, which, based on our results, would explain part of the higher pass-through. We conjecture that this is related to the following factors:

Confounding factors: The currency crisis coincided with poor harvests in the agricultural sector, independently driving food price increases.

Size non-linearities: It is often conjectured that small changes in the exchange rate may translate into prices differently than large changes in the exchange rate (Aron et al., 2014; Jasova et al., 2016), for several reasons. First, menu costs may imply that producers and retailers absorb small changes in the exchange rate through their profit margins, and will only adjust prices once a certain threshold has been surpassed. A large depreciation such as in late 2015 would then surpass these thresholds across sectors and firms, leading to a larger ERPT. Similarly, in times of large exchange rate fluctuations, retailers pay increased attention to the exchange rate and may adjust prices at a higher frequency. Anecdotal evidence suggests that this was the case during 2015, with some retailers informally declaring prices in USD, and converting them into kwachas based on the daily exchange rate for every purchase. This practice mechanically leads to a complete ERPT for these transactions.

4.7 Conclusions

We investigated the dynamics of the exchange rate and its interaction with consumer prices in Zambia in a structural VAR framework, identifying shocks with a combination of theoretically plausible zero-restrictions and sign-restrictions. Crucially, we imposed minimal restrictions on the interaction between the two key

variables (inflation and the exchange rate), which sets our study apart from the earlier literature. Furthermore, building on recent developments in the literature, we explored the possibility of differential exchange rate pass-throughs depending on which shock originally caused the exchange rate to fluctuate. To the best of our knowledge, we are the first to do so in the context of a commodity dependent low-income country. Finally, we contrasted our results with the tumultuous period of late 2015, and outlined the key aspects that led to the collapse of 'normal' macroeconomic relationships..

Our findings suggest that the pass through from the exchange rate to consumer prices is relatively moderate for genuine exchange rate shocks; we estimate a 10% depreciation to induce an increase in consumer prices of 0.7% after one year, corresponding to an ERPT of 7%. This is substantially less than what other studies typically find for small, low income countries, and far below the estimates of over 40% for Zambia by Zgambo (2015), and Choudhri and Hakura (2006). This is likely due to two factors. First, unlike much of the comparable literature, we do not impose any directionality on the contemporaneous relationship between the exchange rate and inflation. When doing so, our estimates of the ERPT are more than doubled. Second, we study more recent periods where Zambia has maintained a fairly low level of inflation judged by historical standards. It is a well-established finding that countries with lower inflation typically have a smaller ERPT, and Zambia may indeed have successfully reduced its vulnerability to global shocks in this respect. From a policy perspective, this result supports the view whereby 'fear of floating' is – in times of normal market conditions – not warranted in Zambia, as resilience to exchange rate shocks appears to be larger than what is commonly believed. Instead, it appears as though the Bank of Zambia's commitment to keeping inflation at moderate levels is bearing fruit and should be maintained.

Variance decomposition suggests that the main drivers of exchange rate fluctuations are copper prices and the money supply. With respect to the copper price, it is worth noting that – while it can be an influential factor – this only becomes salient episodically. While money supply has typically accounted for the largest share of exchange rate fluctuations, the influence of copper is concentrated in individual periods of large price volatility; most notably, it accounts for the largest share of the devaluations in the period from 2008 to 2011. As inflation declined throughout the sample period, so did its importance in the determination of the exchange rate.

We further investigated whether the ERPT varies with the shock that initially caused the exchange rate to fluctuate. Our findings suggest that exchange rate fluctuations caused by copper shocks behave very much like genuine exchange rate shocks in that they are associated with an ERPT of only roughly 7%. In terms

of consumer price stability, the copper price should therefore not be considered a major concern. This result supports the view that trying and anticipate the effect of commodity price price shocks on the economy using monetary policy is better to be avoided. Those fluctuations going back to monetary shocks, on the other hand, are associated with a much larger pass-through of up to 25% (22% in the long run). When considering only food inflation, this figure is even higher with up to 33%. Monetary interventions aimed at stabilising the exchange rate therefore have disproportionately large simultaneous effect on consumer prices, once more supporting a rather passive monetary policy with respect to the exchange rate. Where the goal is to maintain price stability, interventions should target this objective directly, and not seek to offset exchange rate fluctuations.

Lastly, we note that the extent of the sharp depreciation of the exchange rate, as well as the steep increase in consumer prices in late 2015 cannot be explained from the underlying fundamentals included in our system. Instead, we suggest a narrative of market panic paired with non-linearities in the ERPT, and outline the core aspects: thin markets, a deteriorated balance of trade, the breakdown of price revealing mechanisms, and a delayed policy response. We suggest that a dedicated case study is required in order to provide a satisfactory narrative of this episode, and leave this as a fruitful topic for further research.

Chapter 5

Conclusion

This thesis sought to shed light on some of the major controversies surrounding growth and macroeconomic performance in sub-Saharan Africa in three self-contained essays. Chapter 2 did so quite literally by employing data on human-made light emissions to trace GDP growth on the continent over the past two decades. Chapter 3 explored the robustness of macroeconomic inference in SSA with respect to differences between GDP series. Chapter 4 explored the relationship between export commodity prices, exchange rates and consumer prices in Zambia.

Regarding the recent African growth performance, two findings stand out from the analysis in Chapter 2. First, I do not find evidence in support of an 'African growth miracle', according to which African incomes would have grown up to three times faster than what is suggested by official growth rates (Young, 2012). Second, I do observe that countries that recently rebased their GDP tended to report growth rates substantially higher than what is suggested by lights. This lends some support to the hypothesis of 'statistical growth' (Jerven, 2014), whereby economic activity that had previously gone unobserved is increasingly captured as statistical capacity improves, and spuriously attributed to recent economic growth.

The analysis in Chapter 3 shows that inferences on the macroeconomic impact of aid in SSA are sensitive to the GDP data source employed, although not to such an extent as to reject growth analysis based on African macroeconomic data altogether. For approximately one third of the countries, the inference is substantially affected by changes across datasets. With the re-specified models, changes tend to be more substantial in the sense that, for some countries, results remain significant but point in the opposite direction. These reversals apply to the subset of 12 countries for which full data were available and I did not detect any characteristics in the data that explained why signs reversed for some cases but not for others. However, for about half of the countries considered, the results remained stable throughout all exercises.

In the final chapter on Zambia, we find that a 10% devaluation in the exchange rate caused by a shock to the copper price induces consumer prices to increase by only about 0.7%. By contrast, a devaluation of the same amplitude that has been triggered by an increase in money supply is followed by an increase in prices of up to 2.5%. Generally, our results suggest that the Zambian economy may be somewhat more resilient to external shocks than previously thought.

Three main themes emerge from each of these studies. The first is a call for differentiation, in the sense of avoiding across-the-board approaches or uniform assumptions, and one-size-fits-all recommendations. The central point of Chapter 2 was that the assumption of a common (uniform across countries) elasticity between GDP and lights does not hold up to close scrutiny. As economic structures, geographic features, and societies differ across countries, so does the expected amount of light that is emitted for every incremental unit of economic activity. If this is not appropriately taken into account, the growth rates indicated by the growth proxy can be highly misleading. Chapter 3 shows that differences between, in principle, conceptually identical or similar GDP datasets can sometimes significantly alter inference. Where feasible, researchers should therefore avoid basing their conclusions on just one vintage of a particular series; robustness checks with respect to the data should be standard practice, in particular where the data are weak (of poor quality). Chapter 4 emphasises that the same observed fluctuation of a macro variable can have very different implications depending on the context. Specifically, we show that fluctuations in the exchange rate can affect consumer prices to a very different extent depending on the shock that originally caused the exchange rate to move. Conflating all these effects into a single measure, the exchange rate pass-through, is then once more an oversimplification; as Forbes et al. (2015) put it, 'the shocks matter'.

The second theme that emerges is a call for moderation. Novel results and general and often strong conclusions are ubiquitous in the abstracts of top economic journals, arguably at the expense of less eye-catching, but perhaps more accurate analyses. Chapter 2, for instance, fails to find evidence for an 'African growth miracle' as proclaimed by Young (2012), whereby African incomes would have grown at three times the rate suggested by national statistics. Instead, my results suggest that growth rates in SSA between 1992 and 2013 were indeed very high, but, on average, did not exceed the reported figures. Chapter 3, while it does show that changes in the dataset can have a substantial impact on the results, also shows that in the majority of cases (about two thirds of the countries), the results remain robust regardless the GDP dataset employed. Caution is then clearly warranted, but discarding African national accounts data altogether, and the econometric evidence that is based on it, is not. Chapter 4, in turn, suggests

that the vulnerability of the Zambian economy to external shocks at least in relatively normal times may not be as high as previously believed. This adds some nuance to a discourse where the performance of small, resource-rich economies, is too often reduced to the world price of the commodity they export.

Thirdly, and on a more general note, researching macroeconomics in sub-Saharan Africa can be a frustrating endeavour: the data may often be unreliable or difficult to verify reliability, and some conclusions will be sensitive to this. Political and economic regimes change frequently, and this is sometimes hardly documented. As most macroeconomic research focusses on advanced economies, there is comparatively little theoretical guidance when conducting research on the least developed countries in the world. The list of challenges is considerably longer, but low-income countries (where the data are most questionable, especially in SSA), are still home to half of the world's poorest people; careful research targeted at understanding better aspects of their economic performance will always be worthwhile. Macroeconomic policy and performance may be difficult to analyse but remain important to the welfare of individuals; randomised controlled trials may offer insight on which treatments are effective for particular issues in specific contexts, but the associated policy recommendations may all be in vain without the right macroeconomic framework.

Future research will build up on the work presented in this thesis, in particular on Chapters 2 and 4. The method presented in Chapter 2 can be extended in several ways; for instance, weighted approaches to estimating the determinants of the lights-GDP nexus could be explored, giving countries with high quality data more weight. The method could also be explored on a more abstract level, generalising it beyond the narrow application in the context of luminosity as a proxy for growth. The empirical framework of Chapter 4 will be adapted for further research on the exchange rate dynamics in sub-Saharan African countries; specifically, I will explore the exchange rate dynamics associated with inflows of foreign aid.

Appendix A

Appendix Chapter 2

A.1 Sorting of countries in to groups/clusters

Table A.1: Sorting of countries into groups

	Fossil	Econ.	Regime		Fossil	Econ.	Regime
Africa				Swaziland	2	2	1
Angola	1	1	2	Chad	1	1	2
Burundi	2	1	2	Togo	2	1	2
Benin	2	2	3	Tunisia	2	2	2
Burkina Faso	2	2	2	Tanzania	2	1	2
Botswana	2	3	3	Uganda	2	1	2
CAF	2	1	2	South Africa	2	3	3
Cte D'Ivoire	2	2	2	Zambia	2	2	2
Cameroon	2	2	2	Zimbabwe	2	2	2
Congo, DR	2	2	2	**Americas**			
Congo, Republic	1	1	2	Argentina	2	2	3
Djibouti	2	3	2	Bahamas	2	3	-
Algeria	1	1	2	Belize	2	2	-
Egypt	2	2	2	Bolivia	2	2	3
Eritrea	2	2	1	Brazil	2	3	3
Ethiopia	2	1	2	Canada	2	3	3
Gabon	1	1	2	Chile	2	3	3
Ghana	2	1	2	Colombia	2	2	3
Guinea	2	1	2	Costa Rica	2	2	3
Gambia	2	1	2	Cuba	2	3	1
Guinea Bissau	2	1	2	Dom. Rep.	2	2	3
Kenya	2	2	2	Ecuador	2	2	3
Liberia	2	-	2	Guatemala	2	2	3
Libya	1	1	2	Guyana	2	1	3
Lesotho	2	2	3	Honduras	2	2	3
Morocco	2	2	2	Haiti	2	-	2
Madagascar	2	2	3	Jamaica	2	3	3
Mali	2	-	3	Mexico	2	3	3
Mozambique	2	2	2	Nicaragua	2	2	3
Mauritania	2	1	2	Panama	2	3	3
Mauritius	2	3	3	Peru	2	2	3
Malawi	2	2	2	Puerto Rico	2	2	-
Namibia	2	2	3	Paraguay	2	2	3
Niger	2	1	2	El Salvador	2	2	3

Continued on next page

Table A.1: Sorting of countries into groups (continued)

	Fossil	Econ.	Regime		Fossil	Econ.	Regime
Nigeria	1	1	2	Suriname	2	2	2
Rwanda	2	1	2	Trinidad & Tob.	1	3	3
Sudan	2	1	2	Uruguay	2	2	3
Senegal	2	2	2	USA	2	3	3
Sierra Leone	2	1	2	Venezuela	1	2	2
South Sudan	-	-	2				

Continued on next page

Table A.1: Sorting of countries into groups (continued)

	Fossil	Econ.	Regime		Fossil	Econ.	Regime
Asia				**Europe**			
Afghanistan	2	2	2				
UAE	1	-	1	Albania	2	1	3
Armenia	2	2	2	Austria	2	3	3
Azerbaijan	1	1	1	Belgium	2	3	3
Bangladesh	2	2	2	Bulgaria	2	-	3
Bahrain	1	3	1	Bosnia and Herz.	2	2	2
Brunei	1	3	-	Belarus	2	2	2
China	2	2	1	Switzerland	2	3	3
Cyprus	2	3	3	Czech Republic	2	3	3
Georgia	2	2	2	Germany	2	3	3
Indonesia	2	2	2	Denmark	2	3	3
India	2	2	3	Spain	2	3	3
Iran	1	2	2	Estonia	2	3	3
Iraq	1	-	2	Finland	2	3	3
Israel	2	-	3	France	2	3	3
Jordan	2	3	2	United Kingdom	2	3	3
Japan	2	3	3	Greece	2	3	3
Kazakhstan	1	2	2	Croatia	2	3	2
Kyrgyzstan	2	2	2	Hungary	2	3	3
Cambodia	2	2	2	Ireland	2	3	3
Korea	2	3	3	Iceland	2	3	-
Kuwait	1	3	1	Italy	2	3	3
Laos	2	1	1	Lithuania	2	3	3
Lebanon	2	3	3	Luxembourg	2	3	3
Sri Lanka	2	2	2	Latvia	2	3	3
Myanmar	2	1	2	Moldova	2	2	3
Mongolia	2	1	3	Macedonia	2	2	3
Malaysia	2	2	2	Montenegro	2	3	3
Nepal	2	1	2	Netherlands	2	3	3
Oman	1	3	1	Norway	2	3	3
Pakistan	2	2	2	Poland	2	3	3
Philippines	2	2	3	Portugal	2	3	3
Palestine	2	3	-	Romania	2	2	3
Qatar	1	3	1	Russia	1	3	2
Saudi Arabia	1	3	1	Serbia	2	2	3

Continued on next page

Table A.1: Sorting of countries into groups (continued)

	Fossil	Econ.	Regime		Fossil	Econ.	Regime
Syria	1	2	1	Slovakia	2	3	3
Thailand	2	2	2	Slovenia	2	3	3
Tajikistan	2	2	2	Sweden	2	3	3
Turkmenistan	1	2	1	Ukraine	2	2	3
Timor-Leste	2	1	-	**Oceania**			
Turkey	2	2	3	Australia	2	3	3
Uzbekistan	1	1	1	New Caledonia	2	3	-
Viet Nam	2	2	1	New Zealand	2	3	3
Yemen	1	1	2	Papua New G.	2	1	2

A.2 Data Generating Process (detailed)

Income data (Y)

Initial GDP per capita: In order reflect the highly right-skewed distribution of national incomes, we choose an exponential distribution to generate initial values of GDP per capita. We scale up the series to values of a similar order of magnitude of empirically observed income levels as measured in US-Dollars:

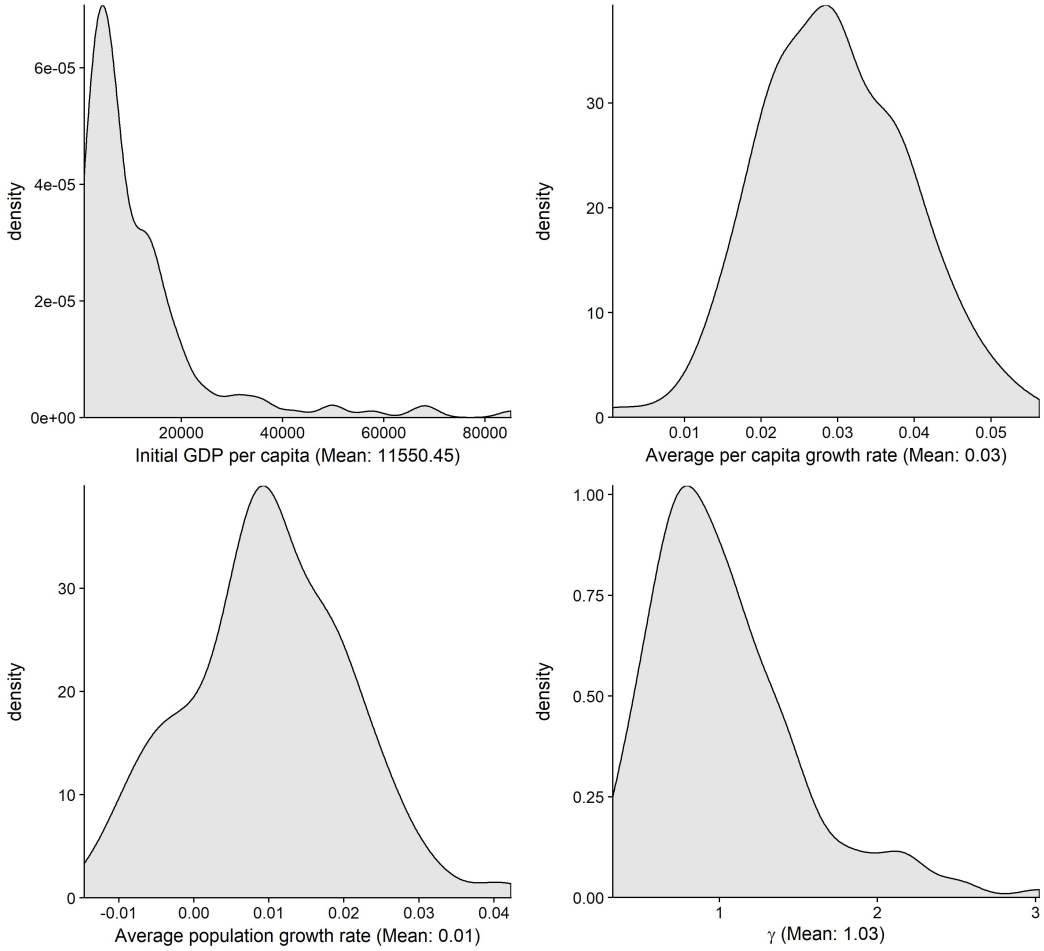
$$y_{init} \sim Exp(1) * 10.000 + 500 \tag{A.1}$$

The resulting distribution is depicted in the top left panel of figure A.1.

Co-determinants of growth and γ: In order to satisfy requirement 4, we introduce three variables that simultaneously determine GDP growth rates of Y and the inverse elasticity between GDP and lights γ for each of the countries. First, we introduce one arbitrary uniformely distributed variable $v_1 \sim U(0.2, 0.8)$ that will both increase GDP growth rates and affect each country's γ (e.g., levels of investment). Second, we introduce some beta-convergence, with countries that have smaller initial GDP per capita values tending to have larger growth rates; GDP per capita will also affect γ. And third, population growth will naturally enter total GDP growth rates, and we include it as another determinant of γ (see below).

Per capita GDP growth: In order to approximately match the mean and variance of empirical per capita growth rates, we choose an unconditional mean of per capita growth rates of 0.03; the expected annual growth rate per capita for

Figure A.1: Distribution of key series of the simulated data



Notes: Kernel density plots of the distribution of key values in the simulated data. The plotted distributions refer to the simulated data discussed in sections 2.4.1 to 2.4.3.

any given country is then $g_{i,det} = 0.03 + (v_1 - 0.5) + ((\mu_{y_{init}} - y_{init})/10 * \mu_{y_{init}})$, where $\mu_{y_{init}}$ is the mean initial GDP per capita across countries; the last term therefore introduces convergence. Beyond this deterministic component, we add some stochasticity, and the average growth rates in each country are are distributed $g_i \sim N(g_{i,det}, 0.01)$. Panel 2 in figure A.1 plots the resulting distribution of average per capita growth rates.

Population growth: To obtain figures similar to empirical figures in their amplitude and distribution, we generate initial population numbers $pop_{init} \sim Exp(1) * 10^6$; these then grow at country-specific rates $pop_{growth} \sim N(0.1, 0.1)$ that remain constant across time. Figure A.1, panel 3 plots the distribution of population growth rates.

The final GDP series is then simply $Y_t = y_t * pop_t$.

Generating an imperfect proxy ℓ (lights)

Next, we generate a proxy ℓ of y, where y induces ℓ with (inverse) elasticity γ and some measurement error. Importantly, we want to allow for γ to vary across units of observation (countries) based on a number of determinants. We therefore first introduce a basic DGP for γ.

DGP for γ: The country-specific component of the inverse elasticities γ are partly determined by the confounding factors (v_1, y_{init} and pop_{growth}) that affect both the elasticities as well as GDP growth, and partly by factors that only affect the elasticities. The latter are each generated as $det_j \sim N(0, SD_{det})$. We also generate a number of identically defined variables that *do not* enter the γs, in order to assess the variable selection properties of our proposed estimator. In the present set-up, we generate 20 of these normally distributed variables, 5 of which enter as determinants into the GDP predictions. The deterministic component of the γs is defined as $\gamma_{det} = c_\gamma + \sum_{j=1}^{5} det_j + f(v_1) + f(y_{init}) + f(pop_{growth})$, with $c_\gamma = 0.5$. We add an element of randomness by generating the final γs as $\gamma_i \sim N(\gamma_{det,i}, SD_\gamma)$. Panel 4 in figure A.1 depicts the resulting distribution.

Proxy series ℓ: Based on the generated unit-specific γ_is, we can now generate luminosity measures from the true income measure Y. In line with section 2.3, we determine $L_{it} = Y_{it}^{\beta_i} * exp(\varepsilon_{it})$, where $\beta_i = \frac{1}{\gamma_i}$. Moreover, $\varepsilon_i t = \epsilon_t + \epsilon_i + \epsilon_{it}$, such that the measurement error in the lights-GDP relationship is a composite of a time specific component (e.g., variability in the sensor sensitivity) that will be accounted for with time fixed effects ($\epsilon_t \sim N(1, SD_{time})$ and an idiosyncratic error ($\epsilon_t \sim N(1, SD_{idio})$).

Figure 2.1 plots Y and L for 12 randomly selected units (countries). The solid line represents the simulated GDP series, and the dashed line represents L; as intended, L follows GDP with some error, and with different elasticities in different countries.

'Bad data' countries In order to simulate misreporting, we create another series of *reported* GDP (z). We therefore divide our sample into two groups: Good data countries, where the reported GDP data is accurate, that is, equal to the true series. And bad data countries, where the reported GDP growth rates are systematically biased. The direction and amplitude of the bias contains an element of randomness. We opt for the systematic error to be distributed uniformly in a range specified by *Range misrep.*. The number of bad data countries is determined by N_{bad}

A.3 Results of simulation exercise

Figure A.2: Predictive power across the parameter range

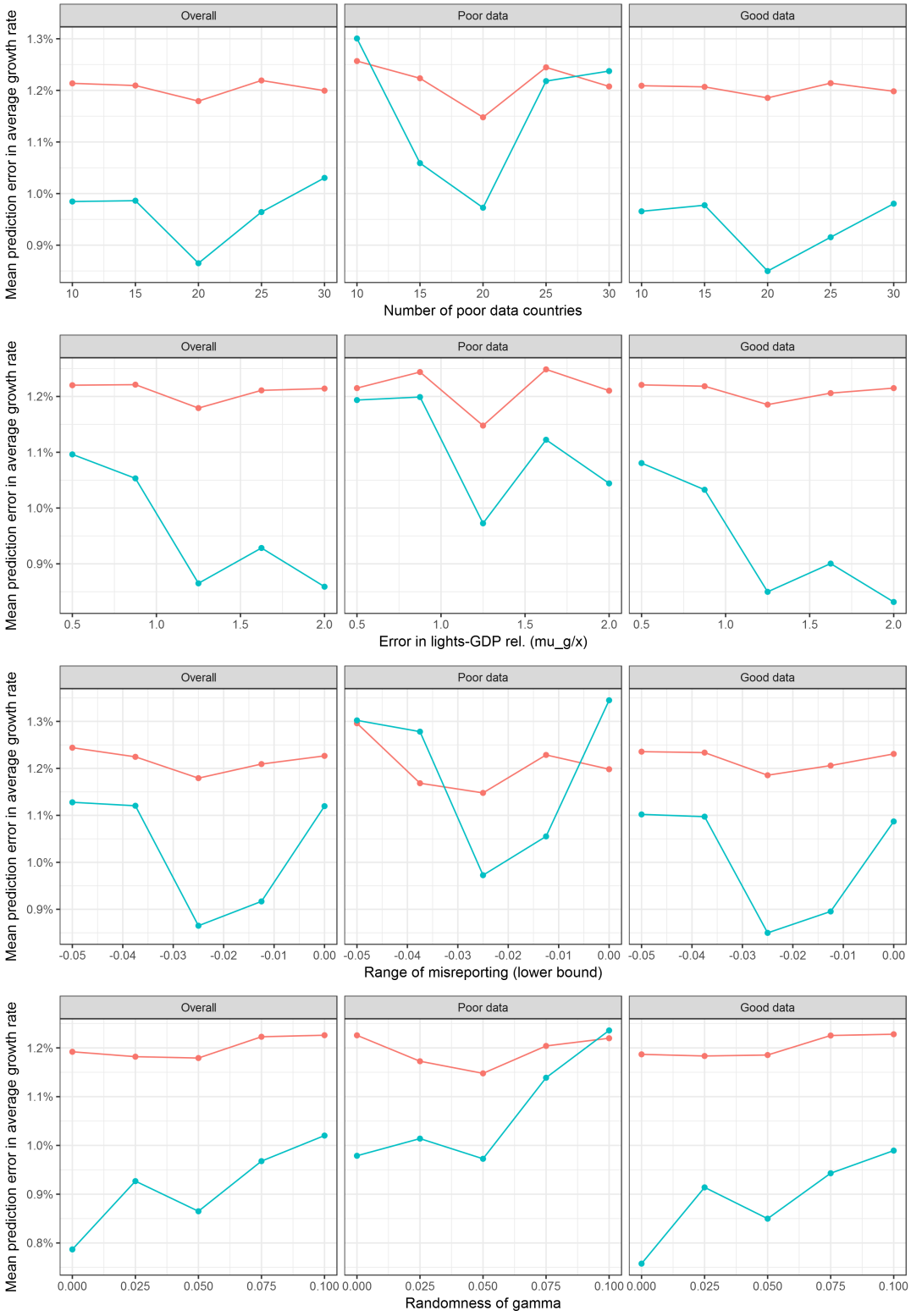


Figure A.2 plots the mean error in the predicted average annual growth rates across the (simulated) sample for a range of parameters. Each row shows how this

measure changes as the respective parameter (see label of the x-axis) is changed, keeping all other parameters constant. The values at which the (respectively) other parameters are kept constant are $N_{bad} = 20$, $SD_{\epsilon} = \mu_g/1.25$, $SD_{\gamma} = 0.05$ and a range of misreporting of $U[-0.25; 0.25]$ (almost identical to the illustratory example in the main body, with minor deviations as the intervals do not exactly coincide with those values). The blue lines refer to the two-stage estimator, the red line to the single-slope fixed effects.

A.4 Estimated country-specific coefficients $\hat{\gamma}_i$

Table A.2: Country-specific $\hat{\gamma}$

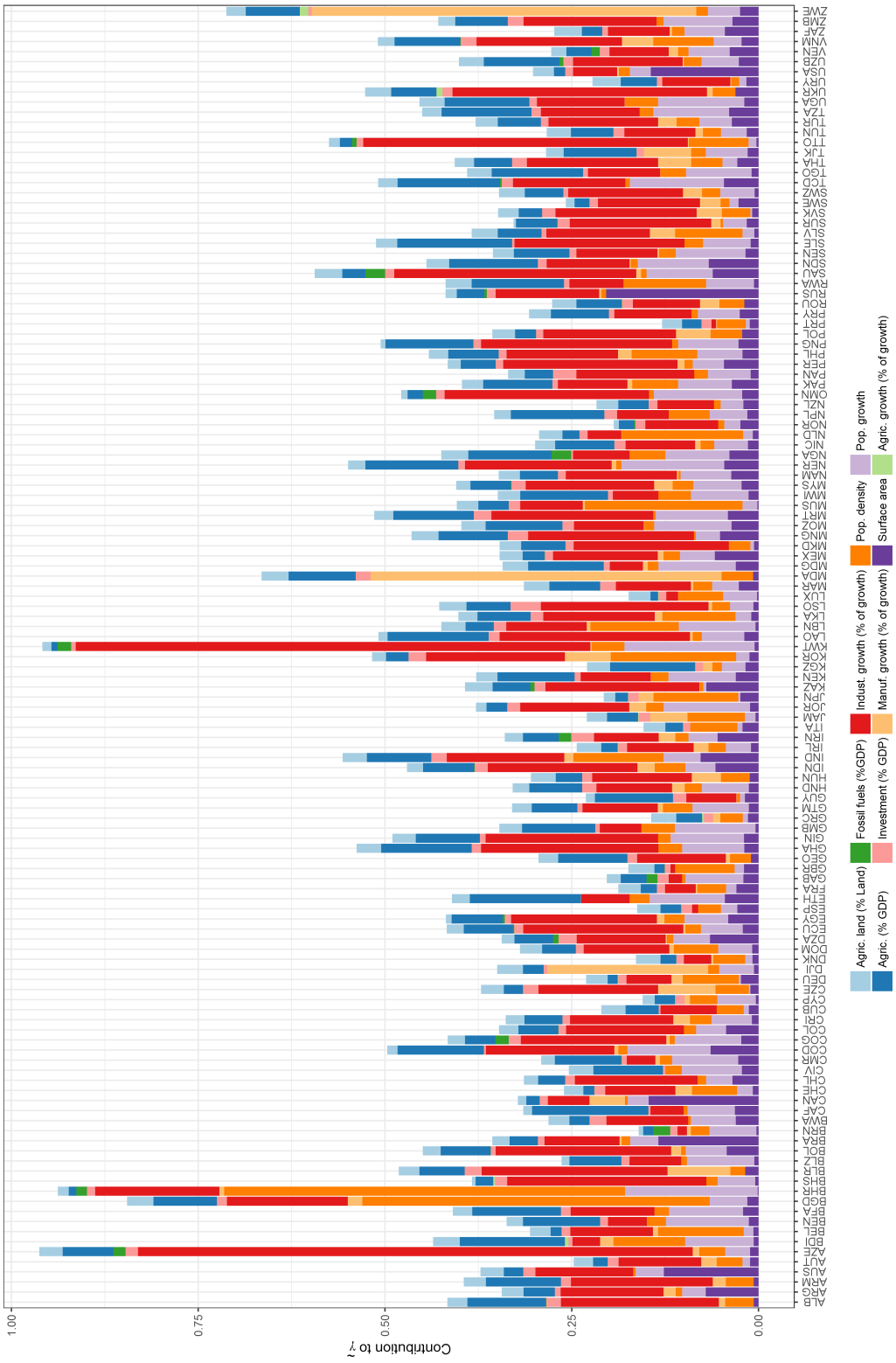
	$\hat{\gamma}$	SE		$\hat{\gamma}$	SE		$\hat{\gamma}$	SE		$\hat{\gamma}$	SE
AGO	0.44	0.02	DEU	-0.16	0.06	KEN	0.23	0.04	POL	0.24	0.03
ALB	0.29	0.03	DJI	0.04	0.03	KGZ	0.12	0.04	PRI	-0.24	0.11
ARE	0.41	0.05	DNK	-0.06	0.04	KOR	0.68	0.07	PRT	-0.26	0.05
ARG	0.16	0.05	DOM	0.52	0.05	KWT	0.79	0.07	PRY	0.07	0.04
ARM	0.50	0.03	DZA	0.14	0.04	LAO	0.36	0.02	ROU	0.04	0.03
AUS	0.27	0.08	ECU	0.12	0.04	LBN	0.14	0.03	RUS	0.24	0.04
AUT	-0.13	0.05	EGY	0.39	0.05	LBR	0.34	0.01	RWA	0.47	0.02
AZE	1.53	0.04	ESP	-0.05	0.08	LKA	0.45	0.04	SAU	0.34	0.04
BDI	0.03	0.03	ETH	0.54	0.03	LSO	0.16	0.04	SDN	0.42	0.03
BEL	0.05	0.07	FIN	-0.02	0.05	LUX	0.19	0.06	SEN	0.20	0.04
BEN	0.20	0.03	FRA	-0.27	0.08	MAR	0.23	0.04	SLE	0.30	0.02
BFA	0.45	0.03	GAB	-0.19	0.04	MDA	0.35	0.02	SLV	0.03	0.06
BGD	0.59	0.05	GBR	0.04	0.11	MDG	0.06	0.04	SUR	0.14	0.04
BGR	0.14	0.05	GEO	0.26	0.02	MEX	0.00	0.07	SVK	0.01	0.02
BHR	1.12	0.12	GHA	0.51	0.04	MKD	-0.08	0.05	SWE	0.00	0.02
BHS	-0.24	0.09	GIN	0.10	0.04	MLI	0.54	0.03	SWZ	-0.04	0.04
BLR	0.32	0.03	GMB	0.07	0.02	MNG	0.46	0.03	TCD	0.40	0.02
BLZ	0.22	0.03	GNB	0.20	0.01	MOZ	0.58	0.03	TGO	0.13	0.04
BOL	0.21	0.04	GRC	-0.20	0.06	MRT	0.27	0.04	THA	0.17	0.03
BRA	0.14	0.05	GTM	0.13	0.03	MUS	0.45	0.06	TJK	-0.09	0.08
BRN	-0.16	0.04	GUY	0.19	0.04	MWI	0.40	0.05	TKM	0.76	0.05
BWA	0.28	0.04	HKG	0.65	0.17	MYS	0.34	0.03	TTO	0.46	0.04
CAF	0.21	0.02	HND	0.13	0.03	NAM	0.36	0.05	TUN	0.32	0.05
CAN	0.05	0.03	HUN	-0.01	0.05	NER	0.24	0.04	TUR	0.22	0.04
CHE	-0.18	0.06	IDN	0.31	0.04	NGA	0.75	0.05	TZA	0.54	0.04
CHL	0.28	0.04	IND	0.86	0.05	NIC	0.20	0.04	UGA	0.62	0.04
CHN	1.05	0.03	IRL	0.57	0.05	NLD	0.00	0.07	UKR	0.38	0.02
CIV	-0.08	0.03	IRN	0.26	0.05	NOR	-0.03	0.02	URY	0.20	0.04
CMR	0.20	0.05	IRQ	0.51	0.04	NPL	0.26	0.04	USA	-0.01	0.09
COD	0.07	0.05	ISL	0.03	0.02	NZL	0.09	0.08	UZB	-0.05	0.05
COG	0.14	0.03	ISR	0.59	0.09	OMN	0.06	0.03	VEN	0.07	0.07
COL	0.28	0.06	ITA	-0.67	0.09	PAK	0.46	0.09	VNM	0.37	0.02
CRI	0.35	0.04	JAM	-0.50	0.09	PAN	0.52	0.04	YEM	0.11	0.03
CUB	0.22	0.04	JOR	0.48	0.04	PER	0.42	0.04	ZAF	0.18	0.08
CYP	0.13	0.07	JPN	-0.33	0.12	PHL	0.33	0.05	ZMB	0.72	0.05
CZE	0.05	0.05	KAZ	0.52	0.02	PNG	0.08	0.06	ZWE	1.22	0.03

A.5 Value and composition of $\tilde{\gamma}_i$ by country

Table A.3: Country-specific coefficients $\tilde{\gamma}$

ALB	0.28	COG	0.32	HUN	0.17	MRT	0.43	SEN	0.24
ARG	0.19	COL	0.20	IDN	0.35	MUS	0.30	SLE	0.36
ARM	0.24	CRI	0.20	IND	0.47	MWI	0.21	SLV	0.24
AUS	0.14	CUB	0.08	IRL	0.09	MYS	0.28	SUR	0.16
AUT	0.01	CYP	-0.04	IRN	0.26	NAM	0.25	SVK	0.22
AZE	0.89	CZE	0.23	ITA	-0.14	NER	0.43	SWE	0.00
BDI	0.35	DEU	-0.07	JAM	0.06	NGA	0.33	SWZ	0.20
BEL	0.08	DJI	0.20	JOR	0.25	NIC	0.17	TCD	0.44
BEN	0.21	DNK	-0.07	JPN	-0.18	NLD	0.07	TGO	0.29
BFA	0.30	DOM	0.19	KAZ	0.31	NOR	-0.07	THA	0.30
BGD	0.77	DZA	0.26	KEN	0.28	NPL	0.25	TJK	0.07
BHR	0.81	ECU	0.29	KGZ	0.04	NZL	0.02	TTO	0.48
BHS	0.20	EGY	0.32	KOR	0.34	OMN	0.36	TUN	0.19
BLR	0.36	ESP	-0.07	KWT	0.76	PAK	0.30	TUR	0.24
BLZ	0.12	ETH	0.32	LAO	0.40	PAN	0.21	TZA	0.34
BOL	0.33	FRA	-0.09	LBN	0.26	PER	0.28	UGA	0.37
BRA	0.16	GAB	0.07	LKA	0.31	PHL	0.33	UKR	0.40
BRN	-0.04	GBR	-0.09	LSO	0.23	PNG	0.29	URY	0.09
BWA	0.21	GEO	0.15	LUX	-0.07	POL	0.22	USA	-0.11
CAF	0.15	GHA	0.43	MAR	0.21	PRT	-0.06	UZB	0.32
CAN	0.06	GIN	0.38	MDA	0.52	PRY	0.17	VEN	0.11
CHE	-0.02	GMB	0.22	MDG	0.22	ROU	0.14	VNM	0.43
CHL	0.19	GRC	-0.08	MEX	0.18	RUS	0.26	ZAF	0.14
CIV	0.16	GTM	0.19	MKD	0.21	RWA	0.29	ZMB	0.35
CMR	0.16	GUY	0.10	MNG	0.40	SAU	0.46	ZWE	0.55
COD	0.37	HND	0.20	MOZ	0.26	SDN	0.36		

Figure A.3: Contribution of determinants to $\tilde{\gamma}$ by country



Notes: The bars depict the absolute contributions of each of the determinants of $\tilde{\gamma}_i$ for each country individualy. Non-linear transformations and the variables in levels are conflated here for readability. 'Industrial growth' etc. correspond to *relative* contributions of the sectors to total GDP growth, that is, they are not actual growth rates.

A.6 Cross-validation

The goal of cross-validation is to find the parameters that lead to the model with the smallest out-of-sample prediction error. In other words, it serves to avoid over-fitting, which may be a serious issue where the number of variables is large compared to the number of observations: Instead of finding a model of general validity, a regression may then simply end up 'memorising' the data in the sample, that is, as in-sample predictions become better, out of sample predictions become worse. As in most empirical applications, testing the out-of-sample predictions is complicated by the fact that we only observe the sample itself.

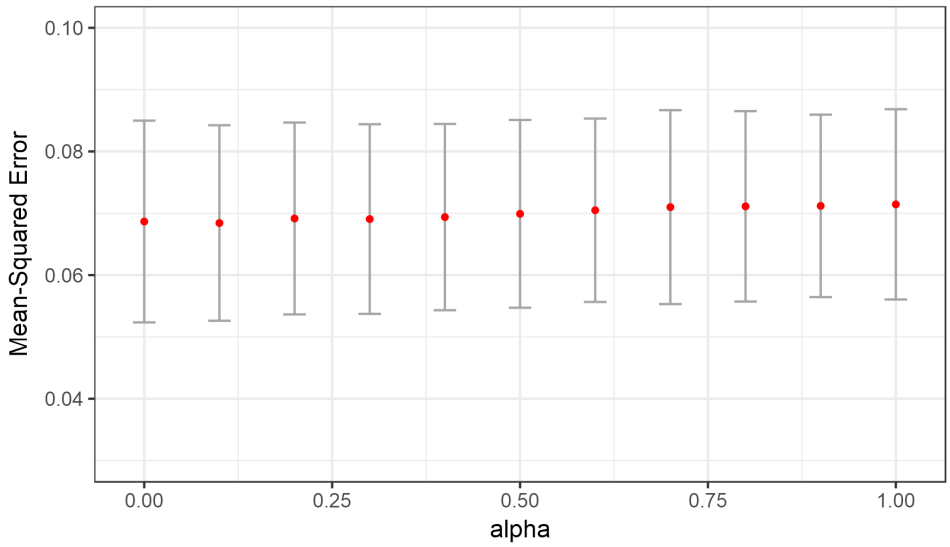
Cross-validation tackles this problem by iteratively dividing the existing sample into a training set and a test set. The training set is the one that is used to estimate the model, while the test set is 'hidden'. Once the model has been estimated, it is then used to predict the dependent variable based on the independent variables (here: $\hat{\gamma}$ based on $\mathbf{\Phi^*}$). The (squared) prediction error – 'out-of-sample', as the test set was not in the sample underlying the estimated model – of this prediction is to be minimised. The key choice to make at this point is about the division of the sample into training and test set. We use 'leave-one-out' cross-validation (LOO), which is particularly rigorous as it uses every country as a test sample once, using the remaining countries as a training sample. For every combination of the parameters to be tuned, the elastic net estimation is therefore carried out N times; thanks to the limited size of our dataset ($N = 129$ in this exercise), this is feasible within reasonable computation times.

We perform this procedure for both *alpha* and λ; note that this notation (with a single λ) refers to a slightly re-parametrised version of equation 2.2:

$$L(\lambda, \alpha, \boldsymbol{\delta}) = |\hat{\boldsymbol{\gamma}} - \mathbf{\Phi^*}\boldsymbol{\delta}|^2 + \lambda[(1-\alpha)|\boldsymbol{\delta}|^2 + \alpha|\boldsymbol{\delta}|_1] \tag{A.2}$$

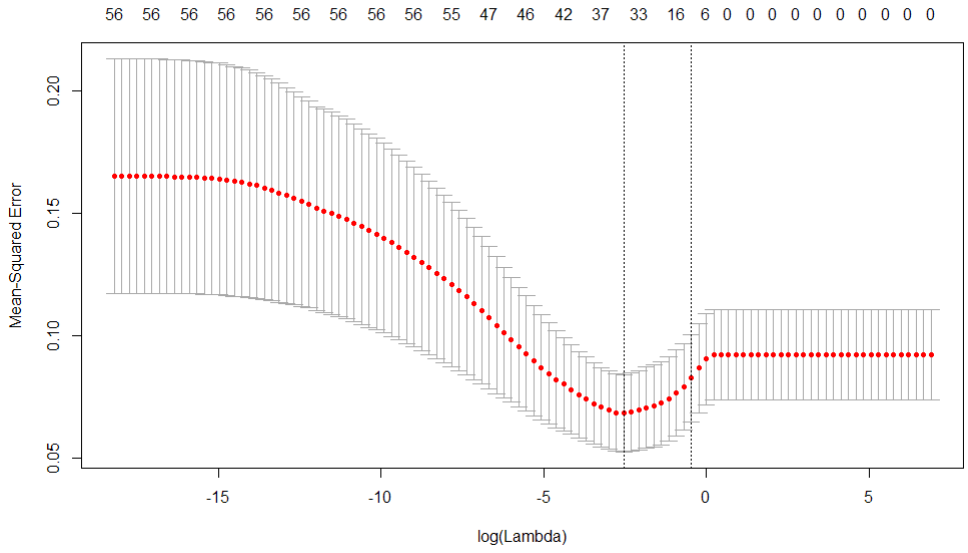
so α still refers to the parameter that dominates the relative importance of the lasso and the ridge penalty, and λ is the overall strength of the penalty. Note that as we tune α, for every value of α we iterate through the entire range of λs. The mean-squared error reported in figure A.4 is then the one resulting from the specification with the optimal λ given that α. Figure A.5 then reports the mean-squared error obtained from different values of λ given the optimal α. As a general observation, relative importance of the lasso and the ridge penalty hardly matters for the predictive properties of the estimated model: The out-of-sample mean-squared error is almost invariable to the choice of α (figure A.4). On the other hand, the choice of λ is very influential, highlighting the importance of the regularisation procedure: With the (virtually) unrestricted model (very small λ, all 56 candidate variables included), the out of sample prediction error is about

Figure A.4: Cross-validation of α



Notes: The points indicate the mean squared error at different values of alpha (see text for details). The grey bars indicate the 95% confidence intervals for these point estimates. The value of α appears to have negligible impact on the quality of the predictions, i.e., the lasso and the ridge penalies have very similar impact.

Figure A.5: Cross-validation of λ



Notes: The points indicate the mean squared error at different values of λ (see text for details). The grey bars indicate the 95% confidence intervals for these point estimates. The numbers in the top row along the x-axis indicate the degrees of freedom, that is, the number of variables with non-zero coefficient. Compared to the case with (virtually) no regularisation, we can halve the out-of-sample prediction error by tuning lambda.

0.16, compared to a mean of the dependent variable, $\hat{\gamma}$, of 0.25. At the optimum, the MSE is 0.07.

A.7 Coefficients $\hat{\delta}$ from elastic net

Table A.4: Estimated coefficients, elastic net

Variable	Coeff.	Variable	Coeff.	Variable	Coeff.
Investment	-	Investment (Sq.)	0.000	Investment (Sqrt.)	-
Consumption	-0.000	Consumption (Sq.)	- 0.000	Consumption (Sqrt.)	-0.012
Agriculture	0.000	Agriculture (Sq.)	-	Agriculture (Sqrt.)	0.007
Industry	-	Industry (Sq.)	- 0.000	Industry (Sqrt.)	-
Sevices	-	Sevices (Sq.)	- 0.000	Sevices (Sqrt.)	-
Pop. Density	0.000	Pop. Density (Sq.)	0.000	Pop. Density (Sqrt.)	0.000
Pop. growth	0.021	Pop. growth (Sq.)	0.001		
Urban pop.	-0.000	Urban pop. (Sq.)	-	Urban pop. (Sqrt.)	-0.001
Forest land	-0.000	Forest land (Sq.)	- 0.000	Forest land (Sqrt.)	-0.000
Agric. land	-	Agric. land (Sq.)	-	Agric. land (Sqrt.)	0.002
Manufacturing	-	Manufacturing (Sq.)	- 0.000	Manufacturing (Sqrt.)	-
Agric. growth	-0.633	Agric. growth (Sq.)	- 2.216		
Indust. growth	2.356	Indust. growth (Sq.)	10.459		
Serv. growth	-	Serv. growth (Sq.)	-		
Manuf. Growth	-0.718	Manuf. Growth (Sq.)	21.902		
GDP level	-0.000	GDP level (Sq.)	-	GDP level (Sqrt.)	-0.000
GDP per capita	-0.000	GDP per capita (Sq.)	-	GDP per capita (Sqrt.)	-0.000
Snow cover	-	Snow cover (Sq.)	-		
Latitude	-	Latitude (Sq.)	-	Latitude (Sqrt.)	-
Surface area	0.000	Surface area (Sq.)	-	Surface area (Sqrt.)	0.000
Fossil fuels	-	Fossil fuels (Sq.)	0.000		

Notes: The reported coefficients are normalised by the respective variables standard deviation. Coefficients reported as (–)0.000 are in fact rounded very small non-zero coefficients, coefficients that have been eliminated (attributed weights that are actually zero) are reported as a dash (-).

A.8 Discrepancy between lights and GDP series

A.8.1 Suggested corrections to growth rates (tabular)

Table A.5: WDI vs. luminosity estimates of growth (1993-2013, 1993-1999)

	Full period (1993-2013)							1993-1999				
	WDI	EN	Δ EN	FE	Δ FE	LD	Δ LD	WDI	EN	Δ EN	FE	Δ FE
AGO	6.28			5.16	-1.12	5.71	-0.57	2.62			2.28	-0.34
BDI	1.03	4.33	3.30	3.87	2.84	3.36	2.34	-3.41	3.08	6.49	2.51	5.92
BEN	4.40	3.75	-0.65	4.19	-0.21	4.29	-0.11	4.75	3.27	-1.48	3.61	-1.14
BFA	5.88	4.87	-1.01	4.69	-1.20	4.86	-1.03	6.08	5.01	-1.06	4.56	-1.51
BWA	4.82	3.82	-1.00	4.29	-0.53	4.46	-0.36	5.26	3.65	-1.61	4.11	-1.15
CAF	0.64	2.39	1.75	2.27	1.63	1.83	1.19	3.15	1.76	-1.38	1.46	-1.68
CIV	2.40	3.72	1.32	4.69	2.29	4.69	2.29	3.68	4.51	0.83	6.50	2.82
CMR	3.29	3.04	-0.24	3.45	0.16	3.36	0.07	2.49	2.29	-0.20	2.45	-0.04
COD	1.59	3.90	2.31	3.50	1.90	3.42	1.83	-4.17	1.72	5.90	1.43	5.60
COG	3.21	4.47	1.25	4.21	1.00	4.24	1.03	0.33	1.72	1.38	1.38	1.05
DJI	1.91	3.50	1.59	3.89	1.98	3.54	1.63	-1.95	1.38	3.32	0.93	2.88
ETH	7.79	5.10	-2.70	4.74	-3.05	4.97	-2.83	5.67	4.76	-0.91	4.13	-1.54
GAB	2.25	2.79	0.53	3.72	1.46	3.69	1.44	2.36	1.97	-0.40	2.30	-0.06
GHA	5.83	5.10	-0.73	4.11	-1.72	4.14	-1.69	4.31	6.21	1.90	4.41	0.10
GMB	3.42	4.19	0.77	4.69	1.26	4.59	1.17	3.01	3.69	0.68	4.01	1.00
GNB	0.92			1.53	0.61	2.26	1.34	0.05			-0.59	-0.65
KEN	3.72	4.03	0.31	4.01	0.29	3.93	0.21	2.52	3.71	1.19	3.47	0.95
LBR	8.83			6.64	-2.19	7.87	-0.96	15.91			6.06	-9.85
LSO	3.87	3.86	-0.01	4.14	0.27	4.03	0.16	3.16	3.70	0.54	3.91	0.75
MDG	2.73	3.78	1.05	4.15	1.42	3.71	0.98	2.60	4.42	1.83	5.01	2.42
MLI	7.22			4.87	-2.36	4.91	-2.32	4.72			4.74	0.02
MOZ	8.54	4.90	-3.65	5.12	-3.42	5.62	-2.92	10.65	3.78	-6.87	3.73	-6.92
MRT	4.33	5.59	1.26	4.45	0.12	4.43	0.10	3.56	6.24	2.68	4.43	0.87
MUS	4.32	3.55	-0.76	3.44	-0.87	3.29	-1.02	4.78	3.99	-0.79	3.62	-1.16
MWI	4.57	3.13	-1.44	3.34	-1.23	3.21	-1.36	4.89	2.52	-2.37	2.56	-2.33
NAM	4.05	3.53	-0.52	3.66	-0.39	3.54	-0.51	2.59	2.89	0.30	2.80	0.21
NER	4.20	5.31	1.10	4.26	0.06	4.18	-0.03	3.44	4.53	1.09	3.28	-0.16
NGA	5.97	3.23	-2.74	3.06	-2.91	2.65	-3.32	1.95	1.63	-0.33	1.33	-0.62
RWA	6.13	4.88	-1.25	4.79	-1.34	4.28	-1.84	2.84	2.99	0.15	2.65	-0.19
SDN	5.19	4.88	-0.32	4.29	-0.90	4.31	-0.89	5.07	3.22	-1.85	2.62	-2.45
SEN	3.71	3.99	0.27	4.28	0.56	4.14	0.42	3.43	3.55	0.11	3.68	0.25
SLE	4.94	5.69	0.74	4.91	-0.03	4.10	-0.84	-1.84	-0.57	1.27	-0.54	1.30
SWZ	2.60	3.40	0.80	3.78	1.18	3.82	1.22	3.26	3.20	-0.06	3.60	0.34
TCD	6.39	7.49	1.10	5.60	-0.79	6.01	-0.38	1.40	4.19	2.79	2.96	1.56
TGO	3.13	3.45	0.32	3.40	0.26	3.30	0.16	4.45	2.61	-1.83	2.37	-2.08
TZA	5.46	4.45	-1.02	4.07	-1.39	3.78	-1.68	3.28	4.79	1.51	4.05	0.77
UGA	7.04	5.11	-1.92	4.41	-2.63	4.23	-2.80	7.63	5.31	-2.32	4.19	-3.44
ZAF	3.05	2.86	-0.19	3.16	0.12	3.01	-0.04	2.48	2.34	-0.14	2.60	0.13
ZMB	5.41	3.79	-1.62	3.48	-1.93	3.42	-1.99	2.20	2.28	0.09	1.91	-0.29
ZWE	0.29	3.32	3.03	2.75	2.46	2.27	1.98	3.65	4.04	0.39	2.65	-1.00

Notes: EN are the estimates obtained from the two-step procedure involving the elastic net estimator (and equation 2.1), FE those obtained from the Fixed Effects estimation with a common slope (equation 2.5), LD those from the long-difference specification (equation 2.6). LD are only obtained for the full period. Δ indicates the difference between the respective estimate and the growth rates implied by WDI.

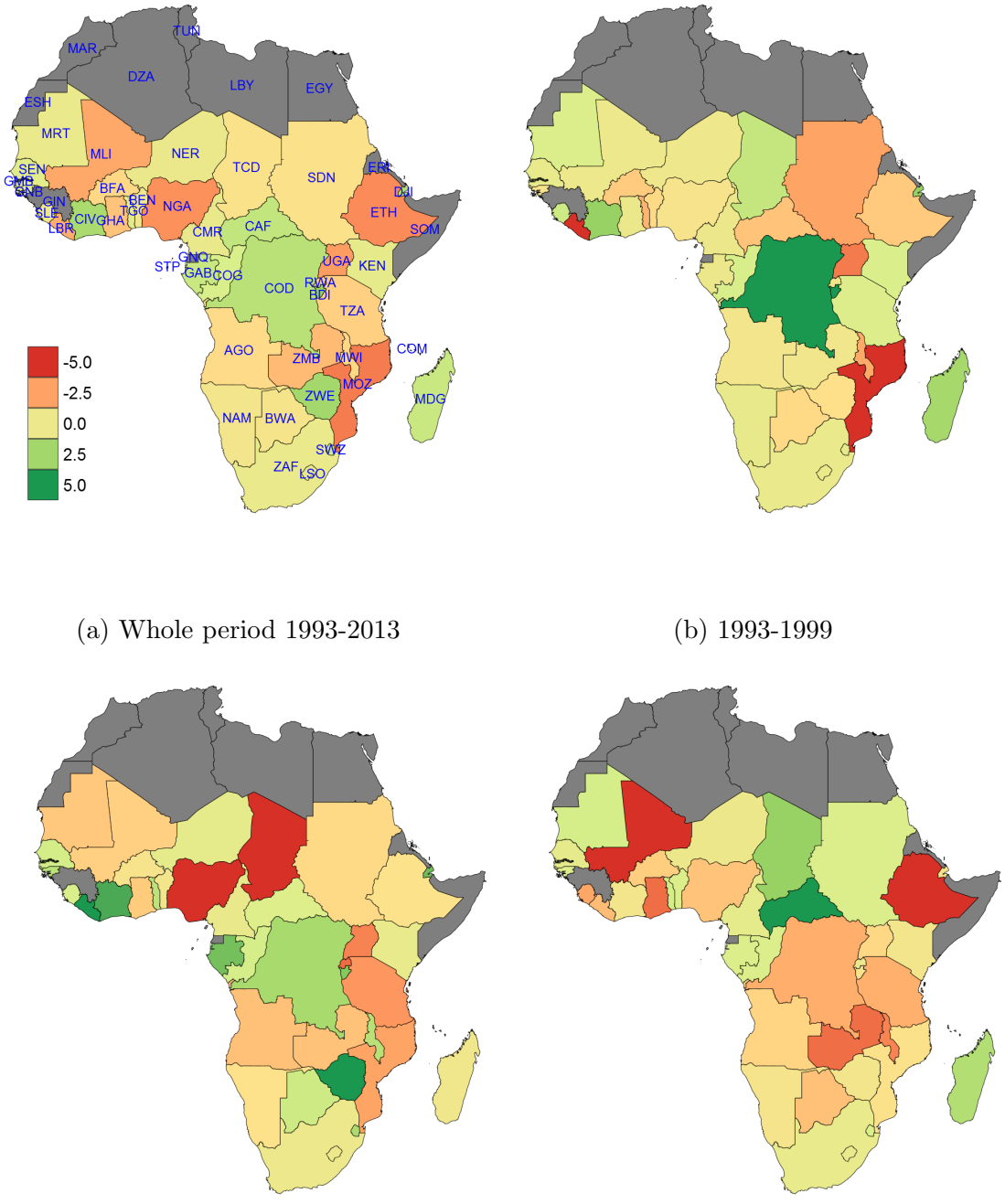
Table A.6: WDI vs. luminosity estimates of growth (2000-2013)

	2000-2006					2007-2013				
	WDI	EN	Δ EN	FE	Δ FE	WDI	EN	Δ EN	FE	Δ FE
AGO	7.92			6.15	-1.76	8.30			7.04	-1.26
BDI	2.22	5.68	3.46	5.45	3.23	4.27	4.21	-0.06	3.65	-0.62
BEN	4.20	3.75	-0.45	3.81	-0.38	4.26	4.23	-0.03	5.16	0.90
BFA	5.71	4.99	-0.72	5.09	-0.62	5.85	4.61	-1.24	4.40	-1.45
BWA	4.08	4.88	0.80	5.40	1.32	5.11	2.93	-2.19	3.35	-1.76
CAF	1.53	3.03	1.49	2.34	0.81	-2.77	2.37	5.14	3.00	5.77
CIV	-0.07	4.05	4.13	4.30	4.37	3.59	2.60	-0.99	3.26	-0.33
CMR	3.71	3.81	0.10	3.86	0.15	3.66	3.03	-0.64	4.05	0.39
COD	2.53	4.96	2.43	4.87	2.34	6.42	5.02	-1.40	4.19	-2.23
COG	4.89	5.99	1.10	5.96	1.07	4.41	5.69	1.28	5.28	0.87
DJI	2.87	5.64	2.77	6.54	3.67	4.82	3.48	-1.34	4.20	-0.62
ETH	7.14	6.42	-0.71	6.33	-0.81	10.57	4.11	-6.46	3.77	-6.80
GAB	0.46	3.94	3.48	4.03	3.57	3.94	2.45	-1.48	4.82	0.88
GHA	5.04	3.16	-1.88	3.55	-1.50	8.13	5.92	-2.20	4.37	-3.76
GMB	3.16	4.60	1.43	4.97	1.81	4.09	4.29	0.20	5.07	0.98
GNB	1.18			3.02	1.84	1.61			2.28	0.67
KEN	3.62	3.72	0.10	3.87	0.25	5.02	4.66	-0.36	4.70	-0.32
LBR	3.01			8.17	5.16	7.57			5.69	-1.88
LSO	3.41	4.09	0.68	4.27	0.86	5.03	3.79	-1.24	4.25	-0.78
MDG	3.25	3.14	-0.12	3.03	-0.22	2.34	3.79	1.45	4.40	2.06
MLI	7.46			6.06	-1.40	9.49			3.79	-5.70
MOZ	8.01	5.26	-2.75	5.57	-2.44	6.97	5.65	-1.32	6.07	-0.90
MRT	5.97	4.39	-1.58	4.38	-1.59	3.46	6.15	2.69	4.55	1.09
MUS	4.04	3.93	-0.12	4.07	0.03	4.12	2.73	-1.39	2.63	-1.49
MWI	2.49	4.14	1.65	4.37	1.88	6.34	2.73	-3.61	3.10	-3.24
NAM	5.08	4.00	-1.08	4.16	-0.92	4.49	3.71	-0.78	4.01	-0.48
NER	3.48	3.68	0.19	3.89	0.40	5.68	7.72	2.03	5.61	-0.07
NGA	9.89	3.25	-6.65	3.46	-6.43	6.06	4.83	-1.24	4.38	-1.68
RWA	7.86	3.91	-3.96	4.04	-3.82	7.67	7.73	0.06	7.66	-0.01
SDN	6.92	6.04	-0.88	5.76	-1.16	3.59	5.37	1.78	4.50	0.90
SEN	4.15	4.88	0.73	5.25	1.10	3.55	3.54	-0.02	3.90	0.35
SLE	7.35	9.34	1.99	8.36	1.02	9.32	8.29	-1.03	6.92	-2.40
SWZ	2.23	4.20	1.97	4.46	2.23	2.30	2.79	0.49	3.27	0.97
TCD	12.23	6.13	-6.10	5.46	-6.76	5.54	12.13	6.60	8.36	2.82
TGO	1.28	2.90	1.62	3.02	1.74	3.67	4.84	1.16	4.81	1.14
TZA	6.52	3.58	-2.95	3.76	-2.76	6.58	4.97	-1.61	4.41	-2.17
UGA	6.78	3.29	-3.49	3.53	-3.25	6.70	6.74	0.04	5.51	-1.19
ZAF	4.14	3.97	-0.17	4.12	-0.02	2.53	2.28	-0.25	2.77	0.24
ZMB	6.12	4.30	-1.82	4.37	-1.75	7.92	4.80	-3.12	4.16	-3.75
ZWE	-6.07	1.47	7.54	2.70	8.77	3.28	4.45	1.17	2.89	-0.40

Notes: See table A.5.

A.8.2 Suggested corrections with single $\hat{\gamma}$

Figure A.6: Growth according to official vs. lights data (common slope)



(a) Whole period 1993-2013

(b) 1993-1999

(c) 2000-2006

(d) 2007-2013

Notes: See figure 2.11

Appendix B

Appendix Chapter 3

B.1 The relative importance of prices and NA data

B.1.1 Construction of alternative GDP series

Using the new price data, I compute the GDP at constant 2005 international (PPP) dollars using:

$$y_t = \frac{c_t^{NA}}{PPP_{c,05}} + \frac{g_t^{NA}}{PPP_{g,05}} + \frac{inv_t^{NA}}{PPP_{i,05}} + \frac{x_t^{NA}}{PPP_{gdp,05}} - \frac{im_t^{NA}}{PPP_{gdp,05}}$$

where y_t is GDP, c_t is consumption, g_t is government expenditure, inv_t is investment, x_t are the exports and im_t are the imports at time t. The superscript NA means that the data are taken from the national accounts tables underlying PWT6.[1] All of these are absolute values in national currencies, adjusted for inflation.[2] Note that the methodology for inflation adjustments can differ between countries, as these are normally carried out by the national statistical offices. The expenditure category specific PPPs are not explicitly reported in PWT7, so I compute them as

$$PPP_{i,05} = (p_{i,05}^{PWT7}/100) * XRAT_{05}^{PWT6} \tag{B.1}$$

where the $i = c, g, inv, x, im$, with meanings identical to above. PWT6 uses a country's overall price level of domestic absorption (a weighted average of p_c, p_i and p_g) for exports and imports, meaning here that $p_x = p_{im} = p$. $XRAT_{05}^{PWT6}$ is the 2005 exchange rate as reported in PWT6; I do not use the exchange rates underlying PWT7 in order to isolate the effect of the prices. The method used to construct the GDP series based on PWT7 NA data and PWT6 prices is identical, with the only exception that equation B.1 becomes obsolete as PPPs are provided in PWT6.[3]

B.1.2 Visual inspection for complete sample

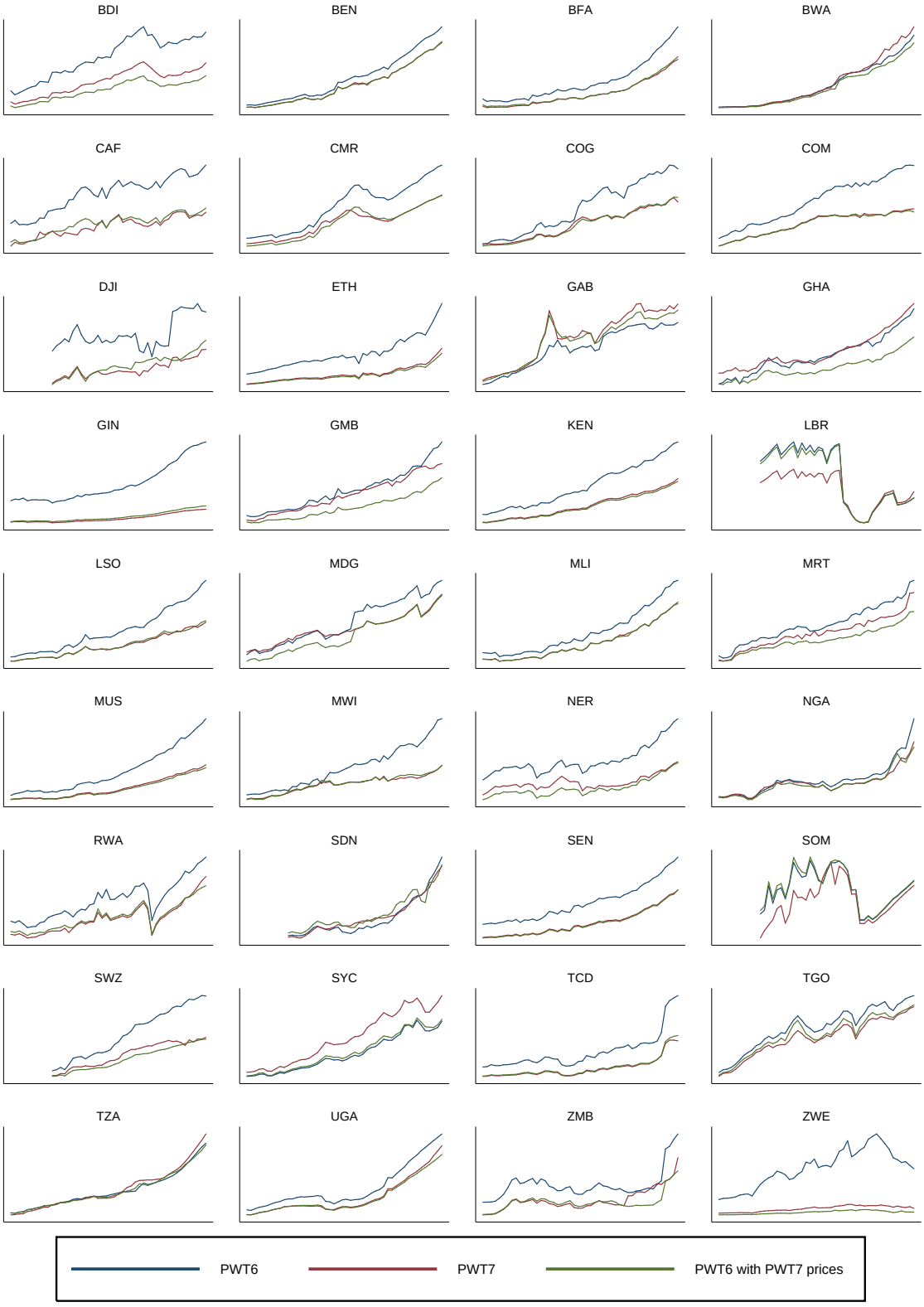
Figure B.1 reproduces figure 3.3 for the entire sample of 36 Sub-Saharan African countries. In almost all countries, the PWT6 series compiled using PWT7 prices matches PWT7 significantly better than the original PWT6 series. Besides Gabon, which is discussed in the main body of the paper, this is particularly pronounced

[1] Provided by the authors at `http://www.rug.nl/research/ggdc/data/pwt`.

[2] That is, they are in real terms; I avoid using the term in order to avoid confusion, as PWT typically use it to describe PPP adjusted values (made comparable across countries rather than time).

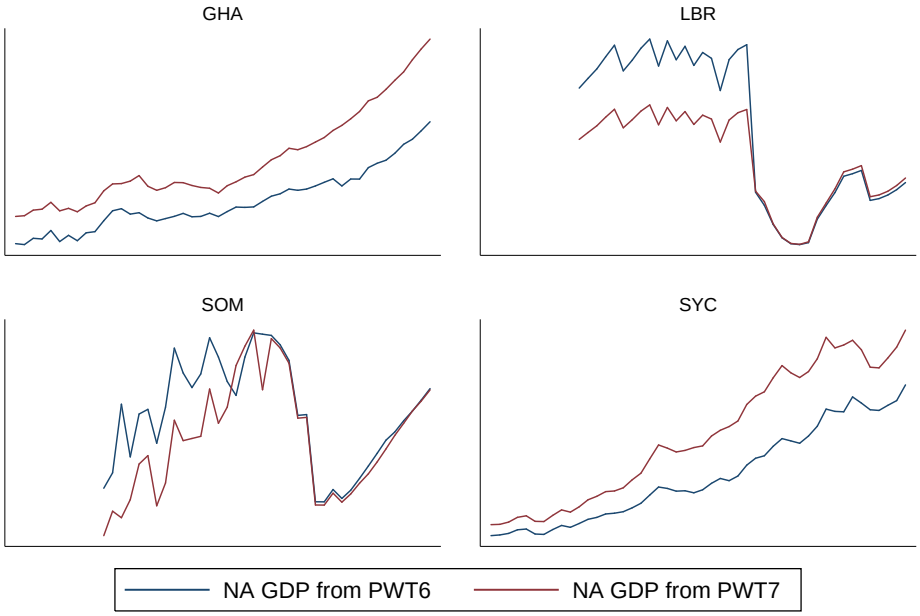
[3] Stata do-files are available from the author upon request

Figure B.1: GDP for 36 countries, 1960-2007



in the cases of Djibouti and Zimbabwe. In these countries, the two versions of the Penn World Table report strikingly different dynamics, judging from visual

Figure B.2: GDP from underlying NA data, 1960-2007



inspection of the graphs. After adjusting the prices, the growth patterns are largely reconciled, again highlighting the importance of the relative prices attributed to the expenditure shares. Indeed, in Djibouti, the estimated price of consumption has almost doubled between the versions, while that of investment has been more than halved. For Zimbabwe, 2005 price figures were revised upwards by a factor of up to 10 or 20, while the exchange rate has been lowered by a factor of 22 between the two vintages; in fact, the PWT7 documentation mentions the country as a problematic case in terms of price collection because of its high inflation rates (Heston et al., 2012).

In a few countries however, namely Ghana, Liberia, Somalia, and Seychelles, the new prices have little discernible impact on the GDP series. While in all of them there clearly is substantial divergence between the datasets, the relative prices of their GDP components have remained relatively stable. Instead, their underlying National Accounts data has been strongly revised: Figure B.2 juxtaposes the constant price GDP series as reported in the NA datasets underlying PWT6 and PWT7 respectively. These are the figures as reported by the statistical offices, prior to any PPP adjustments or other alterations stemming from the PWT methodology. Comparing these to the corresponding plots on figure B.1, it is apparent that the patterns of divergence are very similar.

B.2 Consistency by country

Table B.1: Consistent coefficients by country

Country	*PWT6*	*PWT7*	*PWT8*	*WDI*	*Consistent*
BDI	4	4	2	-	10/12
BEN	**4**	**2**	**2**	**2**	**10/16**
BFA	**4**	**4**	**4**	**4**	**16/16**
BWA	1	0	0	-	1/12
CAF	4	4	0	-	8/12
CMR	**4**	**4**	**4**	**4**	**16/16**
COG	**4**	**4**	**3**	**3**	**14/16**
COM	4	0	3	-	7/12
DJI	2	2	0	-	4/12
ETH	4	1	0	-	5/12
GAB	**4**	**4**	**4**	**4**	**16/16**
GHA	4	2	1	-	7/12
GIN	4	3	4	-	11/12
GMB	4	1	3	-	8/12
KEN	**4**	**4**	**4**	**4**	**16/16**
LBR	4	4	0	-	8/12
LSO	**4**	**0**	**0**	**0**	**4/16**
MDG	**4**	**4**	**3**	**3**	**14/16**
MLI	3	3	3	-	9/12
MRT	**1**	**0**	**1**	**2**	**4/16**
MUS	4	3	3	-	10/12
MWI	4	3	3	-	10/12
NER	4	3	4	-	11/12
NGA	4	2	2	-	8/12
RWA	**3**	**2**	**3**	**2**	**10/16**
SDN	-*	2	2	3	10/16
SEN	**3**	**4**	**4**	**1**	**12/16**
SOM	4	3	0	-	7/12
SWZ	4	2	2	-	8/12
SYC	3	4	0	-	7/12
TCD	4	4	4	-	12/12
TGO	**2**	**2**	**2**	**1**	**7/16**
TZA	3	1	1	-	5/12
UGA	4	1	2	-	7/12
ZMB	3	2	2	-	7/12
ZWE	4	1	1	-	6/12

The table reports the number of coefficients that are consistent with those obtained by JMT under the prior of aid effectiveness, taking into consideration the first and second best choice of rank, using their exact models. The last column reports the total sum of consistent coefficients within each country across all datasets, followed by the number of estimated coefficients. Countries included in the re-specification exercise in section 3.5.3 are in bold characters.

* JMT use WDI data only in the case of Sudan for reasons of data availability, which is why the corresponding PWT6 figure is left out.

B.3 Determinants of consistency

In order to make an assessment of the stability of one's results with respect to different datasets without having to run an entire analysis every time, it would be beneficial to have a metric or a set of metrics that have a reasonable predictive power concerning the stability of the results. It is however difficult to discern an obvious relationship between the consistency of the results and the properties of the underlying data. Figures 3.1 and 3.2 in section 3.3.2 can give a first sense of this. Both the countries with the most stable results in the subsequent analysis (top panels) and those with the least stable results (bottom panels) seem to diverge to comparable extents. This is true both for GDP levels and, as illustrated with shares of investment in figure 3.2, shares of GDP.

Figure B.3 abstracts from differences in levels and plots the growth rates of the same countries over the sample period. Again, it is hard to discern any clear patterns, as both sets of countries (the four with the most stable result at the top, the least stable ones at the bottom) exhibit quite volatile and divergent growth rates. Besides the visual inspection of the graphs, I tried to relate a number of measures to the stability of the results. None of them turned out to be a good predictor. These included:

- Variance (alternatively standard deviation) around the average between the series:

$$\sigma^2_{Y,t} = \frac{1}{4} \sum_{i \in PWT6,PWT7,PWT8,WDI} (Y_{i,t} - \frac{\sum_{i \in PWT6,PWT7,PWT8,WDI} Y_{i,t}}{4})^2$$

- As this measure mechanically increases as GDP increases, the same with GDPs normalised around their mean each period (deviation from mean in percent),

$$Y^{norm}_{i,t} = \frac{Y_{i,t}}{\frac{1}{4}\sum_{i \in PWT6,PWT7,PWT8,WDI} Y_{i,t}} * 100$$

and

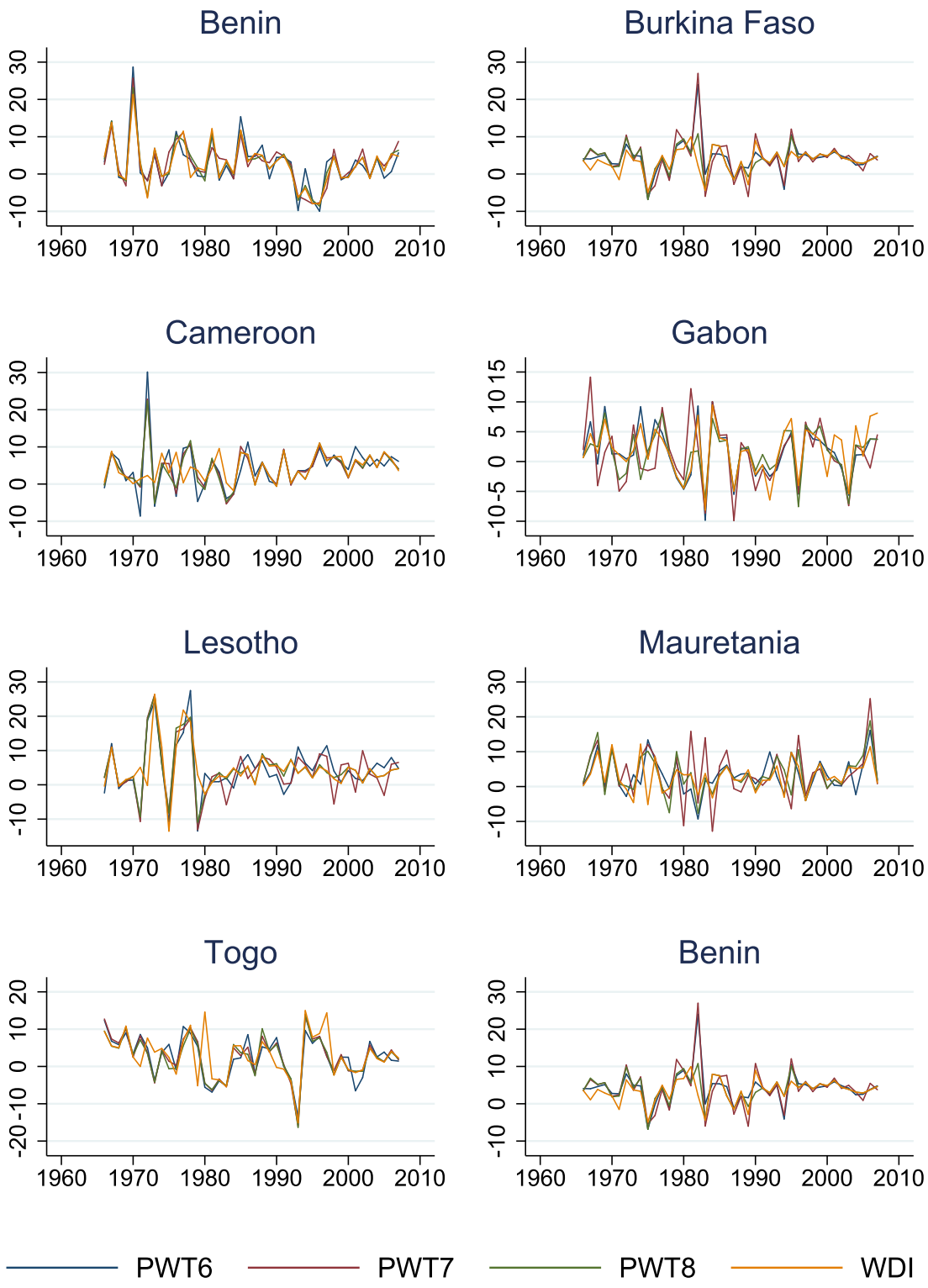
$$\sigma^2_{Y,i,t} = \frac{1}{4} \sum_{i \in PWT6,PWT7,PWT8,WDI} (Y^{norm}_{i,t} - 100)^2$$

.

- Variance and standard deviation across series (as above) in terms of growth rates

- Variance and standard deviation as described above for individual expenditure shares.

Figure B.3: GDP growth rates from four sources

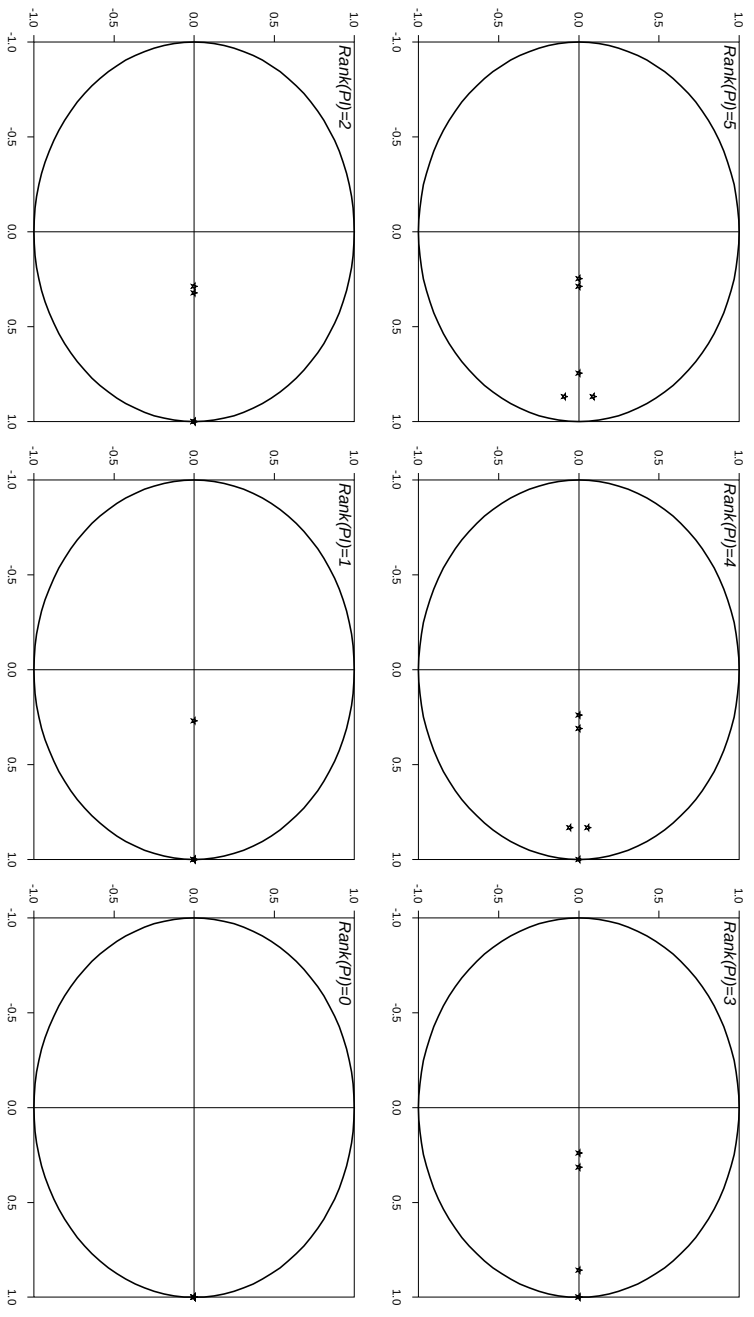


- Statistical capacity as reported by the World Bank. While this has some power in predicting the consistency of individual series over time, it does

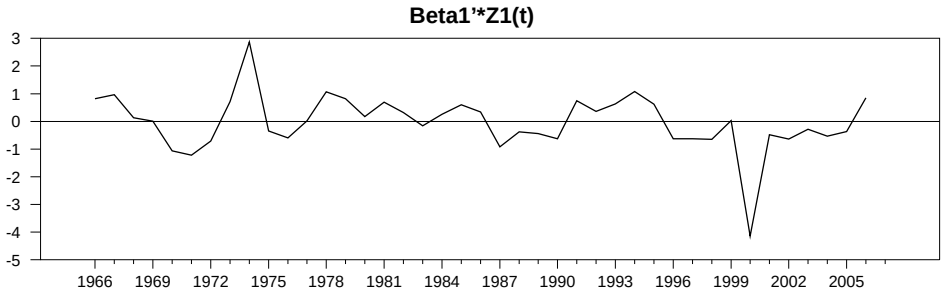
not predict the consistency of the results. This is in line with the findings from the previously mentioned measures.

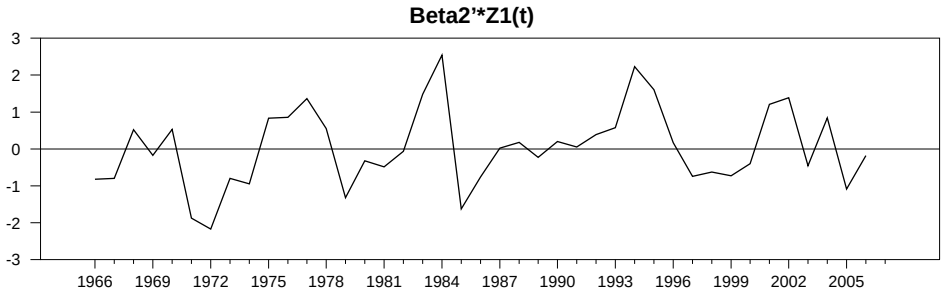
The strongest contributing factor I could identify was not in the data, but in the results. In the present sample, insignificant results were much more likely to persist throughout the datasets: Within the 8 countries that had no single significant coefficient in the first place (PWT6), 80% of the replicated coefficients were also insignificant in the replications. Within all other countries, 50% kept their sign and significance.

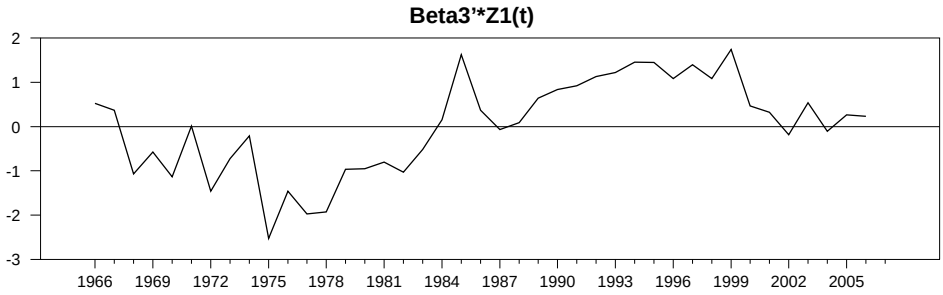
B.4 Roots of the companion matrix (Burkina Faso, PWT6)

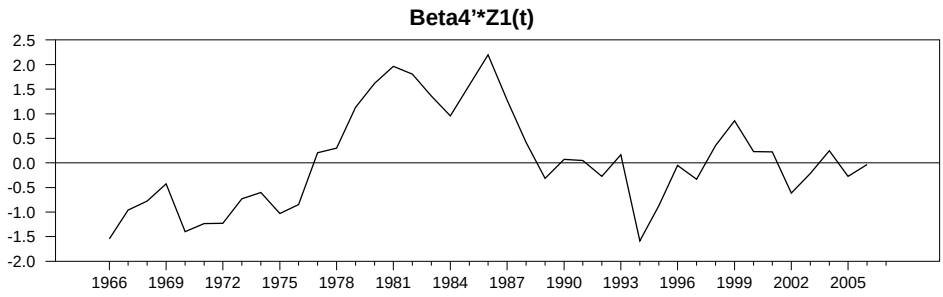


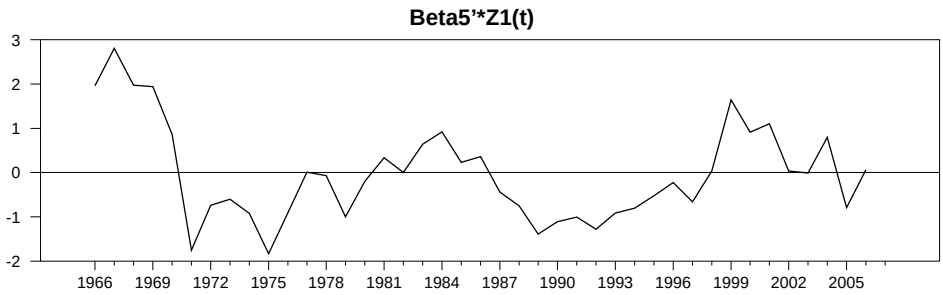
B.5 Graphs of the cointegrating relations (Burkina Faso, PWT6)











B.6 T-ratios of the C_{21}-Matrix

Table B.2: t-ratios of the best and second best choice of rank

	PWT6		**PWT7**		**PWT8**		**WDI**	
	r^*	r'	r^*	r'	r^*	r'	r^*	r'
Burkina Faso	-1.03	-1.08	-1.08	-0.35	-1.03	-1.08	0.65	-0.24
	1.15	-0.13	0.68	-4.00	1.15	-0.13	-3.81	-1.52
	-0.26	-0.43	-0.29	1.78	-0.26	-0.43	2.36	0.72
	0.16	0.89	-0.41	1.70	0.16	0.89	-2.37	1.68
Cameroon	-0.55	-1.19	-0.91	-0.93	-0.89	-0.68	-1.24	-3.01
	0.62	-0.30	-0.22	-1.12	-0.21	-1.20	0.09	0.12
	-0.15	-0.59	-0.82	-0.26	-0.80	-0.03	-0.28	-2.07
	-0.45	-3.85	-0.61	-0.54	-0.27	0.73	-1.48	-0.78
Gabon	0.23	1.02	-0.41	0.14	1.89	1.81	-0.50	0.65
	-0.01	1.02	-0.41	0.19	1.88	1.81	-0.59	0.57
	-2.80	1.02	-0.42	-0.02	0.68	1.22	-0.88	-0.69
	1.18	1.02	-0.42	0.12	1.32	1.55	-0.86	0.77
Kenya	4.03	1.44	2.54	2.84	2.69	3.02	4.75	2.23
	3.38	3.09	2.01	1.88	2.46	1.82	4.53	2.23
	3.83	1.80	2.59	2.96	2.72	3.04	4.41	2.23
	0.67	1.05	-1.13	-0.54	-0.18	-0.15	1.69	-2.23
Congo, Rep.	-0.09	1.36	0.36	-0.55	-0.36	-0.32	-0.13	-0.10
	1.16	1.19	0.84	1.40	0.31	1.09	-0.13	-0.12
	-0.48	0.17	0.84	-0.94	-0.24	-0.91	-0.13	-0.06
	0.93	1.21	0.68	0.52	-0.14	0.34	-0.13	-0.11
Madagascar	2.47	0.66	-1.37	0.00	0.20	1.97	-0.75	-0.80
	2.47	-0.63	-1.37	-2.16	-0.71	16.62	0.70	0.42
	2.47	4.90	1.37	3.05	2.18	2.04	-1.19	-0.46
	-2.47	-6.48	-1.37	-0.98	-0.54	-1.07	0.90	-1.06
Rwanda	1.10	-5.09	1.65	3.23	2.97	3.04	4.33	4.27
	2.10	-3.58	1.84	2.29	1.84	2.17	4.07	1.64
	3.42	1.08	2.40	1.59	2.59	1.46	3.14	3.57
	-1.84	-0.95	-0.23	0.21	-0.07	0.27	1.10	1.59
Senegal	2.01	4.16	1.54	3.05	1.30	4.01	0.47	-0.16
	0.44	2.74	-0.93	0.97	0.58	3.09	1.30	1.17
	2.06	2.25	2.65	3.42	1.98	2.87	3.39	3.39
	2.19	3.99	2.86	3.56	1.93	3.58	1.10	1.74
Benin	1.15	-2.37	-0.63	6.70	-0.89	2.31	-0.75	-1.81
	3.05	-2.37	-4.26	12.78	0.15	9.97	0.70	-4.15
	3.74	2.37	4.19	4.58	2.48	2.43	-1.19	0.76
	-2.84	-2.37	-1.03	6.43	-2.10	-4.38	0.90	2.54

Table B.2: Table B.2 (continued)

	PWT6		**PWT7**		**PWT8**		**WDI**	
Lesotho	2.34	5.11	-5.94	-11.51	-0.12	0.83	1.09	-7.04
	2.22	2.21	1.56	-1.57	0.93	3.72	1.10	-2.94
	3.28	3.46	-6.90	-6.15	-2.55	-7.35	1.47	-2.50
	2.27	0.02	-1.08	-2.43	-7.02	0.49	-1.95	-1.00
Mauritania	-1.30	0.00	1.85	-3.64	-0.93	-2.97	-5.20	-0.31
	2.17	-0.26	0.18	-2.91	0.75	-0.02	-0.75	-26.12
	0.58	0.06	4.51	-0.51	2.93	4.80	0.27	0.91
	2.84	0.08	0.97	3.06	0.90	0.41	2.38	-1.78
Togo	3.64	1.99	11.16	2.47	11.63	1.49	4.22	3.66
	-0.94	1.73	-3.62	-4.98	-3.55	-4.81	2.29	0.81
	4.88	-1.41	5.83	1.63	4.26	1.04	6.45	6.61
	2.22	3.73	1.11	1.13	0.37	0.58	0.88	0.68

Notes: r^* reports the t-ratios for the variables in the second column under the preferred rank specification, r' those obtained under the second best choice of rank.

Source: Author's calculations.

Appendix C

Appendix Chapter 4

C.1 Prior values of contemporaneous matrix H

Table C.1: Prior values of contemporaneous matrix H

		Shock					
	CPI type	Fed	Copper	Money	Output	Exch. rate	Price
Fed funds rate	All Non-Food Food	1	0	0	0	0	0
Copper	All Non-Food Food	8.52 9.42 8.53	1	0	0	0	0
M2	All Non-Food Food	-2.65 -2.53 -2.47	-0.04 -0.04 -0.04	1	0	0	0
Output Gap	All Non-Food Food	-0.49 0.05 -0.64	**0.03** **0.03** **0.04**	-0.02 -0.03 0	1	0	0
Exchange rate	All Non-Food Food	**-1.77** **-2.98** **-1.67**	0.21 0.22 0.18	**-0.76** **-0.79** **-0.77**	0.28 0.16 0.34	1	**-0.85** **-1.3** **-0.47**
Inflation	All Non-Food Food	-0.11 0.25 -0.55	0.02 0.02 0.03	-0.07 -0.04 -0.11	0.04 0.05 0.08	**-0.09** **-0.11** **-0.09**	1

Notes: Priors reported in bold are subject to a sign restriction. The respective sign can be inferred from the reported value of the prior. The priors are obtained from just-identified Cholesky decompositions as described in section 3.1.

C.2 Posterior values of contemporaneous matrix H

Table C.2: Posterior values of contemporaneous matrix H

		Shock					
	CPI type	Fed	Copper	Money	Output	Exch. rate	Price
Fed funds rate	All	1	0	0	0	0	0
	Non-Food						
	Food						
Copper	All	8.19	1	0	0	0	0
	Non-Food	9.51					
	Food	9.19					
M2	All	-3.07	-0.02	1	0	0	0
	Non-Food	-2.19	-0.09				
	Food	-2.48	-0.05				
Output Gap	All	-0.03	**0.01**	-0.19	1	0	0
	Non-Food	0.41	**0.00**	-0.08			
	Food	-0.28	**0.01**	-0.14			
Exchange rate	All	**-2.45**	0.18	**-0.68**	-0.09	1	**-0.82**
	Non-Food	**-2.53**	0.16	**-0.86**	-0.03		**-0.65**
	Food	**-1.80**	0.17	**-0.67**	-0.34		**-0.97**
Inflation	All	-0.22	0.02	0.02	0.11	**-0.02**	1
	Non-Food	0.23	0.01	0.00	0.02	**-0.05**	
	Food	-0.25	0.02	0.03	0.03	**-0.03**	

Notes: Posteriors reported in bold were subject to a sign restriction. The respective sign can be inferred from the reported value of the posterior.

C.3 Augmented Dickey Fuller tests

Table C.3: Augmented Dickey-Fuller tests

		Intercept	Trend	t-stat.	5% CV	p-value
Copper price	Level	yes	no	-1.11	-2.9	0.71
	1st Diff	no	no	-3.71	-1.95	0
Fed funds rate	Level	no	no	-1.5	-1.95	0.13
	1st Diff	no	no	-2.97	-1.95	0
Money	Level	yes	yes	3.47	-3.47	1
	1st Diff	yes	no	-3.2	-2.9	0.02
Output gap	Level	no	no	-3.98	-1.95	0
	1st Diff	no	no	-6.34	-1.95	0
Exchange rate	Level	yes	yes	-1.77	-3.47	0.71
	1st Diff	yes	no	-4.37	-2.9	0
CPI	Level	yes	yes	-2.53	-3.47	0.31
	1st Diff	yes	yes	-5.29	-3.47	0

Notes: All tests have been carried out with trends and intercept where the graphs in section 4 suggested their relevance, and removed when they turned out insignificant in the Dickey Fuller regression. The tests in first differences were carried out with the log differenced series as this corresponds to the series eventually included in out estimation for most variables. The exception is the output gap, where logs cannot be taken because of negative values. All results refer to a test with 4 lags of the variable for consistency with the main analysis.

Bibliography

Addison, T. and Baliamoune-Lutz, M. (2017). Aid, the real exchange rate and why policy matters: The cases of Morocco and Tunisia. *The Journal of Development Studies*, 53(7):1104–1121.

Alesina, A. F., Michalopoulos, S., and Papaioannou, E. (2012). Ethnic inequality. *NBER Working Paper*, (18512).

Aron, J., Macdonald, R., and Muellbauer, J. (2014). Exchange rate pass-through in developing and emerging markets: A survey of conceptual, methodological and policy issues, and selected empirical findings. *Journal of Development Studies*, 50(1):101–143.

Atherton, P., Appleton, S., and Bleaney, M. (2011). Growth regressions and data revisions in Penn World Tables. *Journal of Economic Studies*, 38(3):301–312.

Baldini, A., Benes, J., Berg, A., Dao, M. C., and Portillo, R. (2012). Monetary policy in low income countries in the face of the global crisis: The case of Zambia. *IMF Working Paper*, (19/94).

Baumeister, C. and Hamilton, J. D. (2015). Sign restrictions, structural vector autoregressions, and useful prior information. *Econometrica*, 83(5):1963–1999.

Berg, A., Papageorgiou, C., Pattillo, C. A., Schindler, M., Spatafora, N., and Weisfeld, H. (2011). Global shocks and their impact on low-income countries: lessons from the global financial crisis. *IMF Working Paper*, (11/27).

Bergholt, D. (2014). Monetary policy in oil exporting economies. *CAMP Working Paper Series*, (5/2014).

Bernanke, B. S. (1986). Alternative explanations of the money-income correlation. In *Carnegie-Rochester conference series on public policy*, volume 25, pages 49–99. Elsevier.

Bickenbach, F., Bode, E., Nunnenkamp, P., and Söder, M. (2016). Night lights and regional GDP. *Review of World Economics*, 152(2):425–447.

Blanchard, O. J. and Watson, M. W. (1986). Are business cycles all alike? In Gordon, R. J., editor, *The American business cycle: Continuity and change*, pages 123–180. University of Chicago Press.

Bodart, V., Candelon, B., and Carpantier, J.-F. (2015). Real exchanges rates, commodity prices and structural factors in developing countries. *Journal of International Money and Finance*, 51:264–284.

Bova, E. (2009). Is the Zambian Kwacha a commodity currency? The co-movement between copper prices and the exchange rate. *NCCR Trade Regulations SWISS National Centre of Competence in Research Working Paper*, (2009/11).

Breton, T. R. (2012). Penn World Table 7.0: Are the data flawed? *Economics Letters*, 117(1):208–210.

Breton, T. R. (2015). Changes in the effect of capital and TFP on output in Penn World Tables 7 and 8: Improvement or error? *Center for Research in Economics and Finance (CIEF), Working Papers, https://repository.eafit.edu.co/handle/10784/388*, (15-03).

Breton, T. R. and García, J. J. (2016). ICP 2005 construction prices: Are they underestimated in developing countries? *Review of Income and Wealth*, (62):380–393.

Bwire, T., Lloyd, T., and Morrissey, O. (2017). Fiscal reforms and the fiscal effects of aid in Uganda. *The Journal of Development Studies*, 53(7):1019–1036.

Calvo, G. A. and Mishkin, F. S. (2003). The mirage of exchange rate regimes for emerging market countries. *Journal of Economic Perspectives*, 17(4):99–118.

Cashin, P., Céspedes, L. F., and Sahay, R. (2004). Commodity currencies and the real exchange rate. *Journal of Development Economics*, 75(1):239–268.

Ca'Zorzi, M., Hahn, E., and Sánchez, M. (2007). Exchange rate pass-through in emerging markets. *ECB Working Paper Series*, (739).

Charnavoki, V. and Dolado, J. J. (2014). The effects of global shocks on small commodity-exporting economies: lessons from canada. *American Economic Journal: Macroeconomics*, 6(2):207–37.

Chen, X. and Nordhaus, W. D. (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences*, 108(21):8589–8594.

Chipili, J. M. (2016). Copper price and exchange rate dynamics in zambia. *Journal of International Development*, 28(6):876–886.

Choudhri, E. U. and Hakura, D. S. (2006). Exchange rate pass-through to domestic prices: does the inflationary environment matter? *Journal of International Money and Finance*, 25(4):614–639.

Ciccone, A. and Jarociński, M. (2010). Determinants of economic growth: will data tell? *American Economic Journal: Macroeconomics*, 2(4):222–246.

Collier, P. (2008). *The bottom billion: Why the poorest countries are failing and what can be done about it*. Oxford University Press, UK.

Collier, P. and O'Connell, S. (2009). Opportunities and choices. In Ndulu, B. J., O'Connell, S. A., Bates, R. H., Collier, P., and Soludo, C. C., editors, *The Political Economy of Economic Growth in Africa, 1960–2000*, volume 1, pages 76–136. Cambridge University Press.

Croft, T. A. (1978). Nighttime images of the earth from space. *Scientific American*, 239(1):86–101.

Dai, Z., Hu, Y., and Zhao, G. (2017). The suitability of different nighttime light data for GDP estimation at different spatial scales and regional levels. *Sustainability*, 9(2):305.

Deaton, A. and Heston, A. (2010). Understanding PPPs and PPP-based national accounts. *American Economic Journal: Macroeconomics*, 2(4):1–35.

Deaton, A. and Tten, B. (2017). Trying to understand the PPPs in ICP 2011: Why are the results so different? *American Economic Journal: Macroeconomics*, 9(1):243–264.

Devarajan, S. (2013). Africa's statistical tragedy. *Review of Income and Wealth*, 59(S1):S9–S15.

Devereux, M. B. and Yetman, J. (2010). Price adjustment and exchange rate pass-through. *Journal of International Money and Finance*, 29(1):181–200.

Dezeure, R., Bühlmann, P., Meier, L., Meinshausen, N., et al. (2015). High-dimensional inference: Confidence intervals, p-values and r-software hdi. *Statistical science*, 30(4):533–558.

Doll, C. N., Muller, J.-P., and Morley, J. G. (2006). Mapping regional economic activity from night-time light satellite imagery. *Ecological Economics*, 57(1):75–92.

Doornik, J. A. (2009). Autometrics. In Castle, J. and Shepard, N., editors, *The Methodology and Practice of Econometrics*, chapter 4, pages 88–121. Oxford University Press.

Doornik, J. A. and Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, 70(1):927–939.

Drechsel, T. and Tenreyro, S. (2017). Commodity booms and busts in emerging economies. *Journal of International Economics*, In press.

Easterly, W. and Levine, R. (1997). Africa's growth tragedy: policies and ethnic divisions. *The Quarterly Journal of Economics*, pages 1203–1250.

Easterly, W., Levine, R., and Roodman, D. (2004). Aid, policies, and growth: Comment. *American Economic Review*, 94(3):774–780.

Ebener, S., Murray, C., Tandon, A., and Elvidge, C. C. (2005). From wealth to health: modelling the distribution of income per capita at the sub-national level using night-time light imagery. *International Journal of Health Geographics*, 4(1):5.

Eberhardt, M. and Teal, F. (2011). Econometrics for grumblers: a new look at the literature on cross-country growth empirics. *Journal of Economic Surveys*, 25(1):109–155.

Eichengreen, B. (2006). Can emerging markets float? Should they target inflation. In S, E. E., editor, *Monetary Integration and Dollarization: No Panacea*, chapter 8, pages 149–175.

Elvidge, C. D., Ziskin, D., Baugh, K. E., Tuttle, B. T., Ghosh, T., Pack, D. W., Erwin, E. H., and Zhizhin, M. (2009). A fifteen year record of global natural gas flaring derived from satellite data. *Energies*, 2(3):595–622.

Essers, D. (2013). Developing country vulnerability in light of the global financial crisis: Shock therapy? *Review of Development Finance*, 3(2):61–83.

Feenstra, R. C., Inklaar, R., and Timmer, M. (2013a). Comparing PWT 8.0 with pwt 7.1. *University of California, Davis and University of Groningen.*

Feenstra, R. C., Inklaar, R., and Timmer, M. (2013b). Penn World Table v. 8.0. *University of California, Davis and University of Groningen, http://www.rug.nl/research/ggdc/data/penn-world-table.*

Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2015). The next generation of the Penn World Table. *American Economic Review*, 105(10):3150–3182.

Forbes, K. J., Hjortsoe, I., and Nenova, T. (2015). The shocks matter: improving our estimates of exchange rate pass-through. *MIT Sloan Research Paper*, (5149-15).

Gebregziabher, F. (2014). The long-run macroeconomic effects of aid and disaggregated aid in Ethiopia. *Journal of International Development*, 26(4):520–540.

Ghosh, T., Anderson, S., Powell, R. L., Sutton, P. C., and Elvidge, C. D. (2009). Estimation of Mexico's informal economy and remittances using nighttime imagery. *Remote Sensing*, 1(3):418–444.

Ghosh, T., L Powell, R., D Elvidge, C., E Baugh, K., C Sutton, P., and Anderson, S. (2010). Shedding light on the global distribution of economic activity. *The Open Geography Journal*, 3(1).

Guha-Sapir, D., Below, R., and Hoyois, P. (2014). EM-DAT: International disaster database – `www.emdat.be`. *Université Catholique de Louvain, Brussels: Belgium.*

Hanushek, E. A. and Kimko, D. D. (2000). Schooling, labor-force quality, and the growth of nations. *American Economic Review*, 90(5):1184–1208.

Harttgen, K., Klasen, S., and Vollmer, S. (2013). An African growth miracle? Or: what do asset indices tell us about trends in economic performance? *Review of Income and Wealth*, 59(S1):S37–S61.

Hausmann, R., Pritchett, L., and Rodrik, D. (2005). Growth accelerations. *Journal of Economic Growth*, 10(4):303–329.

Henderson, J. V., Storeygard, A., and Weil, D. N. (2012). Measuring economic growth from outer space. *American Economic Review*, 102(2):994–1028.

Heston, A., Summers, R., and Aten, B. (2009). Penn World Table v. 6.3. *Center for International Comparisons of Production, Income and Prices (Philadelphia: University of Pennsylvania).*

Heston, A., Summers, R., and Aten, B. (2012). Penn World Table version 7.1. *Center for International Comparisons of Production, Income, and Prices at the University of Pennsylvania.*

Hodler, R. and Raschky, P. A. (2014). Regional favoritism. *The Quarterly Journal of Economics*, 129:995–1033.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Hoffmann, M. (2007). Fixed versus flexible exchange rates: Evidence from developing countries. *Economica*, 74(295):425–449.

Hoover, K. D., Johansen, S., and Juselius, K. (2008). Allowing the data to speak freely: The macroeconometrics of the cointegrated vector autoregression. *American Economic Review*, 98(2):251–255.

Hoover, K. D. and Perez, S. J. (2004). Truth and robustness in cross-country growth regressions. *Oxford Bulletin of Economics and Statistics*, 66(5):765–798.

Ilzetzki, E., Reinhart, C. M., and Rogoff, K. S. (2017). Exchange arrangements entering the 21st century: Which anchor will hold? *NBER Working Paper*, (w23134).

IMF (2016). Annual report on exchange arrangements and exchange restrictions 2016. Technical report, International Monetary Fund.

Inklaar, R. and Prasada Rao, D. (2017). Cross-country income levels over time: did the developing world suddenly become much richer? *American Economic Journal: Macroeconomics*, 9(1):265–290.

Jasova, M., Moessner, R., and Takats, E. (2016). Exchange rate pass-through: what has changed since the crisis? *BIS Working Papers*, (583).

Jerven, M. (2011). Users and producers of African income: Measuring the progress of African economies. *African Affairs*, 110(439):169–190.

Jerven, M. (2013a). Comparability of GDP estimates in sub-Saharan Africa: The effect of revisions in sources and methods since structural adjustment. *Review of Income and Wealth*, 59(S1):S16–S36.

Jerven, M. (2013b). For richer, for poorer: GDP revisions and Africa's statistical tragedy. *African Affairs*, 112(446):138–147.

Jerven, M. (2013c). *Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It.* Cornell University Press.

Jerven, M. (2014). African growth miracle or statistical tragedy? Interpreting trends in the data over the past two decades. *WIDER Working Paper 2014/114*.

Jerven, M. (2015). *Africa: why economists get it wrong.* Zed Books, London.

Jerven, M. (2016). Discrepancies: Why do GDP growth rates differ? *Review of Agrarian Studies*, 6(1):63–80.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2):231–254.

Johnson, S., Larson, W., Papageorgiou, C., and Subramanian, A. (2013). Is newer better? Penn World Table revisions and their impact on growth estimates. *Journal of Monetary Economics*, 60(2):255–274.

Johnston, D. and Abreu, A. (2016). The asset debates: How (not) to use asset indices to measure well-being and the middle class in Africa. *African Affairs*, 115(460):399–418.

Jones, B. F. and Olken, B. A. (2005). Do leaders matter? National leadership and growth since World War II. *The Quarterly Journal of Economics*, 120(3):835–864.

Juselius, K. (2006). *The cointegrated VAR model: methodology and applications*. Oxford University Press.

Juselius, K., Møller, N. F., and Tarp, F. (2014). The long-run impact of foreign aid in 36 African countries: Insights from multivariate time series analysis. *Oxford Bulletin of Economics and Statistics*, 76(2):153–184.

Juselius, K., Reshid, A., and Tarp, F. (2017). The real exchange rate, foreign aid and macroeconomic transmission mechanisms in Tanzania and Ghana. *The Journal of Development Studies*, 53(7):1075–1103.

Kenourgios, D. and Padhi, P. (2012). Emerging markets and financial crises: regional, global or isolated shocks? *Journal of Multinational Financial Management*, 22(1-2):24–38.

Keola, S., Andersson, M., and Hall, O. (2015). Monitoring economic development from space: using nighttime light and land cover data to measure economic growth. *World Development*, 66:322–334.

Levin, N. and Zhang, Q. (2017). A global analysis of factors controlling VIIRS nighttime light levels from densely populated areas. *Remote Sensing of Environment*, 190:366–382.

Lopez-Villavicencio, A., Mignon, V., et al. (2016). Exchange rate pass-through in emerging countries: Do the inflation environment, monetary policy regime and institutional quality matter? *CEPII Working Paper*, 7.

MacKinnon, J. G. and White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29(3):305–325.

Magee, C. S. and Doces, J. A. (2015). Reconsidering regime type and growth: lies, dictatorships, and statistics. *International Studies Quarterly*, 59(2):223–237.

Marshall, M. G. and Jaggers, K. (2002). Polity IV project: Political regime characteristics and transitions, 1800-2002. *Center for Systemic Peace*.

Mascagni, G. and Timmis, E. (2017). The fiscal effects of aid in Ethiopia: Evidence from CVAR applications. *The Journal of Development Studies*, 53(7):1037–1056.

Michalopoulos, S. and Papaioannou, E. (2013). Pre-colonial ethnic institutions and contemporary African development. *Econometrica*, 81(1):113–152.

Morón, E. and Winkelried, D. (2005). Monetary policy rules for financially vulnerable economies. *Journal of Development Economics*, 76(1):23–51.

Muhanga, I., Zyuulu, I., and Adam, C. (2014). Monetary policy and the exchange rate in Zambia. In Christopher Adam, Paul Collier, M. G., editor, *Zambia: Building Prosperity from Resource Wealth*, pages 110–128. Oxford University Press.

Mullainathan, S. and Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106.

Ndulu, B. J., O'Connell, S. A., Bates, R. H., Collier, P., and Soludo, C. C., editors (2008). *The political economy of economic growth in Africa, 1960-2000*, volume 2. Cambridge University Press.

NNBS (2013). FAQ on the rebasing / re-benchmarking of Nigeria's gross domestic product (GDP). *Nigerian National Bureau of Statistics, https://tinyurl.com/yb2jukar*.

Obstfeld, M., Ostry, J. D., and Qureshi, M. S. (2017). A tie that binds: revisiting the trilemma in emerging market economies. *IMF Working Paper*, (17/130).

Osei, R., Morrissey, O., and Lloyd, T. (2005). The fiscal effects of aid in Ghana. *Journal of International Development*, 17(8):1037–1053.

Pamu, E. M. (2011). Does the copper price explain the deviation of the real exchange rate from the purchasing power parity equilibrium in zambia. *Issues on the Zambian Economy*, 1(6):38–44.

Pettersson, T. and Wallensteen, P. (2015). Armed conflicts, 1946–2014. *Journal of Peace Research*, 52(4):536–550.

Pinkovskiy, M. and Sala-i Martin, X. (2014a). Africa is on time. *Journal of Economic Growth*, 19(3):311–338.

Pinkovskiy, M. and Sala-i Martin, X. (2014b). Lights, camera,... income!: Estimating poverty using national accounts, survey means, and lights. *NBER Working Paper*, (19831).

Pinkovskiy, M. and Sala-i-Martin, X. (2016). Newer need not be better: Evaluating the Penn World Tables and the World Development Indicators using nighttime lights. *NBER Working Paper*, (22216).

Ponomareva, N. and Katayama, H. (2010). Does the version of the Penn World Tables matter? An analysis of the relationship between growth and volatility. *Canadian Journal of Economics/Revue canadienne d'économique*, 43(1):152–179.

Rahbek, A., Hansen, E., and Dennis, J. G. (2002). ARCH innovations and their impact on cointegration rank testing. *Centre for Analytical Finance Working Paper, University of Copenhagen*, (22:15).

Ram, R. and Ural, S. (2013). Comparison of GDP per capita data in Penn World Table and World Development Indicators. *Social Indicators Research*, pages 1–8.

Ramey, G. and Ramey, V. (1995). Cross-country evidence on the link between volatility and growth. *American Economic Review*, 85(5):1138–1151.

Ramey, V. A. (2016). Macroeconomic shocks and their propagation. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2, pages 71–162. Elsevier.

Rodrik, D. (2018). An African growth miracle? *Journal of African Economies*, 27(1):10–27.

Roodman, D. (2007). The anarchy of numbers: aid, development, and cross-country empirics. *The World Bank Economic Review*, 21(2):255–277.

Sala-i-Martin, X., Doppelhofer, G., and Miller, R. I. (2004). Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach. *American Economic Review*, 94(4):813–835.

Seely, J. C. (2005). The legacies of transition governments: post-transition dynamics in Benin and Togo. *Democratization*, 12(3):357–377.

Shambaugh, J. (2008). A new look at pass-through. *Journal of International Money and Finance*, 27(4):560–591.

Sheldon, K. E. and Penvenne, J. M. (2018). Mozambique. In *Encyclopædia Britannica*. Encyclopædia Britannica, inc.

Shenton, L. and Bowman, K. (1977). A bivariate model for the distribution of $\sqrt{b1}$ and b2. *Journal of the American Statistical Association*, 72(357):206–211.

Siklos, P. L. (2018). Boom-and-bust cycles in emerging markets: How important is the exchange rate? *Journal of Macroeconomics*.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1):1–48.

Stock, J. H. and Watson, M. W. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2, pages 415–525. Elsevier.

Summers, R. and Heston, A. (1991). The Penn World Table (mark 5): an expanded set of international comparisons, 1950–1988. *The Quarterly Journal of Economics*, 106(2):327–368.

Sutton, P. C. and Costanza, R. (2002). Global estimates of market and non-market values derived from nighttime satellite imagery, land cover, and ecosystem service valuation. *Ecological Economics*, 41(3):509–527.

Sutton, P. C., Elvidge, C. D., Ghosh, T., et al. (2007). Estimation of gross domestic product at sub-national scales using nighttime satellite imagery. *International Journal of Ecological Economics & Statistics*, 8(S07):5–21.

Tanaka, K. and Keola, S. (2017). Shedding light on the shadow economy: a nighttime light approach. *The Journal of Development Studies*, 53(1):32–48.

The World Bank (2015). World Development Indicators, retrieved 12 March, 2015.

Thorbecke, E. (2015). The anatomy of growth and development in sub saharan africa. In McKay, A. and Thorbecke, E., editors, *Economic Growth and Poverty Reduction in Sub-Saharan Africa*, chapter 2, pages 20–50. Oxford University Press, Oxford, UK.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure. *Journal of Monetary Economics*, 52(2):381–419.

Ul Haque, N., Pesaran, M. H., and Sharma, S. (1999). Neglected heterogeneity and dynamics in cross-country savings regressions. *IMF Working Paper*, 99/128.

Van de Sijpe, N. (2013). Is foreign aid fungible? Evidence from the education and health sectors. *The World Bank Economic Review*, 27(2):320–356.

Venables, A. J. (2016). Using natural resources for development: why has it proven so difficult? *Journal of Economic Perspectives*, 30(1):161–84.

Wu, J., Wang, Z., Li, W., and Peng, J. (2013). Exploring factors affecting the relationship between light consumption and GDP based on DMSP/OLS nighttime satellite imagery. *Remote Sensing of Environment*, 134:111–119.

Young, A. (2012). The African growth miracle. *Journal of Political Economy*, 120(4).

Zgambo, P. (2015). Exchange rate pass-through to domestic prices in Zambia. *The BoZ Reader*, 2(1):28–55.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.