Dissertation

Three Essays on Growth Econometrics

Abdul Ahad Tariq

This document was submitted as a dissertation in December 2016 in partial fulfillment of the requirements of the doctoral degree in public policy analysis at the Pardee RAND Graduate School. The faculty committee that supervised and approved the dissertation consisted of James Hosek (Chair), Titus Galama, and Krishna Kumar.
Abstract

There are large discrepancies between different panel datasets of cross-country estimates of gross domestic product (GDP). But which set of numbers is most accurate? In the first essay, I develop a general statistical framework for answering this question. I derive two estimators for assessing the relative accuracy of GDP data, and perform Monte Carlo simulations to compare them against existing approaches. Both estimators show that a new panel dataset of GDP from the University of Queensland (UQICD) is generally more accurate than the World Development Indicators (WDI) and the Penn World Tables (PWT), the two most widely-used datasets for growth economics research and policy.

In the second essay, I characterize the differences between several versions of WDI, and PWT, and UQICD. I find that these datasets disagree on the magnitude and direction of growth in a large number of countries. The differences appear to be driven by country geography, economic size and frequency of data collection for purchasing power parity estimates. I quantify the policy impact of GDP differences by simulating counterfactual predictions of World Bank aid allocation and find that development assistance can vary substantially due to changes in underlying income measures. Nevertheless, for the vast majority of the cases, aid allocation is relatively inelastic with respect to changes in GDP data.

The third essay empirically tests Tobin’s Q-theory of investment, a foundational model for modern macroeconomics and growth theory: it is shown that extant measures of average Q suffer from measurement error and a new measure of Q is proposed which adds close to 15 percentage points of explanatory power, more than twice as much as existing measures of Q. From theory, a cash flow-based measure of Q is derived which performs as well as extant Q. Lastly, an empirical measure for marginal Q is derived, but its explanatory power is found to be weak. These results provide support for the neoclassical model of firm investment.
Acknowledgments

This dissertation would not have been possible without the guidance of my wonderful advisors: James Hosek, Titus Galama, Krishna Kumar, and Romain Wacziarg who served as my outside reader.

Jim, who chaired my committee, is a remarkable mentor who combines a formidable intellect with utter warmth and kindness. Under his guidance, there was no roadblock I could not overcome, and I feel incredibly lucky that he agreed to take me on as his student.

I have been working with Titus since my very first weeks at RAND and he taught me what I consider to be the most important lesson in my education: to be intellectually fearless, ask big questions, and push forward, despite doubts and uncertainty. It was an immense privilege to work with and learn from him. I am also grateful for the generous funding he provided so I could explore my research interests in growth econometrics.

Krishna’s guidance was invaluable in writing this dissertation. He always pushed me to ensure that my research had real policy impact, and discussions with him greatly improved the quality of my work.

Finally, meeting and learning from Romain – one of the world’s foremost experts on economic growth – is one of the great fortunes of my graduate career. He selflessly donated many hours of his time to guide me in my research, and provided crucial insights and suggestions for analyses in the first two papers. I cannot thank him enough.

Two other individuals also deserve acknowledgment: D.S. Prasada Rao made the University of Queensland dataset available to me, guided me during the earliest stages of the research and pointed me to relevant literature in the field against which to compare my work. I also benefited tremendously from my discussions with David Powell who has an astounding knack for econometrics.

I want to thank the PRGS administration for making the school feel like a family. I am especially grateful to both Susan Marquis and Rachel Swanger, who have always been in my corner, and to Mary Parker and Kristina Wallace whose ferocious competence makes everything run seamlessly. This dissertation was also made possible through the generosity of Pardee RAND contributors via the Eugene and Maxine Rosenfeld Dissertation Award and the Frederick S. Pardee Dissertation Award for Global Human Progress.
My colleagues made PRGS a wonderful experience: Eric Warner and Kun Gu were the best roommates I could ask for; Prodyumna Goutam, Ujwal Kharel, Michael McGee, Dan Waxman, and Shmuel Abramzon kept life at RAND at once light-hearted and stimulating; Dan Han was a terrific sounding board who patiently suffered through my verbal meanderings; Nono Ayivi-Guedeoussou taught me much about life and our conversations will remain with me forever.

Kfir Mordechay and Yosi Coral ensured that life outside RAND was always hilarious and interesting. I am grateful for their friendship.

Finally, I want to thank my mother, father and brother. Everything I have achieved in life has been possible only because of their sacrifices, their love and their unwavering support. This dissertation is dedicated to them.
Financial Support

Research reported in this publication was supported by the National Institute on Aging of the National Institutes of Health under Award Number R01AG037398. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
# Contents

Abstract ........................................................................................................................................ i
Acknowledgements .................................................................................................................. ii
Financial support ..................................................................................................................... iv

Let There Be Light: Assessing the Accuracy of GDP Estimates Using Night-time Luminosity .............................................................................................................................. 1

1. Introduction ........................................................................................................................................... 2

2. Theory ..................................................................................................................................................... 6
   3.1 Minimum mean-squared prediction error weights (MSPE) ......................................................... 7
   3.2 Minimum variance weights (MVAR) ................................................................................................. 11
   3.2 Some properties of relative weights ................................................................................................. 13
   3.2 Biasedness of MSPE ......................................................................................................................... 14
   3.2 Biasedness of MVAR ......................................................................................................................... 14
   3.2 Biasedness of Pinkovskiy and Sala-i-Martin (PS) ............................................................................... 15
   3.2 Equivalence of MVAR and PS ......................................................................................................... 15
   3.2 Relationship between MVAR and MSPE ......................................................................................... 16

3. Simulation Studies .................................................................................................................................. 17
   3.1 Setup ............................................................................................................................................... 17
   3.2 Sampling .......................................................................................................................................... 19
   3.3 Results ............................................................................................................................................. 20

4. Application to GDP data .................................................................................................................... 25
   4.1 Data .................................................................................................................................................. 25
      4.1.1 ICP ........................................................................................................................................... 25
      4.1.1 WDI ......................................................................................................................................... 26
      4.1.1 PWT ......................................................................................................................................... 27
      4.1.1 UQICD .................................................................................................................................... 28
      4.1.1 Night-lights ............................................................................................................................... 28
   4.1 Results ............................................................................................................................................. 29
      4.1.1 WDI 2015 vs. WDI 2013 ......................................................................................................... 29
      4.1.1 PWT 9.0 vs. PWT 8.1 vs. PWT 7.1 ......................................................................................... 30
      4.1.1 PWT vs. WDI ............................................................................................................................ 32
      4.1.1 UQICD vs. all ........................................................................................................................... 32

5. Conclusion .......................................................................................................................................... 33

A. Appendix ........................................................................................................................................... 37

Small Differences: Discrepancies in GDP Statistics and Implications for Policy .... 41

1. Introduction ........................................................................................................................................... 42

2. Comparison of GDP panels ................................................................................................................ 44
   3.1 Inter-series differences in the direction of GDP growth ............................................................... 45
   3.2 Inter-series differences in GDP growth rates and levels ............................................................. 46
   3.2 The causes of inter-series differences in GDP growth rates ....................................................... 49

3. GDP uncertainty and International Development Assistance .................................................... 52
   3.1 IDA performance-based allocation ............................................................................................... 52
   3.2 Simulation of counterfactual aid allocations ............................................................................... 53
   3.2 Data ............................................................................................................................................... 55
<table>
<thead>
<tr>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2 Results</td>
</tr>
<tr>
<td>4. Conclusion</td>
</tr>
</tbody>
</table>

**In Defense of Neo-Classical Q**

- 1. Introduction | 70 |
- 2. Literature | 74 |
- 3. A theory of firm investment | 78 |
  - 3.1 Some reflection | 80 |
  - 3.2 First-order conditions | 80 |
  - 3.3 Marginal Q | 82 |
- 4. Empirical Strategy | 83 |
  - 4.1 Empirical relations | 83 |
    - 4.1.1 Tobin’s average Q | 83 |
    - 4.1.2 Marginal Q | 85 |
- 5. Data | 87 |
  - 5.1 Variables used in the analysis | 88 |
- 6. Regression results | 90 |
  - 6.1 Assessment of Tobin’s average Q and marginal Q | 91 |
  - 6.2 Measurement error | 94 |
  - 6.3 Summary | 95 |
- 7. Discussion and Conclusion | 96 |
  - A. Equivalence of $q_K(t)$ and marginal Q: general case | 110 |
  - B. Equivalence of marginal Q and average Q: special case | 113 |
  - C. Alternative expression for marginal Q | 115 |
  - D. Measurement error | 116 |
Let There Be Light: Assessing the Accuracy of GDP Estimates Using Night-time Luminosity

Abdul A. Tariq∗

Abstract

There are large discrepancies between different panel datasets of cross-country gross domestic products (GDP). But which set of numbers is most accurate? This paper develops a statistical approach for answering this question. I construct two estimators for computing a (statistically) optimal weighted-average of GDP using available panel datasets. The relative weight accorded to a given dataset is interpreted as being proportional to its relative accuracy. Monte Carlo simulations using plausible parameter values for the data-generating process of the global income distribution show that my preferred estimator has lower bias and mean-squared error than existing estimators for computing weighted averages of noisy data. These simulation results also reveal how data accuracy can govern the choice of the estimator. Applying these estimators to GDP data, I find that: (1) a new panel dataset of GDP from the University of Queensland is generally more accurate than the three most recent vintages the Penn World Tables and the two most recent vintages of the World Development Indicators (2) the newer vintages of the Penn World Tables and the World Development Indicators are not necessarily more accurate than the older vintages; and (3) World Bank GDP estimates based on the most recent international price survey of purchasing power parities (PPP) are about as accurate as estimates based on the 2005 round.

∗RAND Corporation and Pardee-RAND Graduate School
1 Introduction

The gross domestic product (GDP) is a widely-used economic indicator for which estimates are published by a number of sources, including: the World Bank (World Development Indicators); the Organization for Economic Cooperation and Development (the Eurostat-OECD PPP program); the International Monetary Fund (World Economic Outlook); the University of California at Davis (Penn World tables); and recently, the University of Queensland (UQ International Comparisons Data).

While the abundance of data is a good thing, a problem researchers face is that GDP numbers from different sources are inconsistent with each other, and it is unclear which set of estimates should be used. For example, Ram and Ural (2014) find that in nearly one-quarter of the sample in the 2005 versions of the World Development Indicators (WDI) and the Penn World Tables (PWT), per capita GDP levels differ either by more than 25% in relative terms, or by more than $2,500 in absolute terms. GDP estimates appear to be inconsistent even between subsequent versions of the same dataset despite using the same underlying numbers from national accounts data provided by country statistical offices. Breton (2012) points out that even over comparable time-periods, there are discrepancies in per capita GDP between PWT 7.0 and PWT 6.3. In a more detailed study, Johnson et al. (2013) find that annual growth rates differ by more than 1% between successive versions of the PWT in over half the countries, and by more than 2% in more than a quarter of the sample – a marked difference given that the average growth rate in the panel is around 1.5%. Another comparative analysis by Hanousek et al. (2008) concludes that even when estimates of the mean growth rates are similar between WDI, PWT and IMF data, there is low correlation between growth rates from the three time-series, and variability increases and

\[1\text{The absolute magnitudes of the differences are also significant, ranging between \$10,610 and \$9,900. There is no clear relationship between inconsistency of data and country characteristics such as income or geographic location, except that the very richest countries have the most consistent estimates. Moreover data discrepancies in the PWT and WDI appear to be persistent because an earlier analysis by Ram (2009) compared the previous versions of the PWT and WDI to published estimates from the International Comparisons Program (ICP) and found differences of a similar magnitude as in an equivalent comparison in Ram and Ural (2014).}\]

\[2\text{This average is for the PWT 6.2 panel. They use PWT 6.1, PWT 6.2, and PWT 7.0 in their study.}\]
cross-dataset correlation decreases for lower income countries.\textsuperscript{3}

Discrepancies in GDP statistics are not mere numerical curiosities – they can have a meaningful impact on policy analysis and economic research. Ciccone and Jarociński (2010) demonstrate that many variables are highlighted as robust determinants of growth in some vintages of the PWT, but not in others. For example, life expectancy, access to mining resources, relative price of investment goods and geography are found to be robust determinants of economic growth in PWT 6.0 and PWT 6.1, but not in PWT 6.2.\textsuperscript{4} Hanousek et al. (2008) replicate studies of growth determinants and conclude that: “\textit{Key relationships change in size and significance, frequently leading to fundamentally different conclusions were the analysis to be based on seemingly simple changes of data set.}” (p. 1200) Ponomareva and Katayama (2010) find that a purported negative relationship between growth rate volatility and the rate of GDP growth is likewise sensitive to the dataset used. Nalewaik (2014) shows that the severity of the Great Recession may have been obfuscated due to a “Loss of Signal Error” in GDP data, a type of attenuation where noise overwhelms the true component of the underlying variable being measured. Dawson et al. (2001) find that the evidence from cross-country growth studies for the Permanent Income Hypothesis – a statement regarding the relationship between consumption behavior and income shocks – may be an artefact of measurement error in GDP. They conclude by reflecting the consensus view that the “\textit{right thing to do is exclude the low quality data} /\ldots\textit{this implication, of course, is quite disappointing.}” (p. 1008)

In this paper I show that excluding low quality data is neither necessary nor recommended, and that it is possible to take an econometric approach to assess the relative accuracy of GDP estimates. I propose two ways to do so. The first approach is cast as a prediction problem: given several noisy measures of GDP, one can compute a weighted-average which best predicts the true value. Here “best” is defined in a statistical sense, i.e., the weighted-average that minimizes the mean squared prediction error (MSPE) with respect to the true value. GDP data that is accorded the highest weight in such an average is considered the most accurate. Even though true GDP is unobservable,\textsuperscript{3}\textsuperscript{4}Hanousek et al. (2008) contend that the discrepancies in growth rates can be explained, in part, by the “Gerschenkron effect” whereby the growth rate of PPP-adjusted GDP is lower for relatively poor countries, and higher for relatively richer countries. Their argument is based on the work of Nuxoll (1994) who posits that growth rate computations may be more accurate using national accounts data, rather than PPP-adjusted time-series. Nevertheless, inter-version variability is observed even for the WDI, which is more closely-coupled to national accounts, as compared with the PWT.

\textsuperscript{4}Robustness is a concept from the Bayesian model-averaging (BMA) literature. BMA sets a threshold for the probability of including a variable in a growth model, conditional on the data. Variables exceeding this inclusion probability threshold are called \textit{robust}. See Sala-i Martin et al. (2004) for technical details.
I show that an **orthogonal proxy** can be used in its place to derive these weights. Two proxies are defined to be orthogonal to each other if their measurement errors are uncorrelated. I demonstrate the application of this approach using growth in night-time luminosity as an orthogonal proxy for GDP. Night-time luminosity has been shown to be correlated with economic activity by Henderson et al. (2012), but errors in night-lights, likely caused by technical failures, are plausibly unrelated to errors in national accounts-based measures of GDP, which are the result of problems in data collection and statistical capacity.

This approach is similar to Pinkovskiy and Sala-i Martin (2016a) and Pinkovskiy and Sala-i Martin (2016b) in that it uses the ability of a given GDP time-series to predict night-time luminosity as a way to assess its relative accuracy. However, the econometric framework I outline is slightly more general (in which theirs is a special case), and in Monte Carlo simulations I find that the estimator for relative weights that results from my framework has lower bias and mean-squared error than theirs.

Furthermore, the econometric framework I propose enables us to arrive at broad conclusions about the types of variables that can serve as orthogonal proxies: specifically, I show that *any* valid instrumental variable (IV) for GDP can be used as an orthogonal proxy. This is beneficial not only because it addresses criticisms of using night lights in the approach here, but also because it makes the proposed technique more widely applicable to the literature on measurement error. For example, Pinkovskiy and Sala-i Martin (2016a) point out that the criticism facing night-time luminosity is that it tends to favor economic activity that produces relatively more light per unit of output. Therefore, night-lights are better suited as a GDP proxy for countries with relatively larger industrial sectors as compared to countries with relatively larger agricultural sectors, for example. No matter. One can circumvent this criticism by simply estimating the optimal weights separately for these two groups – using an IV for GDP that is more suitable for primarily agrarian countries in the latter case.

Despite its many appealing features, the drawback of the first approach is that it requires a valid IV to serve as an orthogonal proxy in order to identify the weights to be accorded to different series. Is it possible to assess the relative accuracy of GDP estimates without needing any auxiliary

---

5I thank James Hosek (RAND) for providing this insight.
6One such example would be lagged values of rainfall, which Miguel et al. (2004) show to be a valid IV for GDP growth in agricultural countries.
In a second approach (MVAR), I compute relative weights for different GDP series by minimizing the variance of the convex combinations of the underlying data. The intuition for this approach is that variability is approximately inversely related to precision, and is therefore a measure of data accuracy. This idea is conceptually similar to the motivation behind weighted least-squares (in which the inverse of residual variance is taken as an indicator of the quality of a given observation). I show that this approach is formally equivalent to the one taken by Pinkovskiy and Sala-i Martin (2016a) under the assumption – supported by what they observe in different GDP time-series – that the values of GDP in different datasets are equal on average.

I compare the three approaches (MSPE, MVAR and the Pinkovskiy and Sala-i Martin (2016a) approach (PS)) for assessing GDP accuracy using Monte Carlo simulations based on plausible parameter ranges for the global income distribution. I find the MSPE approach to have lower bias and mean-squared error than both MVAR and PS. The MVAR approach is also promising but its performance is driven in large part by the signal-to-noise ratio (SNR) of the data: it performs particularly well in low signal-to-noise environments, but poorly otherwise.\textsuperscript{7}

Applying the MSPE and MVAR approaches to the PWT and the WDI, the two most commonly-used GDP datasets, re-affirms old conclusions: Newer vintages of the data (both from the World Bank and the Penn World Tables) are not necessarily better than older ones; detailed PPP adjustment procedures from the PWT produce estimates that are apparently no more accurate than simpler extrapolations done in the WDI; and GDP estimates based on the more comprehensive PPP price survey in 2011 are no more accurate than the prior round in 2005.

But what’s past is prologue. I find that a new dataset from the University of Queensland substantially outperforms almost all other GDP panels considered in this analysis, and this superior performance appears to be driven by a methodological innovation in computing PPP extrapolations. This is an important conclusion because PPP-adjusted GDP estimates have been something of a pipe dream in economics: everybody agrees they are desirable for wealth comparisons across countries, yet PPP adjustments always seem to introduce error in GDP estimates. The methodology of the Queensland panel appears to be a significant step forward in overcoming this challenge.

\textsuperscript{7}This is not surprising: if the accuracy of the data is high already, then minimizing the variance is actually reducing the variation which contributes to signal.
2 Theory

In the next two sections I derive the MSPE and MVAR estimators for assessing the accuracy of GDP proxies. Here the word “proxy” refers to a given dataset of GDP, produced by the World Bank, the IMF, OECD, etc. I assume GDP estimates in each dataset (i.e. proxy) differ from the true value by some random error, and that the magnitude of this error varies between the series.

Assuming classical measurement error in GDP is consistent with the literature (see: Henderson et al. (2012), Chen and Nordhaus (2011), Pinkovskiy and Sala-i Martin (2016a), Pinkovskiy and Sala-i Martin (2016b)) however this assumption is not always satisfied. For example Rao and Selvanathan (1992) show that standard errors in purchasing power parity (PPP) estimates – and consequently, in PPP-adjusted GDP – are higher in lower-income countries. This could be because less developed countries lack statistical capacity and are likely to have more erroneous estimates of output. Or it could be because GDP is not as precise in countries where a large share of the production is informal, and such countries also tend to be poorer overall.

I should note that the methods discussed in this paper cannot assess this non-classical component of the measurement error in GDP. Instead, I only make judgements about relative accuracy of GDP statistics based on the classical component of measurement error that is different between series. More formally, I assume that for a true value $X$, I have two proxies ($X_1$ and $X_2$), in which measurement error takes the following form:

\begin{align}
    X_1 &= X + \gamma + u_1 \\
    X_2 &= X + \gamma + u_2
\end{align}

In equation (1) and (2) the fundamental component of error is $\gamma$, and includes possibly non-classical error due to country characteristics. I ignore this component by assuming that it is the same for both proxies because both use the same underlying national accounts data.

What differs between proxies is $u_1$ and $u_2$: the components of error resulting from differences in data aggregation procedures used by different agencies. Because they reflect the soundness of the estimation approach of a particular agency, $u_1$ and $u_2$ are plausibly uncorrelated with country characteristics. They might reflect, say, the granularity of raw data used, the sophistication of
the accounting methodology, the quantity of data used for point estimates, and so on. Unless statistical agencies that produce GDP estimates are systematically introducing bias based on country characteristics, assuming that \( u_1 \) and \( u_2 \) are classical errors seems plausible.

### 2.1 Minimum mean-squared prediction error weights (MSPE)

The first approach I use to identify the relative accuracy of GDP datasets is called the mean-squared prediction error (MSPE) approach. The idea behind MSPE weights is to form a weighted average of proxies which best predicts the true GDP. As in Pinkovskiy and Sala-i Martin (2016a), let \( X \) denote the true underlying income for which we have two noisy proxies \( X_1 \) and \( X_2 \) defined as:

\[
X_1 = m_1 X + u_1
\]

\[
X_2 = m_2 X + u_2
\]

I assume that \( u_1 \) and \( u_2 \) classical, mean-zero errors so that:

\[
\text{Cov}(X, u_1) = \text{Cov}(X, u_2) = 0 \quad \text{and} \quad E(u_1) = E(u_2) = 0
\]

However, I allow for the errors to be potentially correlated and the proxies to be potentially biased so:

\[
\text{Cov}(u_1, u_2) \in [0, 1)
\]

\[
m_1 \neq m_2 \quad \text{and} \quad m_1, m_2 \text{ are not necessarily } = 1
\]

---

8Even though I allow for the flexibility that \( m_1 \) and \( m_2 \) are potentially unequal, this is generally not found to be true in the literature for national accounts data. For example Pinkovskiy and Sala-i Martin (2016b) fail to find a statistically or economically significant difference between mean income estimates from several of the most recent vintages of the WDI and PWT in most cases. However, Deaton (2005) does find that income estimates from consumption surveys are about half of what is estimated by national accounts.
Let $Z$ be a convex combination of $X_1$ and $X_2$:  \footnote{A convex combination is just a special case of a weighted average where weights sum to unity. I consider a convex combination here to remain consistent with my approach in section (2.2) on minimum variance weights where convexity is required to ensure that a unique solution exists.}

$$Z = \lambda X_1 + (1 - \lambda)X_2$$  \hspace{1cm} (8)

Here $\lambda$ determines how much relative weight is assigned to $X_1$ versus $X_2$ in forming the combination. Our goal is to find a $\lambda$ such that the resulting $Z$ is an optimal predictor of $X$, i.e., it minimizes the mean squared prediction error between $Z$ and $X$. \footnote{Minimizing the mean-squared prediction error is just one of the many valid optimality conditions (another could be to minimize absolute error, for example). Relative to other optimality criteria, it is appealing because it is mathematically tractable as well as symmetric with respect to over- or under-estimating the true value. Intuitively, it can be interpreted as minimizing the Euclidean distance between the observed and true values.} Now consider predicting $X$ using $Z$ in a least-squares regression:

$$X = \alpha + \beta Z + \epsilon$$  \hspace{1cm} (9)

Here $X$ is a $N \times 1$ vector of true GDP values, $Z$ is an $N \times 1$ vector where each row contains a convex sum of data points $(\lambda x_{1i} + (1 - \lambda)x_{2i})$ with $x_{1i} \in X_1$ and $x_{2i} \in X_2$, and $\epsilon$ is the $N \times 1$ vector of residuals. The squared prediction error (i.e. squared residuals) will be given by:

$$\epsilon^2 = (X - \alpha - \beta Z)^2$$  \hspace{1cm} (10)

For a given value of $\lambda$, using ordinary least-squares on equation (9) we can find estimates $(\hat{\alpha}, \hat{\beta})$ that minimize the sum of squared residuals (or equivalently the mean of the squared residuals): \footnote{Formally, it makes no difference whether the mean or the sum of squared residuals is minimized as one is simply a monotonic transformation of another. For details see Chapter 3 of Lynch et al. (1998) whose approach I follow.}

$$\{\hat{\alpha}, \hat{\beta}\} = \arg \min_{\alpha, \beta} E(\epsilon^2) = E(X - \alpha(\lambda) - \beta(\lambda)Z)^2$$  \hspace{1cm} (11)

Once the regression coefficients $\hat{\alpha}$ and $\hat{\beta}$ have been found and substituted into (11), the minimized mean squared residuals in (11) will only vary as a function of $\lambda$ (as indicated by the parentheses in the equation). The optimal value of $\lambda$ will therefore be the one which yields the infimum of the set of minimized squared residuals given in (11). To find the optimal value, $\lambda^*_E$, we invoke the envelope theorem and simply take the derivative of (11) with respect to $\lambda$ and solve the resulting first-order
To find a closed-form analytic solution for $\lambda^*_E$ given in (12), I follow the steps in chapter 3 (p. 39-41) of Lynch et al. (1998) to first derive an algebraic expression for $E(\hat{\epsilon}^2)$. I then set its first-order condition to zero and solve. Start with the equation for the estimated residuals:

$$\hat{\epsilon} = X - \hat{\alpha} - \hat{\beta}Z$$  \hspace{1cm} (13)

Re-write (13) by adding and subtracting ($\bar{X} + \hat{\beta}\bar{Z}$) on the RHS (where $\bar{X} = E(X)$ and $\bar{Z} = E(Z)$):

$$\hat{\epsilon} = X - \hat{\alpha} - \hat{\beta}Z + (\bar{X} + \hat{\beta}\bar{Z}) - (\bar{X} + \hat{\beta}\bar{Z})$$  \hspace{1cm} (14)

$$= (X - \bar{X}) - \hat{\beta}(Z - \bar{Z}) - (\hat{\alpha} + \hat{\beta}\bar{Z} - \bar{X})$$  \hspace{1cm} (15)

Square both sides to get an expression for $\hat{\epsilon}^2$:

$$\hat{\epsilon}^2 = (X - \bar{X})^2 - 2\hat{\beta}(X - \bar{X})(Z - \bar{Z}) + \hat{\beta}^2(Z - \bar{Z})^2 + (\hat{\alpha} + \hat{\beta}\bar{Z} - \bar{X})^2$$

$$- 2(X - \bar{X})(\hat{\alpha} + \hat{\beta}\bar{Z} - \bar{X}) + 2\hat{\beta}(Z - \bar{Z})(\hat{\alpha} + \hat{\beta}\bar{Z} - \bar{X})$$  \hspace{1cm} (16)

Take the expected value of both sides, remembering the linearity of the expectation operator, and note that $E(X - \bar{X})$ and $E(Z - \bar{Z})$ are zero, by definition. Moreover $E(X - \bar{X})^2$ is the variance of $X$, $E(Z - \bar{Z})^2$ is the variance of $Z$, and $E[(X - \bar{X})(Z - \bar{Z})]$ is the covariance between $X$ and $Z$, also by definition. Therefore:

$$E(\hat{\epsilon}^2) = \text{Var}(X) - 2\hat{\beta}\text{Cov}(X, Z) + \hat{\beta}^2\text{Var}(Z) + (\hat{\alpha} + \hat{\beta}\bar{Z} - \bar{X})^2$$  \hspace{1cm} (17)

I have omitted the expectation operator on the last term because it is a sum of scalar values (this

---

12The two variable case is easy to solve analytically, but the general case with $K$ proxies might not have a closed-form analytic solution. In that case, stochastic optimization methods may be used to find the $K$-dimensional vector of optimal weights by estimating a large number of regressions of the form (9), where the weight vector is randomly varied for each iteration. Then optimal weight vector would be one with the lowest mean of the squared residuals.
is because the derivative will be taken at the optimum values $\hat{\alpha}$ and $\hat{\beta}$, of $\alpha$ and $\beta$, respectively).

From standard results for univariate OLS, the estimated values $\hat{\alpha}$ and $\hat{\beta}$ are given as:

$$\hat{\alpha} = \bar{X} - \hat{\beta}\bar{Z} \quad \text{and} \quad \hat{\beta} = \frac{\text{Cov}(X, Z)}{\text{Var}(Z)} \quad (18)$$

Plugging the values of $\hat{\alpha}$ and $\hat{\beta}$ from (18) into (17) yields an algebraic expression for $E(\hat{\epsilon}^2)$:

$$E(\hat{\epsilon}^2) = \text{Var}(X) - \frac{(\text{Cov}(X, Z))^2}{\text{Var}(Z)} \quad (19)$$

Then, re-expressing $Z = \lambda X_1 + (1 - \lambda)X_2$ gives:

$$E(\hat{\epsilon}^2) = \text{Var}(X) - \frac{(\text{Cov}(X, \lambda X_1) + \text{Cov}(X, (1 - \lambda)X_2))^2}{\text{Var}(\lambda X_1 + (1 - \lambda)X_2)} \quad (20)$$

$$= \sigma_X^2 - \frac{(\lambda \sigma_{XX_1} + (1 - \lambda)\sigma_{XX_2})^2}{\lambda^2 \sigma_{X_1}^2 + (1 - \lambda)^2 \sigma_{X_2}^2 + 2\lambda(1 - \lambda)\sigma_{X_1X_2}} \quad (21)$$

The optimal MSPE weights $(\lambda_E^*)$ are given by setting the derivative of (21) with respect to $\lambda$ equal to zero, and solving:

$$\lambda_E^* = \frac{2\sigma_{X_2}^2 \sigma_{XX_1} - \sigma_{X_1X_2}\sigma_{XX_2}}{2\sigma_{X_1}^2 \sigma_{XX_2} + 2\sigma_{X_2}^2 \sigma_{XX_1} - \sigma_{X_1X_2}(\sigma_{XX_1} + \sigma_{XX_2})} \quad (22)$$

The expression in (22) depends on the unknown covariances $\sigma_{XX_1}$ and $\sigma_{XX_2}$. The equation (22) can be re-written by noting that $\sigma_{XX_i} = m_i \text{Var}(X)$:

$$\lambda_E^* = \frac{2\sigma_{X_2}^2 m_1 - \sigma_{X_1X_2} m_2}{2\sigma_{X_1}^2 m_2 + 2\sigma_{X_2}^2 m_1 - \sigma_{X_1X_2}(m_1 + m_2)} \quad (23)$$

Even though $m_1, m_2$ are unknown, I now show that it is possible to estimate (23) exactly as long as an orthogonal proxy for $X$ exists.

---

13 For example see equation (3.14a) and (3.14b) in Chapter 3 of Lynch et al. (1998).

14 I used WolframAlpha to calculate the derivative and arrive at (22). Moreover, solving the first-order condition based on (21) yields two solutions but the other solution $\lambda_E = \frac{\sigma_{XX_2}}{\sigma_{XX_1}}$ is inadmissible because it leads to a contradiction when the proxies have equal means. To see this suppose that $m_1 = m_2$. Then $\sigma_{XX_1} = \sigma_{XX_2}$, which would result in weights that are undefined.
Definition 1. Y is an orthogonal proxy for X if the measurement error in Y is uncorrelated with the measurement errors in the proxies \( X_1 \) and \( X_2 \). In particular, if \( Y = m_3 X + \nu \), then we require that \( \text{Cov}(\nu, u_{i\in 1,2}) = 0 \) and \( \text{Cov}(X, \nu) = 0 \).

Suppose we have a valid orthogonal proxy \( Y \). Then \( \sigma_{YX_i} = \text{Cov}(m_3 X + \nu, m_i X + u_i) = m_3 m_i \text{Var}(X) \).

Plugging-in \( \sigma_{YX_i} \) in place of \( \sigma_{XX_i} \) in (22) gives:

\[
\lambda^*_E = \frac{2\sigma^2_{X_2} m_3 m_1 \text{Var}(X) - \sigma_{X_1 X_2} m_3 m_2 \text{Var}(X)}{2\sigma^2_{X_1} m_3 m_2 \text{Var}(X) + 2\sigma^2_{X_2} m_3 m_1 \text{Var}(X) - \sigma_{X_1 X_2} (m_1 + m_2) m_3 \text{Var}(X)}
\]

(24)

\[
= \frac{(2\sigma^2_{X_2} m_1 - \sigma_{X_1 X_2} m_2) m_3 \text{Var}(X)}{(2\sigma^2_{X_1} m_2 + 2\sigma^2_{X_2} m_1 - \sigma_{X_1 X_2} (m_1 + m_2)) m_3 \text{Var}(X)}
\]

(25)

Since by definition \( m_3 \text{Var}(X) \) is non-zero, we can factor it out from the numerator and denominator, thus yielding:

\[
\lambda^*_E = \frac{(2\sigma^2_{X_2} m_1 - \sigma_{X_1 X_2} m_2)}{(2\sigma^2_{X_1} m_2 + 2\sigma^2_{X_2} m_1 - \sigma_{X_1 X_2} (m_1 + m_2))}
\]

(26)

And this is exactly equivalent to (23). Hence, if we have data on \( Y \), we can estimate MSPE weight on \( X_1 \) as:

\[
\lambda^*_E = \frac{2\sigma^2_{X_2} \sigma_{YX_1} - \sigma_{X_1 X_2} \sigma_{YX_2}}{2\sigma^2_{X_1} \sigma_{YX_2} + 2\sigma^2_{X_2} \sigma_{YX_1} - \sigma_{X_1 X_2} (\sigma_{YX_1} + \sigma_{YX_2})}
\]

(27)

The weight on \( X_2 \) will just be \( (1 - \lambda^*_E) \). Moreover, any valid instrument for \( X \) will help identify (23). The proof of this is given in the Appendix (A).

### 2.2 Minimum variance weights (MVAR)

The second approach for identifying the relative accuracy of a GDP dataset is termed the minimum variance (MVAR) approach. It is motivated by the fact that we may not always have a suitable orthogonal proxy to use for our purposes. To understand the intuition behind the MVAR approach, temporarily assume that the quantity of error in each proxy is known. Then the MVAR approach says to weight each proxy proportional to its error variance such that the overall error in the resulting sum is minimized. Aruoba et al. (2013) use such an approach, for example, to compute optimal
forecasts of quarterly GDP time-series for the U.S. Henderson et al. (2012) use such an approach to combine night-lights with GDP data into an optimal proxy of income. In the meta-analysis literature, such an approach is used to combine effect estimates from different studies (Hartung et al., 2011).\footnote{Interestingly, to first order, this is also how the primate visual cortex integrates visual information. (Bahrami et al., 2012)} In practice, the error variance is not known so assumptions have to be made in order to apply this variance-minimization heuristic. The MVAR approach proposed here diverges from existing approaches by instead minimizing the overall variance of the convex combination of noisy proxies. As I show later, if the proxies are unbiased, or if their means are equal, minimizing the overall variance is equivalent to minimizing the error variance. The most pathological case is when the means of the proxies are not equal. Here the MVAR heuristic is not appropriate. Nevertheless, I find that the MVAR approach is surprisingly robust, given the properties of GDP time-series data we are likely to deal with.

As before, let \( Z \) be the convex combination of \( X_1 \) and \( X_2 \) with weight \( \lambda \).\footnote{To keep the exposition simple, I consider the case when I have \( K = 2 \) noisy proxies for the unobserved true GDP. For the general case with \( K \) proxies, and a \( K \times K \) variance-covariance matrix \( \Sigma \), and a \( K \times 1 \) weight vector \( \Lambda \), the variance of the \( K \)-dimensional convex hull of data points will be \( \Lambda^T \Sigma \Lambda \). In order to solve for the optimal weights \( \Lambda^* \), one would take the derivative \( \frac{d}{d \lambda} \Lambda^T \Sigma \Lambda \) and solve the resulting homogenous system of \( K \) equations given by \( 2 \Lambda^T \Sigma = 0 \). A possible computational strategy for doing so is given by Paul Rubin at: A Minimum Variance Convex Combination Online at: http://orinanobworld.blogspot.com/2013/09/a-minimum-variance-convex-combination.html. [Accessed: December 19th, 2016].}

\[
Z = \lambda X_1 + (1 - \lambda) X_2 \tag{28}
\]

The variance of \( Z \) will be given as:

\[
\text{Var}(Z) = \lambda^2 \text{Var}(X_1) + (1 - \lambda)^2 \text{Var}(X_2) + 2\lambda(1 - \lambda)\text{Cov}(X_1, X_2) \tag{29}
\]

The minimum-variance optimal weight \( (\lambda^*_V) \) is one which minimizes \( \text{Var}(Z) \):

\[
\lambda^*_V = \arg \min_{\lambda} \text{Var}(Z(\lambda)) \tag{30}
\]

Taking the derivative of (29) with respect to \( \lambda \) and solving the resulting first-order condition yields
the desired value for $\lambda^*_V$:

$$\lambda^*_V = \frac{\sigma^2_{X_2} - \sigma_{X_1} X_2}{\sigma^2_{X_1} + \sigma^2_{X_2} - 2\sigma_{X_1} X_2}$$  \hspace{1cm} (31)$$

Note that (31) can be written as:

$$\lambda^*_V = \frac{m_2(m_2 - m_1)\sigma^2_X + \sigma^2_u - \sigma_{u_1} u_2}{(m_2 - m_1)^2\sigma^2_X + \sigma^2_{u_1} + \sigma^2_u - 2\sigma_{u_1} u_2}$$  \hspace{1cm} (32)$$

If $m_1 = m_2$ then the terms with $\sigma^2_X$ disappear and we are left with an expression we would have obtained as a result of minimizing the convex combination of the error-variance.$^{17}$ However, in general, both proxies will only have approximately equal means ($m_1 \approx m_2$). In Section (3), I quantify the effect of changes in $m_1$ vs. $m_2$ on the bias of the resulting estimates by computing numerical derivatives of (31) with respect to $m_1$ and $m_2$ (and other parameters). I find that for deviations of $m_1$ and $m_2$ from unity that we are likely to observe in real data, the effect on the bias is minimal, and that the minimum-variance heuristic as outlined above, generally performs quite well.

### 2.3 Some properties of relative weights

In this section I establish some properties of the two weighting schemes and of the resulting optimal estimates of GDP, and compare them to existing approaches. In particular, I show that:

1. If the proxies are biased, then their weighted average is also biased regardless of the approach taken.

2. If the proxies are biased, but their means are equal, then MVAR and PS produce the same estimate of the weighted average, namely the sample mean of any of the proxies. The weighted average resulting from MSPE will be the same as MVAR and PS, but the estimates of the weights will be different.

3. Under biased-but-equal means, the simpler MVAR approach will assign the same relative weight to each proxy as the more data-intensive approach using night lights taken by Pinkovskiy and Sala-i Martin (2016a) and Pinkovskiy and Sala-i Martin (2016b).\footnote{For example, see equation (4) on page 5 of Aruoba et al. (2013), where they minimize the error variance of GDP proxies.}

---

\(^{17}\)For example, see equation (4) on page 5 of Aruoba et al. (2013), where they minimize the error variance of GDP proxies.
4. The approach taken in Pinkovskiy and Sala-i Martin (2016a) and Pinkovskiy and Sala-i Martin (2016b) is a special case of the MSPE framework, with MSE’s that are always greater than or equal to the MSE’s yielded by the MSPE approach.

5. The MSPE framework can also be seen as an approximate generalization of the MVAR weights.\(^{18}\)

The third point is a bit nuanced: even though the point estimate of the weighted average under MSPE might be equal to the point estimate by MVAR, their assessment of the relative weight to assign each proxy will be different.

2.3.1 Biasedness of MSPE

Let \(Z^*_E\) be the optimal estimate of GDP produced using MSPE weights. Then the bias will be given as:

\[
E(X) - E(Z^*_E) = E(X) - (\lambda^*_E E(X_1) + (1 - \lambda^*_E)E(X_2))
\]

\[
= E(X) - (\lambda^*_Em_1E(X) + (1 - \lambda^*_Em_2E(X))
\]

\[
= E(X) - \frac{2(m_1^2\sigma^2_{X_2} + m_2^2\sigma^2_{X_1} - m_1m_2\sigma_{X_1X_2})}{m_1\sigma^2_{X_2} + m_2\sigma^2_{X_1} - \sigma_{X_1X_2}(m_1 + m_2)}E(X)
\]

\[
= \left(1 - \frac{2(m_1^2\sigma^2_{X_2} + m_2^2\sigma^2_{X_1} - m_1m_2\sigma_{X_1X_2})}{2m_1\sigma^2_{X_2} + 2m_2\sigma^2_{X_1} - \sigma_{X_1X_2}(m_1 + m_2)}\right)E(X)
\]

Hence MSPE will be unbiased if and only if the term inside the parenthesis is zero, which is only true for \(m_1 = m_2 = 1\).

2.3.2 Biasedness of MVAR

Let \(Z^*_V\) be the optimal estimate of GDP produced with MVAR weights. Then a similar calculation as above yields:

\[
E(X) - E(Z^*_V) = \left(1 - \frac{m_1\sigma^2_{X_2} + m_2\sigma^2_{X_1} - \sigma_{X_1X_2}(m_1 + m_2)}{\sigma^2_{X_1} + \sigma^2_{X_2} - 2\sigma_{X_1X_2}}\right)E(X)
\]

\(^{18}\)I thank James Hosek for showing me this line of reasoning and the related calculation.
Hence MVAR weights will be unbiased if and only if \( m_1 = m_2 = 1 \).

### 2.3.3 Biasedness of Pinkovskiy and Sala-i-Martín (PS)

Note that the procedure used by Pinkovskiy and Sala-i-Martin (2016a) to arrive at optimal weights for different GDP series \((X_1, X_2)\) is to solve the following minimization problem:

\[
\lambda_P^* = \arg \min_{\lambda} E((X - \lambda X_1 - (1-\lambda)X_2)^2) \quad (38)
\]

It is easy to see that equation (38) is a special case of the formulation in section (2.1) by constraining \( \alpha \) to be zero in equation (10). Reformulating equation (11) from Pinkovskiy and Sala-i-Martin (2016a) into the notation I use in this paper, we can see that \( \lambda_P^* \) will equal:\(^{19}\)

\[
\lambda_P^* = \frac{\sigma^2_{X_2}m_1 - \sigma_{X_1}X_2m_2}{\sigma^2_{X_1}m_2 + \sigma^2_{X_2}m_1 - \sigma_{X_1}X_2(m_1 + m_2)} \quad (39)
\]

Now let the optimal estimate of \( X \) using \( \lambda_P^* \) be \( Z_P^* \). Then following the same procedures as before, it is easy to see that bias will be given as:

\[
E(X) - E(Z_P^*) = \left(1 - \frac{m_1^2\sigma^2_{X_2} + m_2^2\sigma^2_{X_1} - 2m_1m_2\sigma_{X_1}X_2}{m_1\sigma^2_{X_2} + m_2\sigma^2_{X_1} - \sigma_{X_1}X_2(m_1 + m_2)}\right)E(X) \quad (40)
\]

Hence the PS weights will be unbiased if and only if \( m_1 = m_2 = 1 \). Moreover, since \( \alpha \) is constrained to be zero, the MSE of the PS estimator will always be greater than or equal to the MSE of the MSPE estimator.

### 2.3.4 Equivalence of MVAR and PS

Suppose that proxies have the same means so that \( m_1 = m_2 = m \). Then it is fairly straight-forward to see that they will yield the same optimal estimate of \( X \):

\[
E(Z_V^*) = E(Z_P^*) = mE(X) \quad (41)
\]

Under such a condition, a stronger result can also be established, namely that MVAR weights and PS weights are identical. To see this, set \( m_1 = m_2 = m \) in equation (39), and factor it out from the

\(^{19}\)This follows by noting that in their equation (11) \( \beta^G = m_1, \beta^S = m_2, \text{Cov}(\epsilon_{i,t}^G, \epsilon_{i,t}^S) = \sigma_{X_1}X_2, \text{Var}(\epsilon_{i,t}^G) = \sigma^2_{X_1}, \text{and Var}(\epsilon_{i,t}^S) = \sigma^2_{X_2}.\)
numerator and denominator. Formally, this results in:

\[ \lambda^*_V = \lambda^*_P \text{ if } m_1 = m_2 \]  

(42)

In other words, if the means are the same, then minimizing the variance of the convex combination of data yields the same weights as the approach of Pinkovskiy and Sala-i Martin (2016a). This is a somewhat surprising result, but bodes well for economic research because it shows that a simpler approach is equivalent to a more complex one, under a very mild equality of means assumption.

Pinkovskiy and Sala-i Martin (2016b) conduct formal tests of the hypothesis that \( m_1 = m_2 \) and fail to reject it in most cases using WDI and PWT data. Moreover, they do not find an economically significant difference in the relative size of \( m_1 \) and \( m_2 \). The same assumption is made by Aruoba et al. (2016), Aruoba et al. (2013), Henderson et al. (2012) and Chen and Nordhaus (2011).

The estimate of the relative weights derived using the MSPE approach in section (2.1) will be different, however, because MSPE guarantees a lower mean-squared error. This boils out of the fact that, by not constraining \( \alpha \) to be zero, the MSPE approach utilizes a more flexible model to derive the weights.

### 2.3.5 Relationship between MVAR and MSPE

From the expressions for MVAR and MSPE weights, it is clear that MVAR accords weights to one proxy based on: the variance in the other proxy and the covariance between the two proxies. However, MSPE weights, in addition to using the covariance between proxies, also take into account the strength of the relationship between a proxy and the true \( X \) (this is reflected by the inclusion of the \( m_1 \) and \( m_2 \) terms in the MSPE weights). Suppose that both proxies have equal means, then by setting \( m_1 = m_2 = m \) and taking out a factor of 2 from the numerator and the denominator, we can see that the MSPE weights will be given as:

\[ \lambda^*_E = \frac{\sigma^2_{X_2} - \frac{1}{2}\sigma_{X_1X_2}}{\sigma^2_{X_1} + \sigma^2_{X_2} - \sigma_{X_1X_2}} \]  

(43)

Comparing this to the expression for the MVAR weights, we find that MSPE weights accord about half as much importance to the covariance between \( X_1 \) and \( X_2 \) as compared with MVAR.

\[20\text{ Refer to footnote 6 in their paper.}\]
3 Simulation Studies

We now have three different estimators for assessing the relative accuracy of GDP statistics: the MSPE and MVAR approaches derived in sections (2.1) and (2.2), respectively and the PS estimator derived in Pinkovskiy and Sala-i Martin (2016a). I now compare the performance of each approach in a simulation study by assessing average bias and mean-squared error (MSE) of each estimator over a plausible range of parameter values, as is standard in the literature (see Chapter 3 of Abu-Mostafa et al. (2012)). I also quantify the effect of changes in parameters on the bias and MSE, through log-log regressions with the bias and MSE as the dependent variables and the parameter values as independent variables. The coefficients in such a regression represent the percent change in the outcome for a 1 percent change in parameter value.

3.1 Setup

Based on Lopez and Servén (2006), I assume that the true underlying GDP follows a normal distribution with mean $\mu_X$ and variance $\sigma_X^2$. I choose $\mu_X = 1$ and three values for $\sigma_X^2$: 0.5, 1.0, and 2.0. Because the mean simply translates a normal distribution without affecting its cumulative distribution function, the specific value of $\mu_X$ does not matter. What is more important is the magnitude of $\sigma_X^2$ relative to the mean, and to the error variances $\sigma_{u1}^2$, $\sigma_{u2}^2$ and $\sigma_\nu^2$ (all three of which I set equal to 1). Setting $\sigma_X^2 = 1$ reflects the base case, since the standard deviation of log per capita GDP (PPP) is about as large as the mean.\(^{21}\) The two additional values for $\sigma_X^2$ are chosen to reflect two measurement error scenarios in addition to the base case: a situation in which measurement error variance is larger than the signal variance (i.e. $\sigma_X^2 = 0.5$), and a situation in which it is smaller (i.e. $\sigma_X^2 = 2.0$). Loosely speaking, increasing $\sigma_X^2$ increases the signal-to-noise ratio (SNR). To formalize the setup, define the true per capita GDP $X$ as:

$$X \sim N(1, \sigma_X) \tag{44}$$

Once $X$ is fixed, two possibly-biased proxies, with mean-zero noise, are simulated as:

\(^{21}\)Based on my calculations using WDI 2015 and data for 2014.
\[ X_1 = m_1 X + u_1 \]  
\[ X_2 = m_2 X + u_2 \]

Here \( m_1 \) and \( m_2 \) control how biased each proxy is. I sample both \( m_1 \) and \( m_2 \) uniformly over the interval \([0.5, 1.5]\). This decision is informed by the work of Deaton (2005) showing that the ratio of income estimates from consumption surveys to income estimates from national accounts range between \([0.6, 0.8]\), and also the work of Inklaar and Rao (2015) which finds that the percentage increase in income levels between ICP 2011 and ICP 2005 was around 25\%. I choose a range that envelopes these values. \( Y \) is a third proxy that is possibly orthogonal to \( X_1 \) and \( X_2 \):

\[ Y = m_3 X + \nu \]

The relationship between (44), (45), (46) and (47) is defined by the following variance-covariance matrix \( M \):

\[
M = \begin{pmatrix}
    u_1 & u_2 & \nu & X \\
    u_1 & 1 & \cdots & \cdots & \cdots \\
    u_2 & \sigma_{u_1 u_2} & 1 & \cdots & \cdots \\
    \nu & \sigma_{\nu u_1} & \sigma_{\nu u_2} & 1 & \cdots \\
    X & \sigma_{X u_1} & \sigma_{X u_2} & \sigma_{X \nu} & \sigma_X^2
\end{pmatrix}
\]  

(48)

These eight elements of the matrix \( M \) define a space, \((\Omega : \{M, m_1, m_2\})\), which spans the possible outcomes of this simplified data generating process. I do not sample values of \( m_3 \) because, as shown in (25), it does not affect the magnitude of the optimal weights. All covariances in \( M \) are picked uniformly over \([0, 0.5]\). Estimates of the covariances between measurement errors and true values do not exist anywhere to my knowledge and this seems to be a plausible range of values. I keep the covariances positive because anti-correlated measures of GDP seem unrealistic.

The pseudo-code for generating and analyzing the simulated data is given below:

1. Generate a random variance-covariance matrix \( M_i \), defined in (48) above by picking its entries from a uniform distribution over \([0, 0.5]\). Note that the simulation is run three times, once for
each value of $\sigma^2_X$.

2. Produce values for $u_1$, $u_2$, and $\nu$ by simulating from a multivariate random normal with zero mean, and the variance covariance matrix $\Sigma = M_i$.

3. Generate a random value for $m_1$ and for $m_2$ from a uniform distribution $U \sim [0.5, 1.5]$.

4. Simulate $X \sim N(1, \sigma^2_X)$, with length $s = 1000$. Here $s$ is the size of the simulated dataset.

5. Generate simulated values for $X_1 = m_1X + u_1$, $X_2 = m_2X + u_2$, and $Y = X + \nu$.

6. Compute $\lambda^*_E$ and $\lambda^*_V$ using the simulated $X_1$, $X_2$, and $Y$.

7. Compute $\text{Bias}(\lambda^*_E)$, $\text{Bias}(\lambda^*_V)$ and $\text{MSE}(\lambda^*_E)$, $\text{MSE}(\lambda^*_V)$, using the simulated data $\{X_1, X_2, Y\}$, and the known true value $X$.

8. Repeat steps (2) - (7) $N = 20,000$ times. Here $N$ is the number of Monte Carlo runs.

This process results in a $N \times K$ matrix, where $K = 8 + 4$: eight columns for the parameters and four columns total for the two bias and two MSE terms. The bias and MSE terms are used as dependent variables in four separate regressions, with the parameter values as independent variables:

\begin{align}
\ln(\text{Bias}) &= \beta_0 + \beta_1 \ln(\sigma_{u_1u_2}) + \beta_2 \ln(\sigma_{u_1\nu}) + \cdots + \beta_7 \ln(m_1) + \beta_8 \ln(m_2) \\
\ln(\text{MSE}) &= \gamma_0 + \gamma_1 \ln(\sigma_{u_1u_2}) + \gamma_2 \ln(\sigma_{u_1\nu}) + \cdots + \gamma_7 \ln(m_1) + \gamma_8 \ln(m_2)
\end{align}  

(49)  

(50)

I use a log-log specification so that the coefficient $\beta_i$ or $\gamma_i$ on a given parameter can be interpreted as an elasticity: that is, the percent change in bias or MSE for a one percent change change in the parameter value. The MVAR coefficients can thus be compared against the MSPE coefficients to assess the relative effect each parameter has on the performance of each weighting scheme.

### 3.2 Sampling

In order to generate values of the parameters of the variance-covariance matrix, I use a Halton sequence over the prescribed parameter ranges. This is a low-discrepancy (pseudo-random) sequence whose values are weakly correlated (Halton, 1964). The correlation ensures even coverage over the
high-dimensional parameter space. If uniform random sampling is used, it may result in closely-clustered points that may miss important regions of the parameter space. For example, suppose that 50 values are sampled for each parameter, then a full factorial Monte Carlo would result in approximately \((50)^8 \approx 3.9 \times 10^{13}\) (around 40 trillion) distinct variance-covariance matrices. However, many of these matrices will be redundant though we have no way of knowing \textit{a priori} which these are. Low-discrepancy sequences solve this problem by guaranteeing well-spaced draws, and thus even coverage, in the parameter space without having to do a full factorial Monte Carlo.

3.3 Results

In figure (1) I plot the kernel density estimates of the distribution of values of MSPE (blue), MVAR (red), and the PS (green) weights. Notice that all three distributions are centered around the value \(\lambda = 0.5\), which is the the ensemble average \(\lambda\) across the \(N = 20,000\) Monte Carlo runs. This is the value of \(\lambda\) one would pick if one didn’t know \textit{anything} about the different datasets. In other words since the average across all Monte Carlo runs represents a situation where every assumption about the data-generating process is relaxed, in such a case one proxy is no better than the other (by symmetry), and should therefore be assigned half the weight.

With this in mind, the dispersion of density estimates can be interpreted to reflect the informativeness of each estimator. In particular, less dispersion implies more informativeness, because such a distribution prefers a certain range values over others. We see that MSPE is uniformly more informative than MVAR or PS, regardless of the signal-to-noise ratio (SNR), given that it is tightly centered around 0.5 in all three panels 1(a) - 1(c). On the other hand the MVAR estimator is more informative than the PS estimator in low SNR environments (i.e. panel 1(a)), but the trend reverses as the SNR improves. In panel 1(c), MVAR weights are essentially uninformative, and assign roughly all weights to the proxies with about equal probability. This establishes the first conclusion about the three weighting schemes: MSPE weights are the best of all worlds, but in low SNR environments the performance of MVAR weights is close enough. This is a useful result because it suggests that MVAR weights can be used in the developing-country contexts with high levels of uncertainty, and where orthogonal proxies are not easily found.

In order to understand how deviations from \(m_1 = m_2\) affect performance, in figure (2) I plot the bias in each estimator as a function of the difference \((m_1 - m_2)\). To help visualize the trend
Figure 1: Kernel density estimates of the distribution of simulated values of $\lambda^*_E$ (blue), and $\lambda^*_V$ (red) for $\sigma^2_X = 0.5, 1.0,$ and $2.0$. Each plot was produced using $N = 20,000$ Monte Carlo samples. ($\sigma^2_X$) denotes the signal variance.

more easily, I use locally-weighted regression (LOESS) and compute polynomial fits of the bias function on subsets of the data. Each point on such a plot is the weighted-average of the fitted points in its neighborhood. This is a widely used data visualization method to assess structural relationships involving complex functions and a high density of points (see Cleveland and Devlin (1988) and Jacoby (2000), for example). The local maximum observed at $m_1 - m_2 = 0$ should not cause alarm: indeed, one does expect a minimum there, but keep in mind that bias is zero if and only if $m_1 = m_2 = 1$. For values of $m_1 = m_2 = m$, the difference will be zero, but the bias can be arbitrarily high (depending on the value of $m$), which results in the bunching there.

Across all three panels, the curve for MSPE weights lies below the curves for MVAR and PS
Figure 2: The LOESS-smoothed scatterplot of bias as a function of the difference in \( m_1 \) and \( m_2 \). The bandwidth for the LOESS-smoother is chosen to be 0.25 (i.e. each point is a local regression of 0.25m closest points) in \( \lambda_E^* \) (blue), and \( \lambda_V^* \) (red) for \( \sigma_X^2 = 0.5, 1.0, \) and 2.0. The shaded regions depict the 95\% confidence intervals.

weights, implying that MSPE yields the least-biased values over the range of parameters considered in this simulation.\(^{22}\) Similar to what we observed before, MVAR weights are preferable to PS weights in low SNR environments, but their performance degrades as the data accuracy improves. There is a region of overlap whose length is defined by \( |m_1 - m_2| \leq q \), where all three estimators are very close in terms of their bias. The size of this region (i.e \( q \)) decreases by about 50\% as we move from panel 2(a) to 2(c), and the SNR improves. Qualitatively speaking, this suggests that the relative performance of the estimators is dependent on the accuracy of the underlying data.

\(^{22}\)This is only an empirical observation. Absent analytical proof, it might well be the case that there are situations where MVAR and/or PS biases are lower than MSPE bias.
Bias is just one aspect of performance. A broader measure is the mean-squared error of estimates generated by each of the estimators. In figure (3), I plot a LOWESS-smoothed estimate of the MSE against \((m_1 - m_2)\). The fact that MSPE weights are well-behaved is evident here: unlike both PS
and MVAR weights, the curvature of MSPE weights remains relatively stable across panels 3(a) to 3(c). This suggests that the overall performance of MSPE weights is quite robust to the vagaries of the data-generating process. The region of overlap $q_i$, defined as before, is quite small to begin with and doesn’t change much from 3(a) to 3(c), meaning that the overall performance of each of the three estimators is quite different, regardless of data accuracy.

To make these qualitative observations more precise, in table (1) and (2), I present the results of the regression specifications given in (49) and (50), respectively. The reported coefficients in both tables are elasticities: the percent change in bias or MSE for a one percent change in a given parameter. Each regression specification is run three times for the three different SNR regimes based on the value of $\sigma^2$. From table (1) we see that the MSPE has the lowest average bias among the three estimators, regardless of the SNR. Moreover, for the MSPE estimator, the average bias decreases as SNR increases while at the same time the bias for MVAR increases. The bias for PS weights remains roughly the same in all three SNR regimes. The change in bias for all three estimators is most sensitive to changes in $\sigma^2_{u_1u_2}$. The sensitivity of the bias to the magnitude of $m_1$ and $m_2$ is interesting. For the MSPE estimator, it declines by an order of magnitude as the signal variance is increased from $\sigma^2_X = 0.5$ to 2. However, neither MVAR nor the PS weights exhibit such sharp drop-offs. Violations in the classical measurement error assumption, i.e. non-zero $(\sigma^2_{Xu_i}(i \in 1, 2))$, or non-zero $(\sigma^2_{X\nu})$ appear to be no more important than other factors, which is also surprising. In the main, parameter coefficients for the MSPE estimator are largest, while MVAR parameter coefficients are in general smallest.

A similar analysis of the MSE is reported in table (2). As before, the MSPE estimator has the lowest MSE of all three, followed by PS in the high SNR regimes and MVAR in the low SNR regime. In all three estimators, mean square error increases as SNR goes up, although the increase is not that large. I suspect that increases in the signal variance also lead to small increases in $\text{Cov}(X, error)$ of various kinds, which results in the overall MSE being slightly larger. As with bias, MSE appears to be most sensitive to increases in the error correlation between proxies of the national accounts type. All other parameters affect the MSE by at least an order-of-magnitude less.

Thus, in summary, the following notes summarize what we have learned about the performance of the estimators:
1. The MSPE estimator of optimal weights has the lowest bias and lowest MSE among the three options considered here.

2. The relative performance of PS and MVAR estimators is governed by the SNR in the data: in low SNR environments, MVAR performs better than PS.

3. In general, lower SNR’s make the bias of each estimator closer to each other, but MSE of the estimators are generally different.

4. Bias and MSE for the MSPE and PS weights are most sensitive to the error correlation of proxies (i.e. $\sigma_{u_1u_2}$), whereas the bias and MSE for the MVAR weight is most sensitive to $m_1$ and $m_2$.

4 Application to GDP data

4.1 Data

The GDP estimates used in this analysis come from three sources, selected to represent different data collection and computation methodologies. I use the 2013 and 2015 data tables of the World Development Indicators (WDI) from the World Bank, which are based on the 2011 and 2005 round of the International Comparisons Program (ICP), respectively. I use three versions of the Penn World Tables (PWT): PWT 9.0 is the latest version released in June, 2016 and it is based on the 2011 round of the ICP; PWT 8.1 and PWT 7.1 are both based on the 2005 ICP. Finally, I use a new panel dataset released through the University of Queensland International Comparisons Data program (UQICD). This dataset is produced using all available ICP benchmarks up to and including the most recent round in 2011. The characteristics and sample-period coverage for each of these datasets is given in Table (3). The night-lights data I use are from Henderson et al. (2012).

4.1.1 ICP

GDP data is published in local currency units by country statistical offices. To make cross-country comparisons possible, these numbers are expressed in a common currency, either by using currency market exchange rates, or by using purchasing power parity (PPP) exchange rates. A PPP is a conversion factor which states how much currency is required in a given country to purchase an
equivalent basket of goods in the currency of a numeraire country. Because PPPs are computed using the consumption and production patterns across countries, they are preferred over market exchange rates for currency comparisons because the latter only reflect the traded sectors of an economy, and these might play a smaller part in the economies of developing nations (Callen, 2007).

Collecting information on the consumption and prices of goods and services around the world is the purview of the ICP, which has been holding regular price surveys once every five or six years starting in 1972. The year of the ICP survey is called the benchmark year, and participating countries are called benchmark countries. There were 199 economies that participated in the 2011 ICP, up from 146 that participated in the 2005 round. Since ICP information is only available for the benchmark years (and countries), it must be extrapolated to non-benchmark years and countries. As Johnson et al. (2013) show, details of the extrapolation procedure is one of the primary sources of variability across the datasets. As such, I discuss the procedure used by each series in some detail below.

4.1.2 WDI

I use the per capita PPP GDP series \((NY.GDP.PCAP.PP.KD)\) from WDI 2015 and WDI 2013, which are expressed in constant 2011 and 2005 PPP dollars, respectively. The World Bank starts with the PPPs for a benchmark year and extrapolates backwards and forwards in time by using rates of inflation experienced over that time period, relative to the United States. For example, suppose that \(PPP_t^c\) is the PPP for country \(c\) in benchmark year \(t\). Then in year \(t + \tau\), this country’s \(PPP_{t+\tau}^c\) will be:

\[
PPP_{t+\tau}^c = PPP_t^c \cdot \left( \frac{G_{t+\tau}^c}{G_{US}^c} \right) \frac{G_{t+\tau}^c}{G_{US}^c}
\]

(51)

Here \(G_t^c\) is the GDP deflator for country \(c\) in period \(t\). The WDI tables are tethered to the most recent round of the ICP. Whenever a new ICP round is released, the WDI are updated (i.e. \(PPP_t^c\) is changed) and data from older rounds is discarded. In other words, in WDI 2015, the PPP GDP for 2005 is extrapolated backwards from the 2011 ICP and does not use the 2005 ICP data at all.

\(^{23}\)The details of computing PPPs, which are an example of a price index, are nuanced and the reader is referred to World Bank (2013) for an in-depth discussion of the available choices and their advantages and disadvantages.
4.1.3 PWT

I use two different series of data provided in the PWT. The PPP-adjusted GDP series is given by the variable \( RGDPE \) in PWT 9.0 and PWT 8.1 and by the variable \( RGDPCH \) in PWT 7.1. The PWT approach in general can be considered a fine-grained version of the WDI approach wherein the PPPs for each component of GDP are extrapolated based on inflation rates. By definition, GDP is the sum of consumption (C), investment (I), government expenditures (G), and the trade balance (exports (X) minus imports(M)):

\[
GDP = C + I + G + (X - M)
\]  

(52)

These \( K = 5 \) items are provided annually by statistical offices of each country. Using the ICP data for the benchmark year \( t \) and country \( c \), it is possible to compute the PPP value of each component \( K \) of GDP \( (PPP^K_{ct}) \) in the benchmark year. Then the PPP for component \( K \) in country \( c \) at a non-benchmark year \( t + \tau \) is given in the PWT as:

\[
PPP^K_{c,t+\tau} = PPP^K_{c,t} \cdot \frac{P^{NA,K}_{c,t+\tau}}{P^{NA,K}_{c,t}}
\]  

(53)

Here \( P^{NA,K}_{c,t} \) is the price deflator of a given component \( K \) of the national accounts. The second major difference between PWT and WDI is that starting with PWT 8.0, multiple benchmarks are used to compute PPPs for any given non-benchmark year. An example by Feenstra et al. (2015) should serve to illustrate the process. Suppose that a country participates in the 2005 and 2011 ICP benchmarks. For this country, the PWT creates a complete time-series of PPP from, say, 2000 till 2016 as follows: for years prior to 2005 and subsequent to 2011, data is extrapolated based on the formula in (53), using the “closest” ICP round (so in this example ICP 2005 would be used for years 2004 and prior, and ICP 2011 would be used for years 2012 onwards). For the intervening years 2006-2010, the PWT interpolates using data from both benchmarks, as follows:

\[
PPP^K_{c,t} = PPP^K_{c,2005} \times \frac{P^{NA,K}_{c,t}}{P^{NA,K}_{c,2005}} \left( \frac{2011 - t}{2011 - 2005} \right) + PPP^K_{c,2011} \times \frac{P^{NA,K}_{c,t}}{P^{NA,K}_{c,2011}} \left( \frac{t - 2005}{2011 - 2005} \right)
\]  

(54)

The equation (54) imposes the constraint that extrapolated prices must match benchmark prices in
the benchmark year. In PWT 7.1 and prior vintages, data from a single ICP round was extrapolated forwards and backwards, and is therefore somewhat closer to the WDI.

4.1.4 UQICD

I use the $CGDP_{PC}$ series from UQICD 2.1. The construction of the UQICD is similar to the PWT, but an additional layer of sophistication is added. Rather than only using data from adjacent ICP benchmarks in a given non-benchmark year, the UQICD panel dataset computes a weighted-average of extrapolations from all available benchmarks. Conceptually, the extrapolations are the same as PWT and WDI in that price deflators for components of GDP are used to backcast and forecast benchmark PPPs, however: (1) unlike the PWT where the weight assigned to a given benchmark is a function of the distance to it in time, the UQICD panel determines these weights endogenously using Kalman filters; (2) the UQICD panel does not constrain extrapolated prices to match benchmark year prices. As such, in one sense it is closer to PWT 7.1 and WDI, because implicitly their extrapolations are similarly unconstrained. However, the UQICD is similar to PWT 8.1 and PWT 9.0 in that multiple benchmarks are used. Rao et al. (2010) provide details on the econometric formulation for generating weights for the series in this way.

4.1.5 Night-lights

Data on night-time luminosity are taken from Henderson et al. (2012), and the reader is referred to their paper for details on the construction. Briefly, they take processed data of satellite imagery from the US Air Force Defense Meteorological Satellite Program (DMSP) as provided by the National Oceanic and Atmospheric Administration. The processed dataset represents light intensity as a six bit digital number for a 0.86 square kilometer-sized grid. The digital number is an integer ranging between 0 (no light) to 63 (maximum sensor intensity). Shapefiles representing the geographic boundaries of the countries of the world are applied to the satellite images and for each country-year observation, the area-weighted digital number is calculated. The final series from their dataset that I use in this analysis is the log of the area-weighted average digital number for each cell assigned to a country-year observation.
4.2 Results

In tables (4), (5), and (6) I present the results of weights assigned to the selected data series by the MSPE, MVAR and PS estimators. To calculate the weights, I use first-differences for both log lights and log GDP as my input vectors, as follows:

\[ X_j = \ln \frac{X^it_j}{X_j^{i,t-1}} \]  
\[ Y = \ln \frac{Y^it}{Y^{i,t-1}} \]  

(55)

(56)

In the above expression, the subscript \( j \) refers to a given dataset, while \( i \) indexes countries and \( t \) indexes the year. For small changes, the expressions for \( X_j \) and \( Y \) are approximately equal to percent-changes in GDP and night-lights. I prefer this specification over using the logarithm of growth rates of GDP and night-lights for two reasons. First, it ensures that there are no country-year observations where we must compute the logarithm of a negative growth rate. Moreover, the partial correlation between GDP and night-lights is highest when both variables are logged, as shown by Henderson et al. (2012).

4.2.1 WDI 2015 vs. WDI 2013

In the top row of Table (4) I compare WDI 2015 (based on the 2011 ICP) against WDI 2013 (based on the 2005 ICP). Because the only difference between WDI 2015 and WDI 2013 is the version of the ICP used, this comparison is really an assessment of which ICP round is the more accurate reflection of the true trends in economic output. For the 2005 ICP round, there was concern that the price imputation for non-participating countries may have biased the PPP estimates downward. Inklaar and Rao (2015) conduct a counterfactual analysis in which ICP 2011 methodology is applied to ICP 2005 data and find that these implied prices suggest much smaller changes between 2005 and 2011 ICP as compared to what was actually observed when results of ICP 2011 were published. Given this, I would expect two outcomes from the comparison of WDI 2015 against WDI 2013: (1) greater weight is assigned to WDI 2015 by both MVAR and MSPE because the ICP 2011 data is more accurate; (2) there is a large disparity between MVAR and MSPE weights because the systematic biases of ICP 2005 discussed by Inklaar and Rao (2015) imply that \( m_1 \neq m_2 \).
However, these are not the trends we see in table (4). According to MSPE, WDI 2015 and WDI 2013 both receive about equal weight, while MVAR and PS respectively estimate the weight for WDI 2015 at about 37% and 30%. The difference in MVAR and PS estimates suggests that WDI 2015 and WDI 2013 have different means. Though it doesn’t tell us the direction, this does confirm the view in the literature that ICP 2005 might have been biased. What is surprising, however, is that if we accept the assumptions of the MSPE estimator, then despite improved country coverage and better measurement, the ICP 2011 round is about as accurate as the 2005 round. The analogous analysis by Pinkovskiy and Sala-i Martin (2016b) finds WDI 2015 to receive almost a hundred percent of the weight relative to WDI 2013. But their conclusions are reversed to favor WDI 2013 when year and country fixed-effects are included, similar to what I observe with my MVAR estimator.\footnote{There are two important differences between the approach I take and the data and estimation procedures used by Pinkovskiy and Sala-i Martin (2016b). First, they use five additional years of night-time luminosity in their calculations (up to 2013). Second, they use a log-level regression (variously including country and year fixed-effects) for both GDP and night lights. I refrain from using fixed-effects because they tend to inflate variance and violate monotonicity of coefficient attenuation in classical measurement-error models.}

The greater weight assigned to WDI 2013 in my analysis may be explained, in part, by the fact that I only include country-year observations up to 2008 because that is the last year for which I have comparable data on night-lights. Moreover, I only use the subset of observations which are non-missing across all three datasets. This can potentially favor WDI 2013 over WDI 2015 because the WDI methodology does not constrain itself to using multiple ICP benchmarks, and the further we move from a given ICP year, the less reliable that data gets. As a result a substantial portion of the newer (and presumably more reliable) part of the WDI 2015 sample is omitted in my analysis. To test this hypothesis, I re-run the MVAR estimates of the weights – which don’t require night-lights – on all available data from WDI 2015 and WDI 2013, but I find that WDI 2015 still gets a lower MVAR weight (around 0.22) as compared to WDI 2013 (around 0.78).

Overall, this suggests that newer vintages of the WDI are not necessarily better than older vintages for calculating growth rates of GDP.

4.2.2 PWT 9.0 vs. PWT 8.1 vs. PWT 7.1

The second through the fourth rows of table (4) present head-to-head comparisons of the PWT vintages. The MSPE weights are about equal for PWT 9.0 and PWT 8.1. This is not surprising given that both vintages use the same PPP extrapolation methods. Moreover, as shown in WDI
comparisons, it appears that ICP 2011 is not unambiguously better than ICP 2005, and so PWT 9.0 (which incorporates the new ICP 2011 data) does not have a clear advantage over PWT 8.1 (which only includes data up to ICP 2005). The differences between the MVAR versus the PS weights are also small, suggesting that the PWT 8.1 and PWT 9.0 have similar means. This bodes well for their use in economic research because unlike vintages prior to PWT 7.1, the results here indicate that the data are at least consistent between revisions.

When both PWT 9.0 and PWT 8.1 are compared against PWT 7.1, the latter receives between twice (MSPE) to six times (MVAR) as much weight as either of the newer vintages. This is surprising because one would expect the multiple benchmark procedure used in both PWT 9.0 and PWT 8.1 to deliver more accurate results. Similar results are obtained by Pinkovskiy and Sala-i Martin (2016b), who additionally subject PWT 7.1 to greater scrutiny against other vintages of the PWT. They find that the only head-to-head comparison in which PWT 7.1 is unambiguously worse than another PWT vintage is if it is compared against PWT 8.1 in the sub-sample of post-Communist countries.

The good news is that MVAR and PS estimates appear to be about equal across all the comparisons, which suggests that revision-related biases appear to have been resolved starting with PWT 7.1. Nevertheless, it is puzzling why the newer PWT vintages do not perform better than the older vintage.

In rows five through eight, the PPP-adjusted GDP numbers in each vintage are compared against their national-accounts-based counterparts. Specifically, this set of comparisons is between $RGDP_{E}$ and $RGDP_{NA}$ for PWT 9.0 and PWT 8.1, and $RGDP_{CH}$ and $RDGPL2$ for PWT 7.1. These comparisons serve to highlight the advantage (if any) of PPP-adjusted economic output over the more coarse-grained inflation-adjusted output from national accounts. As we discussed earlier and as noted by Feenstra et al. (2015), one of the main advantages of PPP-based GDP numbers is that comparisons can be made between countries over time, while national accounts numbers may be more suitable for cross-sectional comparisons of countries at one point in time.

The results in table (4) are mixed, however: for both PWT 9.0 and PWT 8.1, the national accounts time-series get between two thirds to almost full weight, relative to the PPP-based series. However, in PWT 7.1, the PPP-based GDP receives between one-third to the same weight as the national accounts-based GDP using the PS and MSPE estimators, respectively. From my investigation of the documentation for PWT 7.1., it is unclear why this series is so robust. The methodology
from PWT 7.1 has only been improved in subsequent PWT rounds so perhaps there was something unique about the data collection in national in that year which leads to this robustness. This should be investigated further.

In the main, however, these results suggest a role for national accounts-based measures of GDP. Despite the added effort of computing price equivalences, it seems that the coarse measure of economic activity as captured by national accounts is adequate.

4.2.3 PWT vs. WDI

The series of comparisons in Table (5) assesses vintages of the PWT (both the PPP-based figures and the national accounts-based figures) against the WDI. A similar question was considered in the influential study by Johnson et al. (2013), who concluded that WDI is the preferred source of GDP information because the extrapolations conducted in PWT introduce too much error. We can take a use our econometric approach to identify whether WDI or PWT is more accurate.

The left panel of the table compares PPP-based numbers from the PWT against WDI. In head-to-head comparisons against both WDI 2015 and WDI 2013, all three estimators assign lower weight to PWT 9.0 and PWT 8.1 (with wide discrepancies between the estimators). On the other hand, PWT 7.1 is assigned roughly equal weight by both MSPE and PS estimators, and about a third of the weight per the MVAR estimator.

When the national-accounts based GDP estimates are used from the PWT, as expected, the differences with respect to the WDI are minimized. In these comparisons, all three estimators assign both series about equal weight, and in certain cases PWT even receives more weight.

Thus, it seems that as Johnson et al. (2013) concluded, the WDI figures appear to be more accurate than the PPP-based numbers in PWT 9.0 or PWT 8.1 figures. However, the PPP-based numbers in PWT 7.1 are roughly equally as accurate as WDI. Morever, the discrepancy between the weights assigned by MVAR versus the MSPE estimators suggests that the mean values of GDP are quite different between estimates compiled by WDI and PWT.

4.2.4 UQICD vs. all

In table (6), the UQICD panel is compared against PWT and WDI. Based on the MSPE estimates, the UQICD series are accorded more weight in head-to-head comparisons against all series except
for PWT 7.1, for which both series receive about equal weight. The superior performance of the UQICD to WDI 2015 and 2013 is of note, as it is the only PPP-based series in this entire analysis to receive more weight than WDI estimates. This is a promising result for two reasons: first, it shows that PPP adjustments do not lead to some systematic bias in GDP numbers (which is the conclusion reached by Johnson et al. (2013) using only PWT data); second, it shows that prior benchmarks are informative regarding income estimates and that a sensible aggregation procedure can be used to integrate that prior information into current estimates. We also notice that MVAR and MSPE estimates differ from each other only when UQICD is compared against a WDI vintage. This discrepancy between estimators when WDI vintages are compared against PPP-based vintages has been persistent in this analysis. There are likely theoretical reasons, perhaps related to how WDI is constructed which leads to consistent deviations from PPP-based figures and further investigations are warranted.

5 Conclusion

In this paper I developed two estimators for computing a statistically optimal weighted average of GDP time-series data. The magnitude of the weights is interpreted as being proportional to the accuracy of a given time-series. The first approach (MSPE) derived these weights by minimizing the mean-squared prediction error a given GDP time-series and the true value of GDP. The second approach (MVAR) derived these weights by minimizing the variance of the convex combination of different GDP estimates. In Monte Carlo studies of both estimators I find that the MSPE approach yields lower bias and mean squared error than the MVAR approach. However, in conditions where data accuracy is low, the performance of the MVAR estimator is close to the MSPE estimator. Applying these estimators to GDP panels datasets I find that a new panel dataset of GDP from the University of Queensland is more accurate than GDP estimates in the two most recent vintages of the World Development Indicators, as well as the three most recent vintages of the Penn World Tables.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\sigma^2_X = 0.5$</th>
<th>$\sigma^2_X = 1.0$</th>
<th>$\sigma^2_X = 2.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_{u_1 u_2}$</td>
<td>0.208 0.222 0.231</td>
<td>0.211 0.217 0.232</td>
<td>0.207 0.202 0.233</td>
</tr>
<tr>
<td>$\sigma^2_{u_1 \nu}$</td>
<td>0.194 0.184 0.162</td>
<td>0.200 0.170 0.161</td>
<td>0.212 0.158 0.162</td>
</tr>
<tr>
<td>$\sigma^2_{u_2 \nu}$</td>
<td>0.204 0.202 0.187</td>
<td>0.210 0.188 0.189</td>
<td>0.222 0.162 0.186</td>
</tr>
<tr>
<td>$\sigma^2_{X u_1}$</td>
<td>0.187 0.179 0.168</td>
<td>0.191 0.166 0.165</td>
<td>0.198 0.152 0.164</td>
</tr>
<tr>
<td>$\sigma^2_{X u_2}$</td>
<td>0.193 0.186 0.177</td>
<td>0.201 0.178 0.178</td>
<td>0.201 0.157 0.177</td>
</tr>
<tr>
<td>$\sigma^2_{X \nu}$</td>
<td>0.199 0.191 0.179</td>
<td>0.200 0.181 0.174</td>
<td>0.206 0.161 0.173</td>
</tr>
<tr>
<td>$m_1$</td>
<td>0.157 -0.514 0.261</td>
<td>0.113 -0.696 0.272</td>
<td>0.016 -0.802 0.278</td>
</tr>
<tr>
<td>$m_2$</td>
<td>0.177 -0.486 0.275</td>
<td>0.122 -0.652 0.295</td>
<td>0.014 -0.768 0.304</td>
</tr>
<tr>
<td><strong>Average Bias</strong></td>
<td>0.169 0.181 0.188</td>
<td>0.165 0.204 0.189</td>
<td>0.162 0.248 0.190</td>
</tr>
</tbody>
</table>

Table 1: Results from the log-log regression of bias versus simulation parameters with $N = 20,000$ Monte Carlo runs.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\sigma^2_X = 0.5$</th>
<th>$\sigma^2_X = 1.0$</th>
<th>$\sigma^2_X = 2.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_{u_1 u_2}$</td>
<td>0.116 0.114 0.127</td>
<td>0.112 0.104 0.124</td>
<td>0.105 0.078 0.119</td>
</tr>
<tr>
<td>$\sigma^2_{u_1 \nu}$</td>
<td>0.019 0.018 0.007</td>
<td>0.019 0.007 0.005</td>
<td>0.017 -0.017 0.0006</td>
</tr>
<tr>
<td>$\sigma^2_{u_2 \nu}$</td>
<td>0.019 0.018 0.007</td>
<td>0.019 0.007 0.006</td>
<td>0.017 -0.017 0.0015</td>
</tr>
<tr>
<td>$\sigma^2_{X u_1}$</td>
<td>0.019 0.016 0.014</td>
<td>0.018 0.005 0.010</td>
<td>0.014 -0.019 0.004</td>
</tr>
<tr>
<td>$\sigma^2_{X u_2}$</td>
<td>0.020 0.017 0.014</td>
<td>0.018 0.006 0.011</td>
<td>0.015 -0.017 0.005</td>
</tr>
<tr>
<td>$\sigma^2_{X \nu}$</td>
<td>0.020 0.017 0.014</td>
<td>0.018 0.007 0.010</td>
<td>0.014 -0.016 0.004</td>
</tr>
<tr>
<td>$m_1$</td>
<td>-0.004 -0.047 -0.004</td>
<td>-0.006 -0.083 0.002</td>
<td>-0.019 -0.128 0.012</td>
</tr>
<tr>
<td>$m_2$</td>
<td>-0.008 -0.043 -0.002</td>
<td>-0.004 -0.077 0.008</td>
<td>-0.019 -0.120 0.019</td>
</tr>
<tr>
<td><strong>Average MSE</strong></td>
<td>0.697 0.716 0.739</td>
<td>0.710 0.811 0.765</td>
<td>0.745 1.09 0.816</td>
</tr>
</tbody>
</table>

Table 2: Results from the log-log regression of mean-squared error (MSE) versus simulation parameters with $N = 20,000$ Monte Carlo runs.
<table>
<thead>
<tr>
<th>Dataset</th>
<th>Number of Countries and/or Regions</th>
<th>Time Period</th>
<th>ICP Round</th>
</tr>
</thead>
<tbody>
<tr>
<td>World Development Indicators 2015&lt;sup&gt;a&lt;/sup&gt;</td>
<td>214</td>
<td>1990-2015</td>
<td>2011</td>
</tr>
<tr>
<td>World Development Indicators 2013&lt;sup&gt;a&lt;/sup&gt;</td>
<td>220</td>
<td>1980-2012</td>
<td>2005</td>
</tr>
<tr>
<td>Penn World Tables 9.0&lt;sup&gt;b&lt;/sup&gt;</td>
<td>182</td>
<td>1950-2014</td>
<td>2011 (interpolated)</td>
</tr>
<tr>
<td>Penn World Tables 8.1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>167</td>
<td>1950-2011</td>
<td>2005 (interpolated)</td>
</tr>
<tr>
<td>Penn World Tables 7.1</td>
<td>190</td>
<td>1950-2010</td>
<td>2005</td>
</tr>
<tr>
<td>UQICD 2.1</td>
<td>181</td>
<td>1970-2012</td>
<td>All</td>
</tr>
<tr>
<td>Night-lights (Henderson et al. (2012))</td>
<td>228</td>
<td>1992-2008</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table 3: (a) The world development indicators also provide aggregated data for various country groupings based on geography and country characteristics, which is why the number of countries and regions reported there is higher. (b) PWT 9.0 and PWT 8.1, are not based on a single round of the ICP but rather interpolate between adjacent ICP rounds, as described in the data section.

<table>
<thead>
<tr>
<th>Data Series</th>
<th>MSPE  ( \lambda^*_E )</th>
<th>MVAR  ( \lambda^*_V )</th>
<th>PS  ( \lambda^*_PS )</th>
</tr>
</thead>
<tbody>
<tr>
<td>WDI 2015 vs. WDI 2013</td>
<td>0.48</td>
<td>0.37</td>
<td>0.31</td>
</tr>
<tr>
<td>PWT 9.0 vs. PWT 8.1</td>
<td>0.50</td>
<td>0.55</td>
<td>0.50</td>
</tr>
<tr>
<td>PWT 9.0 vs. PWT 7.1</td>
<td>0.32</td>
<td>0.15</td>
<td>0.18</td>
</tr>
<tr>
<td>PWT 8.1 vs. PWT 7.1</td>
<td>0.32</td>
<td>0.16</td>
<td>0.20</td>
</tr>
<tr>
<td>PWT 9.0 (PPP) vs. PWT 9.0 (NA)</td>
<td>0.31</td>
<td>-0.03</td>
<td>0.17</td>
</tr>
<tr>
<td>PWT 8.1 (PPP) vs. PWT 8.1 (NA)</td>
<td>0.29</td>
<td>-0.01</td>
<td>0.15</td>
</tr>
<tr>
<td>PWT 7.1 (PPP) vs. PWT 7.1 (NA)</td>
<td>0.49</td>
<td>-0.22</td>
<td>0.31</td>
</tr>
</tbody>
</table>

Table 4: Comparison of the newer versus older vintages of the World Development Indicators (WDI), and the Penn World Tables (PWT). WDI 2015 is based on the 2011 ICP, while WDI 2013 is based on the 2005 ICP. PWT 9.0 and PWT 8.1 are based on the 2011 ICP, while PWT 7.1 is based on the 2005 ICP. The last three rows of the table compare PPP-adjusted GDP time-series in the PWT (\( RGDPE \) in PWT 9.0/8.1 and \( RGDPCH \) in PWT 7.1) against the national-accounts based data in the same (\( RGDPNA \) in PWT 9.0/8.1 and \( RGDPL2 \) in PWT 7.1)
<table>
<thead>
<tr>
<th>Data Series</th>
<th>PPP-adjusted GDP</th>
<th>National Accounts GDP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSPE $\lambda_E^<em>$ MVAR $\lambda_V^</em>$ PS $\lambda_{PS}$</td>
<td>MSPE $\lambda_E^<em>$ MVAR $\lambda_V^</em>$ PS $\lambda_{PS}$</td>
</tr>
<tr>
<td>PWT 9.0 vs. WDI 2015</td>
<td>0.30 0.02 0.17</td>
<td>0.48 0.62 0.32</td>
</tr>
<tr>
<td>PWT 9.0 vs. WDI 2013</td>
<td>0.29 0.03 0.16</td>
<td>0.47 0.53 0.20</td>
</tr>
<tr>
<td>PWT 8.1 vs. WDI 2015</td>
<td>0.30 0.05 0.19</td>
<td>0.52 0.64 0.60</td>
</tr>
<tr>
<td>PWT 8.1 vs. WDI 2013</td>
<td>0.29 0.05 0.18</td>
<td>0.50 0.56 0.47</td>
</tr>
<tr>
<td>PWT 7.1 vs. WDI 2015</td>
<td>0.48 0.22 0.50</td>
<td>0.49 0.25 0.53</td>
</tr>
<tr>
<td>PWT 7.1 vs. WDI 2013</td>
<td>0.46 0.21 0.45</td>
<td>0.47 0.25 0.47</td>
</tr>
</tbody>
</table>

Table 5: Comparison of the World Development Indicators (WDI) against the Penn World Tables (PWT).

<table>
<thead>
<tr>
<th>Data Series</th>
<th>MPSE ($\lambda_E^*$)</th>
<th>MVAR ($\lambda_V^*$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UQICD 2.1 vs. WDI 2015</td>
<td>0.59</td>
<td>0.20</td>
</tr>
<tr>
<td>UQICD 2.1 vs. WDI 2013</td>
<td>0.53</td>
<td>0.19</td>
</tr>
<tr>
<td>UQICD 2.1 vs. PWT 9.0</td>
<td>0.80</td>
<td>0.80</td>
</tr>
<tr>
<td>UQICD 2.1 vs. PWT 8.1</td>
<td>0.78</td>
<td>0.79</td>
</tr>
<tr>
<td>UQICD 2.1 vs. PWT 7.1</td>
<td>0.55</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Table 6: This table compares UQICD 2.1 versus all other series in the dataset.
A Appendix

To prove that any valid IV is a valid orthogonal proxy, I will show that (22) can be identified using any valid IV for X. Let Y be a valid instrument for the true income X, then the following conditions must hold. Cov(Y, u_{i \in 1,2}) = 0 (exogeneity), and Cov(Y, X) ≠ 0 (relevance). Therefore σ_{YX_i} = Cov(Y, m_iX + u_i) = m_iσ_{YX}. Now, consider (22), and plug-in σ_{YX_i} in place of σ_{XX_i}:

\[ \lambda_E^* = \frac{2\sigma^2_{X_2}\sigma_{YX_1} - \sigma_{X_1X_2}\sigma_{YX_2}}{2\sigma^2_{X_1}\sigma_{YX_2} + 2\sigma^2_{X_2}\sigma_{YX_1} - \sigma_{X_1X_2}(\sigma_{YX_1} + \sigma_{YX_2})} \]

\[ = \frac{2\sigma^2_{X_2}m_1\sigma_{YX} - \sigma_{X_1X_2}m_2\sigma_{YX}}{2\sigma^2_{X_1}m_2\sigma_{YX} + 2\sigma^2_{X_2}m_1\sigma_{YX} - \sigma_{X_1X_2}(m_1\sigma_{YX} + m_2\sigma_{YX})} \]

\[ = \frac{(2\sigma^2_{X_2}m_1 - \sigma_{X_1X_2}m_2)\sigma_{YX}}{2\sigma^2_{X_1}m_2 + 2\sigma^2_{X_2}m_1 - \sigma_{X_1X_2}(m_1 + m_2)} \]

\[ = \frac{(2\sigma^2_{X_2}m_1 - \sigma_{X_1X_2}m_2)}{(2\sigma^2_{X_1}m_2 + 2\sigma^2_{X_2}m_1 - \sigma_{X_1X_2}(m_1 + m_2))} \]

Where the last equation is exactly equivalent to (23)
References


Small Differences: Discrepancies in GDP Statistics and Implications for Policy

Abdul A. Tariq*

Abstract

In this paper I compare stylized differences between the most recent vintages of three widely-used sources of GDP data: the Penn World Tables (PWT 9.0, PWT 8.1, and PWT 7.1), the World Development Indicators (WDI 2015 and WDI 2013), and the University of Queensland International Comparisons data (UQICD 2.1). I find that there are large and persistent differences in the growth rates and levels of GDP between these datasets. In particular, growth rate differences are larger for smaller economies, for countries located in sub-Saharan Africa and Latin America, and for country-year observations which are further away from the years of the ICP benchmark surveys. To demonstrate the potential consequences of variability in estimates of GDP levels, I simulate counterfactual predictions of aid allocations for the World Bank development assistance program (IDA) using alternative datasets. For most countries, the change in allocations is relatively inelastic with respect to the choice of dataset of income per capita, but for a few recipients aid allocations can change substantially. More perniciously, I note that uncertainty in GDP can potentially affect eligibility of countries to receive assistance in the first place. I offer some options to mitigate some of these concerns.

*RAND Corporation and Pardee-RAND Graduate School
1 Introduction

There are large disparities between estimates of gross domestic product (GDP) published by different sources. For example, Liberia is the second, seventh, or 22nd-poorest country in sub-Saharan Africa in terms of per capita GDP, depending on whether one uses data from the the Penn World Tables, the World Bank, or the Maddison project.\(^1\) Moreover, Liberia is not an isolated example. There are a number of countries for which estimates of the rate of growth are highly variable between datasets. Different sources disagree even on the direction of growth, much less the magnitude of the estimates.

These discrepancies are problematic because the effectiveness of research and policy for economic development and poverty alleviation depends upon accurate estimates of GDP. For example, as a high-level measure of the overall size and health of an economy, GDP enables policymakers to assess the impact of their interventions, and track long-run trends in growth. It informs the national budgeting process, and often the relative importance of particular government programs and policies is measured by their GDP share.

The numeric values of the estimates are also used directly in setting policy. For example, the International Monetary Fund (IMF) uses GDP estimates in calculating country quota shares. These quotas determine the amount a country will contribute to and be able to draw from the Fund. The World Bank uses gross national income (GNI) – a concept intimately tied to GDP through an accounting identity – to determine the eligibility and amount of the development assistance to provide to countries. The importance of GDP as a measure is summed-up by the economist Diane Coyle (Coyle, 2015):

“[This] single measure of “the economy” tends to dominate political contests, and governments’ fortunes seem to rise and fall with the difference between plus 0.2 percent and minus 0.1 percent in one quarter’s GDP numbers.”

Given its central role in economic policy, it is important to understand the nature, causes and consequences of statistical disparities in GDP estimates – issues that I take up in this paper. By understanding the magnitude of the differences and the contributing factors, steps can be taken to improve the quality of data, which will ultimately improve the efficacy of economic development efforts.

These issues have been examined previously in a small but growing literature which collectively concludes that estimates for the same country – from purportedly comparable sources – can deviate enough to paint a picture of entirely different economic realities (Bentzen, 2015; Ram and Ural, 2014; Johnson et al., 2013; Togo, 2011; Ram, 2009; Hanousek et al., 2008).

The literature takes two different approaches to investigate the causes. One group of studies uses statistical methods to uncover sources of discrepancies. For example, Johnson et al. (2013) conclude that for the Penn World Tables (PWT) the differences in adjacent versions are the result of three factors: (1)
changes in underlying national accounts data (newer PWT versions have more recent national accounts data); (2) economic size (poorer countries have higher inter-version variability because of greater uncertainty in their purchasing power parity (PPP) estimates); and (3) the extrapolation of PPPs from the International Comparison Program (ICP) to non-benchmark years (historic data becomes more variable in every PWT revision). Hanousek et al. (2008) make a similar point when comparing PPP-adjusted GDP in the PWT, the World Development Indicators (WDI), and the International Financial Statistics (IFS) from the IMF. They argue that the discrepancies arise because “data sets differ in whether and how they adjust for changes in relative prices across countries.”

The work of the economist Morten Jerven provides another approach for uncovering differences. He examiness the data collection processes and the in-depth history of growth statistics in several countries in sub-Saharan Africa, comparing the raw data from national statistical offices to the harmonized data published by WDI and PWT. In the sub-Saharan case, Jerven et al. (2016) supplies three over-arching reasons for the discrepancies: (1) revisions to GDP data from benchmark years; (2) clerical errors, oversight and lack of quality control when putting series together; and (3) political negotiations in putting growth statistics together. In particular, he finds that differences in estimates can result from differences in timing and procedures used by WDI and PWT for incorporating updates to the underlying data from country statistical offices. As a concrete example, Jerven (2011) notes that Tanzania revised its GDP accounting methods in 1997 to better reflect economic activity in its informal sectors. These revisions were reflected in subsequent versions of the WDI but not in the PWT. The reason for this oversight is unclear, but (Jerven, 2010; Jerven et al., 2016) suspect that PWT did not revise the data so as to not introduce breaks in the data. On the other hand WDI dropped data for Tanzania prior to 1987 (the earliest year for which Tanzania produced revised estimates). As Jerven et al. (2016) notes, the consequences of this were not insignificant: using PWT data from Tanzania, Durlauf et al. (2005) erroneously concluded that “negative output shocks are common in low income countries” because of the large drop in Tanzania’s GDP in 1987, the year in which the revised methodology was implemented.

In the spirit of Johnson et al. (2013) and Hanousek et al. (2008), the goal of this paper is to statistically characterize the differences in several different datasets of GDP time-series, which represent multiple institutional and methodological perspectives on producing growth estimates. In addition, I attempt to capture the policy consequences of these differences on the allocation of development assistance by the World Bank. Previous literature has focused almost exclusively on the impact of GDP uncertainty on the theory of economic growth. The policy simulation exercise done here attempts to quantify the effect on the policies for growth.

In the next section, I present a series of analyses of the differences between the growth rates and levels of GDP based on data from the World Bank, the Penn World Tables, and a new panel dataset from the
University of Queensland. I find that different estimates vary not just in magnitudes but even in the basic direction of growth (positive or negative) for up to ten percent of the set of all non-missing observations across the three sources of data, and up to 30% in low income countries. I find that PWT consistently over-estimates growth rates relative to WDI and UQICD, suggesting that there are methodological reasons for these disparities specific to a given dataset. The size of the economy, geography and frequency of price surveys appear to be the dominant contributing factors for inter-series variability. In section 3, I present estimates of the uncertainty in aid allocation that might result from uncertainty in gross national income (GNI). Even though the magnitude of the misallocations is small for most recipients of World Bank development assistance, for a few countries the differences are quite substantial. As such, I provide some suggestions for the World Bank to safeguard its aid allocation process against these problems. Section 4 concludes the discussion.

2 Comparison of GDP panels

The analyses in this paper are based on six different datasets of GDP at PPP: three datasets (WDI 2015, PWT 9.0, and UQICD 2.1) are based on PPP figures derived from the 2011 round of the ICP, while the other three (WDI 2013, PWT 8.1, PWT 7.1) are based on the 2005 ICP. The 2011 ICP was an improvement over the 2005 ICP in terms of country coverage and methodology (Deaton and Aten, 2014). Moreover, within each round of the ICP (also referred to as an ICP “benchmark”), compilers of WDI, PWT and UQICD differ in how they apply PPP conversion factors to the raw GDP data: all three use the same PPP in the benchmark year (in this case 2005 and 2011). For non-benchmark years, WDI extrapolates PPPs by assessing how much inflation a country experienced over that time period, relative to the United States. PWT uses growth rates of the components of national accounts (i.e. consumption, investment, and government expenditures) and interpolates between different ICP benchmarks (Feenstra et al., 2015). UQICD uses a similar procedure as PWT but rather than simply interpolating between benchmarks, a weighted average of data from different ICP rounds is used in which weights are determined endogenously to minimize the error between extrapolated and observed PPPs (Rao et al., 2010).

To the extent possible, the analysis proceeds *ceteris paribus*: I generally only make within-ICP comparisons so as to reduce any bias and confounding that might result from the intrinsic differences in the 2011 and 2005 ICP rounds. Moreover, unless noted otherwise, I only use country-year observations which are

---

2 The International Comparisons Program (ICP) is an international statistical effort for collecting and publishing data on prices of goods baskets in most countries of the world. The data on price and consumption of goods from ICP surveys are used to calculate PPP conversion factors through which GDP valued in local currencies can be expressed in internationally comparable currency units.

non-missing across all three series based on a given ICP group. While this reduces the sample size (because WDI has shorter temporal coverage than both UQICD and PWT), it ensures that comparisons are based on data from a consistent period in time. There is an unfortunate but necessary trade-off here. On the one hand, having comparable sample sizes and countries across the different series gives more meaning to the comparisons.\footnote{Even simple statistics such as correlation coefficients between two datasets are irrelevant if the length and coverage of datasets in one pairwise comparison is different from that in another. For example, since sample variance depends on sample size, even if all else is equal, the longer sample (with more observations) would have lower variance than a shorter sample.} On the other hand, only using “complete” country-year observations (i.e. observations non-missing across all three sources) might induce sample selection effects wherein, say, observations dropped in one series but not another are indicative of important characteristics. Nevertheless, in this analysis comparability takes precedence over selection effects because generalizability is less important than understanding precisely why the differences occur in the first place.

2.1 Inter-series differences in the direction of GDP growth

In table (1) and (2) I present the panel-average growth rates of sub-Saharan countries, as computed in the ICP 2011 and the ICP 2005-based datasets, respectively. It is clear that the same set of ICP data leads to widely different estimates of growth rates between the different series. The Republic of Congo, for example, had an impressive average growth rate of more than six percent according to PWT 9.0, yet according to both WDI 2015 and UQICD 2.1, it’s performance was more mediocre, averaging less than 0.5 percent over the same period. These numbers paint vastly different economic realities.

An even more pernicious example is Zambia. Based on ICP 2005 growth rates presented in table (2), we see that Zambia grew by between 1 and 3 percent on average according PWT 8.1 and PWT 7.1. Yet over the same period it \textit{shrunk} by close to 0.3 percent on average according to WDI 2013. Jerven (2010) notes that there were changes in Zambia’s official accounting practices in the 1990s in response to structural adjustments in the economy.\footnote{There were also changes in the base years used to compute the national accounts data in 1970 and 1976, but these predate the sample period.} PWT and WDI might take different approaches for incorporating these historical changes – they are different organizations with possibly different data philosophies – which might explain the discrepancy in this case. However, even \textit{within} data from the same source, there are differences for which a similar explanation is implausible. For example, according to PWT 7.1 Cameroon averaged a growth rate of $-0.566\%$ over the sample period, but according to PWT 8.1 it grew by $0.866\%$. These instances are all the more problematic because the growth rates are computed using PPP-adjusted GDPs which, in theory, ought to be more stable than non-adjusted figures (Feenstra et al., 2015).

Zambia and Cameroon are not the only pathological cases. We see that twelve countries in table (1) and eight countries in table (2) have growth rates of opposite signs. Of these, four are common to both lists: Central African Republic, Gabon, Guinea, and Guinea-Bissau. This suggests that for some countries, data
quality issues can persist over time, even after newer survey data is available. Lastly, the fact that there are a larger number of countries with inconsistent growth rates in the newer ICP 2011 (table 1) as compared to the older ICP 2005 (table 2) is as puzzling as it is problematic.

Similar to the analysis undertaken in Hanousek et al. (2008), tables (3) and (4) report the fraction of the sample in which a given pair of series report opposite signs on the average rate of growth over the period covered in the panel. The results are tabulated for the full sample, and also disaggregated by World Bank income classifications. For the ICP 2011-based series, table (3) shows that PWT 9.0/UQICD 2.1, and PWT 9.0/WDI 2015 disagree on the direction of growth for nearly one-tenth of the sample overall. The sign on the average growth rate is more consistent between WDI 2015 and UQICD 2.1, where only five percent of the sample reports opposite signs overall. The level of agreement in growth directions clearly depends on country incomes: between a fifth to a third of the low-income countries have opposite signs across all pairwise comparisons of ICP 2011-based datasets, whereas nearly all high-income countries have the same growth direction. Comparing the results of table (4) to table (3), we find that the ICP 2005-based datasets have fewer disagreements – about half that of ICP 2011-based datasets. Although both groups are more consistent on the higher end of the income classification.

2.2 Inter-series differences in the GDP growth rates and levels

To make the discussion more precise, in this section I report the results of statistical tests of equality of GDP growth rates and levels, and the cross-correlations of growth rates and levels between the different series. For growth rates the results for ICP 2011 and ICP 2005-based datasets are presented in tables (5) and (6), respectively.

The first column of table (5) reports the difference in means between WDI 2015 and PWT 9.0. The estimate of the difference is negative for the full sample, as well as for the sub-samples by income group. This means that WDI 2015 tends to understate growth rates relative to PWT 9.0. The differences are statistically significant for all except the low income countries, where the magnitude of the difference is also smaller.

In the second column of table (5), I report the comparison between WDI 2015 and UQICD 2.1. The overall difference in growth rates between the two series is small but statistically significant, and WDI 2015 tends to underestimate growth rates relative to UQICD 2.1. The magnitude of the differences decreases as one moves to higher income countries, and the differences are statistically insignificant for all but the upper middle income group.

The third column of table (5) compares PWT 9.0 with UQICD 2.1 and finds that these two series have the largest overall difference in growth rates, with PWT 9.0 reporting larger average rates of growth than UQICD 2.1 in general. Moreover, the differences are roughly equal in size across the income groups (though
not statistically significant for the low income group).

I persistently find a lack of a statistically significant difference in the low income group in all pairwise growth rate comparisons (for both ICP 2011 and ICP 2005-based datasets), except WDI 2015 vs. UQICD 2.1. This finding goes against the putative notion that there must be greater inter-series variability among lower income countries. Moreover, I find that two plausible reasons for this lack of significance are both unsupported in the data. First, the lack of significance does not result from lack of statistical power. One might suspect that among low income countries, there is larger measurement error, which leads to loss of significance. However, in preliminary analyses (not reported here) when I compute estimates of the measurement error, conditional on the income grouping, I do not find that the lower income groups are necessarily more error prone.\(^6\)

The lack of significance also does not seem to result from sample selection effects. In other words, it is not the case country-year observations are dropped more frequently (to achieve inter-series completeness) among the low income group. In table (9) and table (10), I report the fraction of observations per income group, and fraction of complete observations per income group, for ICP 2005 and ICP 2011, respectively. From these tables it is clear that the fraction of complete observations is roughly equal across income groups. This finding of the lack of significance in lower income countries warrants further investigation.

The fourth through sixth columns of table (5) report correlations between the series. Overall, UQICD 2.1 and WDI 2015 are quite tightly correlated while PWT 9.0 is less correlated with either series. This runs counter to expectations because UQICD 2.1 and PWT 9.0 are methodologically more similar to each other than either is to WDI 2015. The pattern of correlations by income group is also interesting: while cross-correlation increases as a function of income, the series are least correlated for the lower middle, rather than the low income group of countries. One might speculate that perhaps countries on either end of the income range (low or high) are on their balanced growth path, whereas the lower middle income group is comprised of transition economies which have not yet settled into equilibrium. Because transition might be captured differentially in each series, and might be harder to characterize in general, the temporal correlation between series for these countries is lower. I test for this by comparing the fraction of countries which retain their income group classification between WDI 2013 and WDI 2015. The idea is that country classifications should be more stable over time (i.e. from WDI 2013 to WDI 2015) for the high and low income groups, but less so for the lower middle income group. I find that the fraction of observations which don’t change their income grouping from WDI 2013 to WDI 2015 is essentially similar: low income (85% retain ranking), lower middle income (92% retain ranking), upper middle income (95% retain ranking), and high income (96% retain ranking).

\(^6\)Based on the approach of Blackwell et al. (2015), I compute estimates of the measurement error as follows. Let \(X_1 = X + u_1\) and \(X_2 = X + u_2\) represent each of the datasets (say WDI 2015 and PWT 9.0). Assume that: \(\text{Cov}(X_1, u_1|W) = 0\), where \(W\) represents income groupings, and \(\text{Cov}(u_1, u_2) = 0\), and \(E(u_i) = 0\). Then the estimate of measurement error in series 1 (say WDI 2015) will be \(\text{Var}(u_1|W) = \text{Var}(X_1|W) - \text{Cov}(X_1, X_2|W)\).
The analysis of growth rates and cross-correlations for the set of series based on ICP 2005 is presented in table (6). In the first two columns, WDI 2013 is compared with PWT 8.1 and PWT 7.1, respectively. Generally speaking the World Bank series estimates of the average growth rates are statistically significantly lower than the Penn World Tables, as was the case with the ICP 2011-based series. However the differences are significant (and larger) when WDI 2013 is compared against PWT 8.1 than when it is compared against PWT 7.1. In fact, PWT 7.1 and WDI 2013 appear to be statistically indistinguishable overall and for each of the income groups. When the PWT 8.1 and PWT 7.1 are compared, the more recent vintage has higher average rates of growth than the older vintage. The difference is statistically insignificant in low and lower middle income groups, while it is significant in upper middle and high income groups. As before, when looking at correlations between the different series, WDI 2013 and PWT 7.1 are much more tightly correlated than either of the series is with PWT 8.1. Correlations trend positively with income level, however, as before, the lowest correlations are noted for the lower middle income group.

In general, growth rates are more prone to errors so I also compare inter-series differences in GDP levels. Tables (7) and (8) report the average difference in GDP levels, as a fraction of the average GDP in the corresponding income group. In table (7) we see that for the ICP 2011-based datasets, all but one of the pairwise comparisons by income group are statistically significant and the differences range between 9 and 30 percent. WDI 2015 and UQICD 2.1 have the largest differences in GDP levels, which is opposite to their high degree of agreement in growth rates. The overall correlation between the levels series is quite high. Even when disaggregated by income group, the correlations are typically higher than 0.8 for all but the low income countries. The average per capita GDP estimates are highest for WDI 2015, followed by PWT 9.0.

The GDP levels comparisons for ICP 2005-based series are presented in table (8). For the overall sample, the inter-series differences in levels are smaller than for the ICP 2011-based series. When disaggregated by income groups, the differences for the low income group of countries are quite significant, ranging between 10 to 40 percent in absolute value; however, for the other income groups, the differences are much smaller and range between one and seven percent in absolute value. Moreover, no single series consistently under or over estimates GDP levels relative to another. The correlations between the ICP 2005 series are also higher than what was observed for the ICP 2011-based series.

The foregoing analysis thus presents a picture of notable differences between datasets. In the main, differences between GDP time-series are prominent in both growth rates and levels, regardless of the round of the ICP. The magnitude of these differences is sometimes positively and at other times negatively correlated with the income level of a country. That said, lower middle income countries generally appear to have the least consistent data. In a not insignificant portion of the sample, different series suggest opposing trends in growth, and these inconsistencies persist over subsequent rounds of the ICP. Methodological similarities in

---

7I thank Krishna Kumar (RAND) for pointing out the reasoning that error in even a single year can propagate over time when calculating growth rates but not levels.
the construction of GDP datasets do not appear to confer data consistency: for example, adjacent vintages of the PWT are less similar to each other than to the WDI. A puzzling observation for future research is that series tend to be indistinguishable from each other when looking at the sub-sample of low income countries, despite our intuition that data quality is inversely related to income. This finding does not appear to be driven by lack of statistical power or bias from sample selection. Correlations in GDP levels are higher, but the differences in point estimates of the means are still quite significant, although smaller for most income groups in the ICP 2005-based series.

2.3 The causes of inter-series differences in GDP growth rates

Given that there are differences in alternate data sources of GDP, a natural question is to ask why. In the spirit of Johnson et al. (2013), in this section I estimate a regression model in which the dependent variable is a measure of inter-series variability in GDP growth rates, and several characteristics that are plausibly related to growth rate variability are used as explanatory variables. The effect size and significance of the coefficients is used to identify factors which might contribute to differences between the series.

One can think of the point estimate of GDP growth in a given year for a given country as a random variable for which we have three realizations: one from each of the three GDP datasets (within an ICP round). Hence the standard deviation of these point estimates – for a given country in a given year – is a measure of variability. Formally, define the inter-series variability $\sigma_{it}^{GDP}$ as:

$$\sigma_{it}^{GDP} = \sqrt{\frac{1}{3} \left( \sum_{j=1}^{3} (GDP_{it}^j - \mu_{it}^{GDP})^2 \right)}$$ (1)

where

$$\mu_{it}^{GDP} = \frac{1}{3} \sum_{j=1}^{3} (GDP_{it}^j)$$ (2)

Here $GDP_{it}^j$ is the growth rate for country $i$ at time $t$ as provided by series $j$, and $\mu_{it}^{GDP}$ is the average of the three growth rates for a given country in a given year.

This variability metric is used as a dependent variable in a regression of the form:8

$$\sigma_{it}^{GDP} = \beta_0 + \beta_1 \log(GDP_{it}) + \beta_2 \ast Region_i + \beta_3 \ast Distance_{it} + \beta_4 \ast Type_{ICP}$$ (3)

In the model, $\log(GDP_{it})$ is at PPP and taken from WDI 2015 for the ICP 2011-based comparison, and from

---

8Johnson et al. (2013) conduct a very similar regression exercise in which the dependent variable is the absolute difference in GDP growth rates (or the absolute difference in GDP levels) between different series. I use the standard deviation across series (in a given year) because it enables the simultaneous comparison of multiple datasets in a single model.
WDI 2013 for the ICP 2005-based comparison. Doing so ensures a common methodological reference point when inter-ICP comparisons are made. The first explanatory variable I control for is $\log GDP_{it}$. Based on the literature, it is unclear whether economic size (proxied by $\log GDP_{it}$) should be positively or negatively related to inter-series variability. Rao and Selvanathan (1992) show that standard errors of PPP estimates are inversely related to the total consumption expenditure of a given country. Consequently, in an analysis such as this one, which uses PPP GDP, I would expect smaller countries to have more variability because a less precise PPP conversion factor is applied to their GDP estimates. On the other hand, an argument can also be made that larger GDP leads to higher variability. This is because countries with higher GDPs presumably have more economic sectors. If each sector is measured independently and measurement error is not correlated across sectors, simply having more sectors would lead to greater inter-series variability.

$Region_i$ is a dummy variable for geographic region in which a country is located. Regional categorizations are taken from the World Development Indicators, which groups countries into seven regions: East Asia & Pacific, Europe & Central Asia, Latin America & Caribbean, Middle East & North Africa, North America, South Asia, and sub-Saharan Africa. I control for this regional location because the precision of PPP estimates is governed by geographic characteristics.

The variable $Distance_{it}$ is the same as that used by Johnson et al. (2013) (see their equation 4) and it is the absolute value of the distance (in years) of an observation from the year of the ICP benchmark. I expect $Distance_{it}$ to be negatively related to data variability because PPP extrapolations become less reliable the further out an observation is from the benchmark year. Finally, $Type^{ICP}_i$ is a dummy variable indicating the year of the most recent ICP survey information available for a given country. For the 2011 ICP, this variable can indicate one of four things: (1) country participated in the 2011 ICP; (2) country participated in the 2011 ICP but only data for individual consumption expenditure by households is available; (3) country has a rolling PPP program (i.e. the available consumption and expenditure data is more recent than 2011); and (4) country did not participate in the 2011 ICP. For the 2005 ICP, this variable indicates: (1) country participated in the 2005 ICP; (2) country has PPP information available from 2008; (3) country has a rolling PPP program. Intuitively on expects that benchmark countries that have more recent ICP data should have lower data variability as compared to non-benchmark nations or nations where the ICP information is older.

The results for the ICP 2011 data series are presented in table (11). Across all models tested, I find that log GDP is statistically significant and inversely related to inter-series variability in growth rates, while

---

9I thank James Hosek (RAND) for showing me this line of reasoning.

10The precision of PPPs depends upon geography because consumption and expenditure patterns, which are used to compute PPPs, necessarily depend upon country location (for example, agricultural availability (and consumption patterns of food, say) in Scandinavian countries is bound to be different from that in tropical nations). The ICP expends a tremendous amount of effort to ensure that the product list used to create PPPs is representative of the consumption patterns in all the countries surveyed. Nevertheless, any such “structured product list”, to use terminology fo the ICP, is necessarily imperfect, and this affects the precision of the estimated PPPs.

11For details, see page 3 of World Bank (2015) which notes that in the 2011 ICP a limited amount of price survey information was gathered from certain Pacific Islander nations.
distance to the benchmark year is positively related to inter-series variability in growth rates. The impact of regional differences, first shown in column 2, are not surprising: sub-Saharan Africa has higher variability while North America has lower variability. The interesting case is Europe and Central Asia which has a negative and insignificant coefficient in column 2, but a positive and statistically significant coefficient in column 5, even though one would expect European data to have lower variability, seeing as how it is composed of relatively advanced economies. To examine whether this effect is being driven by the coarseness of the WDI regional classification, which lumps together geographically heterogeneous areas (in that all of Europe and Central Asia is a single category), I re-estimate the model in column 5 using geographical classifications from the United Nations Statistical Division (UNSTAT). These classifications assign countries to more fine-grained geographical classifications (22 versus seven as in the WDI). The detailed geographic categories separate out Europe from Central Asia yet I find that the coefficients for all European geographies (Eastern Europe, Western Europe, Northern Europe, and Southern Europe) are statistically significant and positive. Likewise, the coefficients for Western Asia and for South-central Asia are also positive and statistically significant. Therefore, regional heterogeneity does not seem to play a role in data variability and European and Asian countries do have higher variability, even when economic size and data characteristics are controlled for.

In column four of table (11), I assess whether PPP information has any bearing on inter-series variability. As expected, countries for which rolling PPPs are available have lower variability, presumably because more frequently available PPP information confers greater precision to PPP data (which is then reflected in the PPP GDP data). The final column of table (11) presents the full specification which yields effect sizes and significances similar to the sub-models in columns 1-3.

The equivalent analysis for the ICP 2005-based datasets is provided in table (12). The patterns of variability here are quite different than before. One major difference is that temporal distance from the benchmark year is not consistently positively correlated with variability across all models. One explanation might be that there are about 660 country-year observations for which there is additional PPP price survey data available for 2008. In results not reported in the table, I take the model in column 4 but add an interaction term for Distance_{it} and Type_{ICP}^{t} and find that while the coefficient on Distance_{it} is still significant, negative (and similar in magnitude), the coefficients on all interaction terms are positive, of similar magnitude as Distance_{it}, and statistically significant for the indicator for 2008 data and for the indicator for non-benchmark countries. However, when \( \log(GDP) \) is added to this model, all interaction terms become statistically insignificant. This provides only weak evidence that, conditional on the timing of the benchmark, observations away from it indeed have greater variability. The relationship of variability in GDP growth rates to economic size and to geography is similar to what we saw with ICP 2011-based

\(^{12}\)http://unstats.un.org/unsd/methods/m49/m49regin.htm
comparisons.

In summary, variability in GDP growth rates appears to be driven primarily by economic size and geography, and secondarily by characteristics of the PPP data: economically larger countries have lower growth rate variability; data appears to be consistently more variable for countries in sub-Saharan Africa and in Latin America and the Caribbean. For ICP 2011-based series, observations closer to the ICP benchmark are less variable than those further out and observations for which more recent PPP data are available are less variable than those for which it is not.

3 GDP uncertainty and International Development Assistance

Previous work has shown that GDP uncertainty impacts a variety of theoretical conclusions about economic growth (Johnson et al., 2013; Ciccone and Jarociński, 2010; Hanousek et al., 2008; Ponomareva and Katayama, 2010; Dawson et al., 2001). However, to my knowledge, the impact of GDP uncertainty on policy outcomes has not been examined. In this section I conduct a policy simulation to quantify the amount by which the World Bank’s aid allocations would change as a result of changes in the gross national income data upon which these allocations are based.

The international development association (IDA) is a program of the World Bank which provides funds to the world’s poorest countries to boost economic growth, reduce inequality and improve social welfare.$^{13}$ A country is eligible for IDA funds if its per capita gross national income (GNI) is below a cut-off value. The numeric value of this cut-off is revised every year and for fiscal year 2016 it was set at $1,215 (current USD), making 77 countries eligible to receive aid.$^{14}$ The current funding cycle (also termed the “IDA replenishment”) will disburse $52.1 billion dollars over a three year period ending in June 30, 2017.$^{15}$ These funds will be used by recipient countries to undertake a wide variety of projects in health, education, sanitation, business policy, agriculture, and infrastructure. Given the size and scope of the IDA, the optimal allocation of funds – that is, one which closely matches country needs – has significant implications for global social welfare.

3.1 IDA performance-based allocation

IDA determines grant amounts based on a formula consisting of three factors: a country performance rating, used to assess the effectiveness of a country’s institutions for utilizing aid; country population; and gross national income (GNI) per capita, which is used to take country needs into account. If there are $N$ IDA-

---


eligible countries, then the performance-based allocation (PBA) for country $i$ is given as: $^{16}$

$$PBA_i = \frac{(CPR_i)^5 \times Pop_i \times (GNI per capita_i)^{-0.125}}{\sum_{i=1}^{N}(CPR_i)^5 \times Pop_i \times (GNI per capita_i)^{-0.125}} \times Annual IDA envelope \quad (4)$$

The annual IDA envelope is the total amount of money available through the IDA in a given year. $CPR$ is the country performance rating and as evidenced by its exponent of 5 (versus 1 for population and -0.125 for GNI per capita), CPR is the most important factor in determining how much assistance a country will get. CPR is derived from the World Bank’s Country Policy and Institutional Assessment (CPIA) measure. $^{17}$ The CPIA is a broad indicator of how effective a country’s policies and institutions are for enabling development and growth. It consists of 16 indicators grouped into four equally weighted clusters: economic management (cluster A), structural policies (cluster B), policies for social inclusion (cluster C), and public sector management and institutions (cluster D). Each country’s CPR is calculated as a weighted average of the four CPIA clusters as follows:

$$CPR_i = (0.249 \times CPIA_{i}^{A-C} + 0.68 \times CPIA_{i}^{D} + 0.08 \times PORT_i)$$

$PORT_i$ is the portfolio performance indicator for how well IDA-funded projects are performing in a given country. The initial formula-based PBA’s are subsequently modified based on a variety of country characteristics such as debt distress and risk, identification as a post-conflict nation, credits and deductions for performance, and several other subjective criteria. $^{18}$

### 3.2 Simulation of counterfactual aid allocations

The goal of the simulation is to quantify how aid allocations change due to uncertainty in GDP numbers. However, IDA uses GNI rather than GDP in order to allocate aid, and only a single GNI series is available from the World Bank. Therefore, the first step of the simulation is to generate a set of alternative GNI figures by using the fact that GNI and GDP are related through an accounting identity so that uncertainty in GDP can be propagated to uncertainty in GNI. In the second step, these alternative GNI values are plugged into an estimated version of the PBA formula (4) to compute alternative allocation amounts.

Recall that GNI for country $i$ is defined as:

$^{16}$See annex 2 of World Bank (2010).


$^{18}$An archive of World Bank reports on the current and several previous IDA replenishments are given at: http://ida.worldbank.org/financing/replenishments/ida18-replenishment-1.
\[ GNI_i = GDP_i + NFI_i \] (5)

In the above equation \( NFI_i \) stands for net foreign income, and the equation is typically applied using figures in current USD for both GDP and NFI. I estimate a log-log transformed version of (5), in which current GDP is replaced with PPP GDP from from UQICD 2.1, PWT 9.0, and WDI 2015:\(^{19}\)

\[
\log(GNI_i) = \beta_1 \log(GDP^j_i + NFI_i) 
\] (6)

Here \( j \in \{1, 2, 3\} \) and indicates the source of PPP GDP data (WDI 2015, UQICD 2.1, or PWT 9.0). The model has been log-log-transformed because it leads to an improved fit.\(^{20}\) The predicted values from the model (6) will be \( \log(G\hat{N}I_j) \).\(^{21}\) Now, also note that the PBA formula (4), can be re-written after a log-log transformation as:\(^{22}\)

\[
\log(PBA_i) = \alpha_1 \log(CPR_i) + \alpha_2 \log(Pop_i) + \alpha_3 \log(GNI \text{ per capita}_i) 
\] (7)

Were we to estimate this with values of \( PBA \) as computed by the World Bank, we would get \( \{\hat{\beta}_1 = 5, \hat{\beta}_2 = 1, \hat{\beta}_3 = -0.125\} \). However, the WDI 2015 tables only report the final IDA grant amounts, which are slightly different from the interim \( PBA \) amounts used in their calculations. Therefore, I replace the dependent variable with actual IDA grant amounts (IDAG) from WDI 2015, in the specification above and instead estimate:

\[
\log(IDAG) = \alpha_1 \log(CPR) + \alpha_2 \log(Pop) + \alpha_3 \log(GNI \text{ per capita}_i) 
\] (8)

As long as the relationship between \( PBA \) and \( IDAG \) is approximately linear, the predicted values of the grant amounts from (8) will be close to the actual amounts granted. I estimate the model above using values

\(^{19}\)Using PPP GDP in place of current price GDP is justified as long as the relationship between the two quantities is linear.

\(^{20}\)To elaborate: the goal of estimating (6) is to faithfully generate values of GNI that might have resulted in alternative instantiations of the world. Because these alternative instantiations are not realized, I use the variation in GDP, and the linear relationship between GDP, NFI, and GNI to predict alternative GNI amounts. As such, any transformation that improves model fit is welcome, because we are interested in prediction and not inference. In Figure (1) I produce the goodness-of-fit plots for each of the three models estimated.

\(^{21}\)Note that the dependent variable in (6) does not have a subscript \( j \). This is to say that I use the single available GNI series to estimate the model, and then use predicted GNIs (with subscript \( j \)) from this model for each PPP GDP series.

\(^{22}\)Dropping the scalar denominator and the scalar multiplier (Annual IDA Envelope) is consistent with how the World Bank estimates the log-log transformed version of the PBA allocation formula for its own policy simulations as well. See page 12 (item 28) of World Bank (2009).
of GNI per capita in current US dollars so that \( \hat{\alpha}_3 \) will capture the effect of GNI per capita on the final grant amounts. Then, the estimated coefficients from (8) can be used in conjunction with the predicted values \( \log(\hat{GNI}_j) \) from (6) based on alternative values of \( GDP_j \), to produce predicted values of the grants amounts \( (\hat{IDAG}_j) \) as:

\[
\hat{IDAG}_j = \exp(\hat{\alpha}_1 \log CPR_i + \hat{\alpha}_2 \log Pop_i + \hat{\alpha}_3 \log \hat{GNI}_j)
\]  

(9)

Following the procedure above for each of the PPP GDP time-series (WDI 2015, UQICD 2.1, and PWT 9.0) will yield \( j = 3 \) different estimates of aid allocations. There will also be an estimate of aid allocation based on GNI per capita in current USD, when the model in (8) is first estimated – call this the reference allocation. To quantify the uncertainty in aid that results from changing data, estimates of aid allocation based on PPP-GDP will be compared to the reference allocation and to actual IDA grant amounts in the data, in order to assess the role of model-based uncertainty. The measure of uncertainty in aid allocations I use is the standard deviation of the four estimates: three estimated values of \( IDAG \) plus either the reference allocation or the actual IDA allocation.

### 3.3 Data

This simulation is based on data from fiscal year 2012 because that is the most recent year for which I have complete data on all the necessary indicators, much of which is provided in series published in the WDI 2015. The final IDA grant amounts (i.e. the amount of money granted after adjustments to the calculated PBA) are provided in current USD by the WDI 2015 series called \( DT.DIS.IDAG.CD \); per capita GNI is provided in current USD by the series called \( NY.GNP.PCAP.CD \); and population figures are provided by the series called \( SP.POP.TOTL \). Net foreign income is provided by the series called \( NY.GSR.NFCY.CD \). CPR figures for 2012 are available online at the IDA portal of the World Bank.\(^{24}\) Values of PPP GDP per capita are taken from UQICD 2.1 (\( CRGDP_PC \)), PWT 9.0 (\( RGDE \)), and WDI 2015 (\( NY.GDP.PCAP.PP.KD \)).

### 3.4 Results

The results of this exercise are presented in table (13). The total amount of aid given to a particular country is provided in the first column. I report two measures of aid uncertainty. The first measure of aid uncertainty, reported in column 2, is calculated as the standard deviation of the aid amounts predicted using PPP GDP

\(^{23}\)The predicted values of the final grant amounts can also be written in a way to show their explicit dependence on GDP as: \( \hat{IDAG}_j = \exp(\hat{\alpha}_1 \log CPR_i + \hat{\alpha}_2 \log Pop_i + \hat{\alpha}_3 \hat{GNI}_j \)). Here I have simply substituted-in equation (6) in place of \( \log(\hat{GNI}_j) \).

from the three datasets and the reference allocation. The second measure of uncertainty, reported in column 4, is calculated as the standard deviation of predicted aid amounts and the actual aid granted. The latter measure of aid uncertainty includes model-based uncertainty so it is generally larger than the measure in column 2. For both measures, uncertainty as a fraction of the aid granted is reported in column 3 and column 6, respectively. The countries are ordered in descending order of uncertainty in their aid estimates (reported in the second column).

The good news is that the magnitude of uncertainty as a fraction of total aid is quite small for the majority of the sample. This makes sense because even though GNI is a component of the performance-based allocation, it is given far less weight (an elasticity of 0.125) than population (elasticity of 1) and country performance (elasticity of 5). Only if the measurement error in GNI is substantial would the aid allocations be ultimately affected.

Nevertheless, for a few large recipients of aid, the uncertainty is substantial. The largest absolute uncertainty is for Ethiopia whose allocation could be affected by between 6% to 14% as a result of altering the underlying dataset. Likewise for Ghana, the aid allocations could change by between 5% to 28%, which represent a significant change, given that Ghana received around $111 million in aid. In fact the potential revisions for a number of large recipients of aid are quite substantial, and include countries such as: Burkina Faso, Malawi, and Rwanda. Aid uncertainty is substantial for recipients of smaller aid amounts as well. For example Cameroon received $2.34 million in 2012, but this figure could be revised by at least 20 percent.

In addition, an even more pernicious problem than aid amounts, however, is that the alternative datasets may affect country eligibility for IDA funds. This is a problem not just for false negatives (country should be eligible, but dataset shows it is not), but false positives as well (country should not be eligible, but dataset shows it is). IDA allocation are a zero-sum process so the correct sequence of funding is necessary for optimal social welfare. For example, the WDI cut-off for IDA assistance was $1,205 in 2012, making the following countries eligible for IDA assistance: Bangladesh, Kyrgyz Republic, Cambodia, Nepal, and Tajikistan. However, for these countries the average of the three predicted GNIs greater than the IDA threshold. Assuming that predicted GNIs approximate the uncertainty in GNI reasonably well, it is therefore possible that though these countries received aid, they might not have needed it. On a more promising note, however, in my analysis I don’t find any countries whose predicted GNI is less than the cut-off value but who did not receive aid.

4 Conclusions

In this paper I studied six major sources of GDP data and found there to be large and persistent differences between the rates of economic growth they imply. I found that even simple consistency checks for whether the direction of growth matches across datasets failed to hold in around five to ten percent of the cases.
Consistency across datasets appears to be lowest for countries classified by the World Bank as lower middle income. Geography and the extent of the availability of recent price surveys seemed to be positively correlated with precision. In simulated counterfactual predictions of development assistance by the World Bank, I found that uncertainty in the underlying GNI data can have large effects on the aid given to some countries. However, the magnitude of the misallocation is small relative to the total amount of aid disbursed for most countries.

To protect against these concerns, IDA can take several steps. First, rather than having strict cut-offs for aid, IDA could have a graduated cut-off approach which would recognize that any point estimate of per capita income is necessarily imprecise: if a country falls a bit above the cut-off, it does not necessarily mean that the country is ineligible. Second, IDA could use multiple measures of per capita income to increase robustness. Third, IDA could revise its PBA formula to explicitly account for uncertainty in income estimates. The World Bank continues to study aid disbursement policies and many such revisions have taken place for the IDA over the years. An explicit recognition of uncertainty in GNI estimates would allay fears of sub-optimal allocation, and potentially improve data collection as well. Fourth, alternative measures of wealth that are inherently less prone to uncertainty could be used. For example, Henderson et al. (2012) show that using night-time lights is a promising alternative to national accounts-based income measures.
<table>
<thead>
<tr>
<th>Country</th>
<th>WDI 2015</th>
<th>PWT 9.0</th>
<th>UQICD 2.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benin</td>
<td>0.962</td>
<td>1.85</td>
<td>0.853</td>
</tr>
<tr>
<td>Botswana</td>
<td>2.59</td>
<td>3.79</td>
<td>3.40</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>2.87</td>
<td>2.18</td>
<td>2.77</td>
</tr>
<tr>
<td>Burundi</td>
<td>-1.65</td>
<td>-0.344</td>
<td>0.881</td>
</tr>
<tr>
<td>Cameroon</td>
<td>-0.117</td>
<td>0.340</td>
<td>-0.134</td>
</tr>
<tr>
<td>Cape Verde</td>
<td>6.24</td>
<td>5.28</td>
<td>4.34</td>
</tr>
<tr>
<td>Central African Republic</td>
<td>-0.047</td>
<td>-0.610</td>
<td>0.851</td>
</tr>
<tr>
<td>Chad</td>
<td>2.97</td>
<td>2.11</td>
<td>4.37</td>
</tr>
<tr>
<td>Comoros</td>
<td>-0.499</td>
<td>-1.24</td>
<td>-0.846</td>
</tr>
<tr>
<td>Republic of Congo</td>
<td>0.427</td>
<td>6.46</td>
<td>0.369</td>
</tr>
<tr>
<td>Cote d'Ivoire</td>
<td>-0.659</td>
<td>1.37</td>
<td>-0.481</td>
</tr>
<tr>
<td>Equatorial Guinea</td>
<td>20.7</td>
<td>28.1</td>
<td>15.4</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>3.15</td>
<td>1.85</td>
<td>3.89</td>
</tr>
<tr>
<td>Gabon</td>
<td>-0.305</td>
<td>3.59</td>
<td>-0.223</td>
</tr>
<tr>
<td>The Gambia</td>
<td>0.206</td>
<td>-0.361</td>
<td>-1.67</td>
</tr>
<tr>
<td>Ghana</td>
<td>3.00</td>
<td>3.41</td>
<td>3.04</td>
</tr>
<tr>
<td>Guinea</td>
<td>0.298</td>
<td>-2.13</td>
<td>0.924</td>
</tr>
<tr>
<td>Guinea-Bissau</td>
<td>-0.204</td>
<td>-0.041</td>
<td>7.08</td>
</tr>
<tr>
<td>Kenya</td>
<td>0.562</td>
<td>1.23</td>
<td>0.357</td>
</tr>
<tr>
<td>Lesotho</td>
<td>2.81</td>
<td>3.13</td>
<td>2.94</td>
</tr>
<tr>
<td>Liberia</td>
<td>2.01</td>
<td>4.87</td>
<td>2.28</td>
</tr>
<tr>
<td>Madagascar</td>
<td>-0.744</td>
<td>2.58</td>
<td>-0.714</td>
</tr>
<tr>
<td>Malawi</td>
<td>1.64</td>
<td>0.629</td>
<td>0.518</td>
</tr>
<tr>
<td>Mali</td>
<td>1.45</td>
<td>3.21</td>
<td>1.77</td>
</tr>
<tr>
<td>Mauritania</td>
<td>1.11</td>
<td>2.11</td>
<td>2.20</td>
</tr>
<tr>
<td>Mauritius</td>
<td>3.66</td>
<td>2.72</td>
<td>3.41</td>
</tr>
<tr>
<td>Mozambique</td>
<td>4.66</td>
<td>3.19</td>
<td>5.03</td>
</tr>
<tr>
<td>Namibia</td>
<td>2.06</td>
<td>3.13</td>
<td>2.76</td>
</tr>
<tr>
<td>Niger</td>
<td>-0.120</td>
<td>0.024</td>
<td>0.008</td>
</tr>
<tr>
<td>Nigeria</td>
<td>2.77</td>
<td>10.8</td>
<td>6.15</td>
</tr>
<tr>
<td>Rwanda</td>
<td>3.71</td>
<td>2.80</td>
<td>3.47</td>
</tr>
<tr>
<td>Sao Tome and Principe</td>
<td>2.85</td>
<td>4.43</td>
<td>1.84</td>
</tr>
<tr>
<td>Senegal</td>
<td>0.743</td>
<td>0.207</td>
<td>0.743</td>
</tr>
<tr>
<td>Seychelles</td>
<td>2.50</td>
<td>2.12</td>
<td>2.62</td>
</tr>
<tr>
<td>Sierra Leone</td>
<td>0.828</td>
<td>0.480</td>
<td>3.23</td>
</tr>
<tr>
<td>South Africa</td>
<td>0.834</td>
<td>1.69</td>
<td>0.898</td>
</tr>
<tr>
<td>Sudan</td>
<td>3.62</td>
<td>3.66</td>
<td>2.85</td>
</tr>
<tr>
<td>Swaziland</td>
<td>0.926</td>
<td>1.87</td>
<td>0.468</td>
</tr>
<tr>
<td>Tanzania</td>
<td>1.97</td>
<td>2.97</td>
<td>3.80</td>
</tr>
<tr>
<td>Togo</td>
<td>-0.011</td>
<td>-0.409</td>
<td>0.298</td>
</tr>
<tr>
<td>Uganda</td>
<td>3.57</td>
<td>3.81</td>
<td>4.39</td>
</tr>
<tr>
<td>Zambia</td>
<td>1.97</td>
<td>4.95</td>
<td>0.829</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>-1.59</td>
<td>-2.73</td>
<td>1.15</td>
</tr>
</tbody>
</table>

Table 1: Growth rates of countries in sub-Saharan Africa as computed using ICP 2011-based datasets.
<table>
<thead>
<tr>
<th>Country</th>
<th>WDI 2013</th>
<th>PWT 8.1</th>
<th>PWT 7.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angola</td>
<td>2.59</td>
<td>4.05</td>
<td>4.59</td>
</tr>
<tr>
<td>Benin</td>
<td>0.729</td>
<td>0.959</td>
<td>1.06</td>
</tr>
<tr>
<td>Botswana</td>
<td>4.19</td>
<td>7.23</td>
<td>3.77</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>2.14</td>
<td>1.67</td>
<td>1.27</td>
</tr>
<tr>
<td>Burundi</td>
<td>-0.712</td>
<td>-0.551</td>
<td>-0.645</td>
</tr>
<tr>
<td>Cameroon</td>
<td>-0.035</td>
<td>0.860</td>
<td>-0.566</td>
</tr>
<tr>
<td>Cape Verde</td>
<td>6.02</td>
<td>4.99</td>
<td>4.46</td>
</tr>
<tr>
<td>Central African Republic</td>
<td>0.113</td>
<td>-0.234</td>
<td>-0.821</td>
</tr>
<tr>
<td>Chad</td>
<td>3.05</td>
<td>4.35</td>
<td>3.24</td>
</tr>
<tr>
<td>Comoros</td>
<td>-0.552</td>
<td>-1.13</td>
<td>-1.51</td>
</tr>
<tr>
<td>Republic of Congo</td>
<td>0.950</td>
<td>2.81</td>
<td>0.819</td>
</tr>
<tr>
<td>Cote d’Ivoire</td>
<td>-1.36</td>
<td>-0.076</td>
<td>-0.477</td>
</tr>
<tr>
<td>Equatorial Guinea</td>
<td>11.7</td>
<td>17.1</td>
<td>15.0</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>1.58</td>
<td>1.49</td>
<td>1.69</td>
</tr>
<tr>
<td>Gabon</td>
<td>-0.580</td>
<td>2.61</td>
<td>-0.578</td>
</tr>
<tr>
<td>The Gambia</td>
<td>0.151</td>
<td>0.514</td>
<td>0.230</td>
</tr>
<tr>
<td>Ghana</td>
<td>1.38</td>
<td>1.32</td>
<td>1.69</td>
</tr>
<tr>
<td>Guinea</td>
<td>0.352</td>
<td>-2.20</td>
<td>0.451</td>
</tr>
<tr>
<td>Guinea-Bissau</td>
<td>0.674</td>
<td>0.193</td>
<td>-0.434</td>
</tr>
<tr>
<td>Kenya</td>
<td>0.249</td>
<td>-0.047</td>
<td>0.341</td>
</tr>
<tr>
<td>Lesotho</td>
<td>2.44</td>
<td>2.05</td>
<td>2.34</td>
</tr>
<tr>
<td>Liberia</td>
<td>-0.257</td>
<td>-0.911</td>
<td>-0.643</td>
</tr>
<tr>
<td>Madagascar</td>
<td>-1.22</td>
<td>-0.039</td>
<td>-0.986</td>
</tr>
<tr>
<td>Malawi</td>
<td>0.006</td>
<td>1.00</td>
<td>-0.097</td>
</tr>
<tr>
<td>Mali</td>
<td>1.13</td>
<td>3.31</td>
<td>2.21</td>
</tr>
<tr>
<td>Mauritania</td>
<td>0.365</td>
<td>1.47</td>
<td>1.08</td>
</tr>
<tr>
<td>Mauritius</td>
<td>4.07</td>
<td>3.26</td>
<td>4.10</td>
</tr>
<tr>
<td>Mozambique</td>
<td>2.28</td>
<td>1.90</td>
<td>2.39</td>
</tr>
<tr>
<td>Namibia</td>
<td>0.766</td>
<td>0.456</td>
<td>0.793</td>
</tr>
<tr>
<td>Niger</td>
<td>-1.16</td>
<td>-1.92</td>
<td>-1.46</td>
</tr>
<tr>
<td>Nigeria</td>
<td>0.865</td>
<td>3.66</td>
<td>0.474</td>
</tr>
<tr>
<td>Rwanda</td>
<td>1.73</td>
<td>1.19</td>
<td>2.97</td>
</tr>
<tr>
<td>Sao Tome and Principe</td>
<td>1.79</td>
<td>3.87</td>
<td>3.32</td>
</tr>
<tr>
<td>Senegal</td>
<td>0.428</td>
<td>0.777</td>
<td>0.892</td>
</tr>
<tr>
<td>Sierra Leone</td>
<td>-0.245</td>
<td>0.267</td>
<td>1.20</td>
</tr>
<tr>
<td>South Africa</td>
<td>0.308</td>
<td>0.773</td>
<td>0.775</td>
</tr>
<tr>
<td>Sudan</td>
<td>1.88</td>
<td>2.78</td>
<td>2.28</td>
</tr>
<tr>
<td>Swaziland</td>
<td>2.51</td>
<td>5.34</td>
<td>1.53</td>
</tr>
<tr>
<td>Tanzania</td>
<td>2.09</td>
<td>2.00</td>
<td>2.57</td>
</tr>
<tr>
<td>Togo</td>
<td>-0.853</td>
<td>-0.715</td>
<td>-1.49</td>
</tr>
<tr>
<td>Uganda</td>
<td>2.50</td>
<td>2.52</td>
<td>2.53</td>
</tr>
<tr>
<td>Zambia</td>
<td>-0.257</td>
<td>2.67</td>
<td>1.13</td>
</tr>
</tbody>
</table>

Table 2: Growth rates of countries in sub-Saharan Africa as computed using ICP 2005-based datasets.
<table>
<thead>
<tr>
<th>Income group</th>
<th>WDI 2015 vs. UQICD 2.1 (%)</th>
<th>WDI 2015 vs. PWT 9.0 (%)</th>
<th>PWT 9.0 vs. UQICD 2.1 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Same sign</td>
<td>Opposite sign</td>
<td>Same sign</td>
</tr>
<tr>
<td>All countries</td>
<td>94</td>
<td>6</td>
<td>91</td>
</tr>
<tr>
<td>Low</td>
<td>72</td>
<td>28</td>
<td>81</td>
</tr>
<tr>
<td>Lower middle</td>
<td>97</td>
<td>3</td>
<td>87</td>
</tr>
<tr>
<td>Upper middle</td>
<td>100</td>
<td>0</td>
<td>98</td>
</tr>
<tr>
<td>High</td>
<td>98</td>
<td>2</td>
<td>93</td>
</tr>
</tbody>
</table>

Table 3: This table reports the fraction of the sample in which any two series have different signs on the 21-year panel-average growth rate, for the ICP 2011-based datasets.

<table>
<thead>
<tr>
<th>Income group</th>
<th>WDI 2013 vs. PWT 8.1 (%)</th>
<th>WDI 2013 vs. PWT 7.1 (%)</th>
<th>PWT 8.1 vs. PWT 7.1 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Same sign</td>
<td>Opposite sign</td>
<td>Same sign</td>
</tr>
<tr>
<td>All countries</td>
<td>93</td>
<td>7</td>
<td>95</td>
</tr>
<tr>
<td>Low</td>
<td>85</td>
<td>15</td>
<td>84</td>
</tr>
<tr>
<td>Lower middle</td>
<td>91</td>
<td>9</td>
<td>95</td>
</tr>
<tr>
<td>Upper middle</td>
<td>93</td>
<td>7</td>
<td>100</td>
</tr>
<tr>
<td>High</td>
<td>100</td>
<td>0</td>
<td>98</td>
</tr>
</tbody>
</table>

Table 4: This table reports the fraction of the sample in which any two series have different signs on the 29-year panel-average growth rate, for the ICP 2005-based datasets.

<table>
<thead>
<tr>
<th>Income group</th>
<th>Mean Percentage-Point Difference in Growth Rates</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WDI 15 vs. PWT 9.0</td>
<td>WDI 15 vs. UQICD 2.1</td>
</tr>
<tr>
<td>All</td>
<td>-1.03* (3888)</td>
<td>-0.019* (3732)</td>
</tr>
<tr>
<td>Low</td>
<td>-0.087* (591)</td>
<td>-0.747* (574)</td>
</tr>
<tr>
<td>Lower middle</td>
<td>-1.21* (946)</td>
<td>-0.241* (993)</td>
</tr>
<tr>
<td>Upper middle</td>
<td>-1.25* (1077)</td>
<td>0.013* (999)</td>
</tr>
<tr>
<td>High</td>
<td>-1.15* (1274)</td>
<td>-0.057* (1166)</td>
</tr>
</tbody>
</table>

Table 5: The table reports mean percentage-point differences in the growth rates, and the correlations between the different ICP 2011-based GDP datasets. Asterisks (*) mark results that are significant at the 5 percent level in a paired t-test of the difference in mean growth rates. The number of country-year observations used for each test are reported in the parentheses below each estimate.
### Table 6: The table reports mean percentage-point differences in the growth rates, and the correlations between the different ICP 2005-based GDP datasets. Asterisks (*) mark results that are significant at the 5 percent level in a paired t-test of the difference in mean growth rates. The number of country-year observations used for each test are reported in the parentheses below each estimate.

<table>
<thead>
<tr>
<th>Income group</th>
<th>WDI 13 vs. WDI 8.1</th>
<th>WDI 13 vs. PWT 7.1</th>
<th>PWT 8.1 vs. PWT 7.1</th>
<th>WDI 13 vs. WDI 8.1</th>
<th>WDI 13 vs. PWT 7.1</th>
<th>PWT 8.1 vs. PWT 7.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>−0.544∗</td>
<td>−0.052</td>
<td>0.359∗</td>
<td>0.61</td>
<td>0.80</td>
<td>0.64</td>
</tr>
<tr>
<td>Low</td>
<td>0.043</td>
<td>0.025</td>
<td>−0.053</td>
<td>0.69</td>
<td>0.77</td>
<td>0.75</td>
</tr>
<tr>
<td>Lower middle</td>
<td>−0.738∗</td>
<td>−0.029</td>
<td>0.275</td>
<td>0.41</td>
<td>0.68</td>
<td>0.45</td>
</tr>
<tr>
<td>Upper middle</td>
<td>−0.615∗</td>
<td>−0.020</td>
<td>0.476∗</td>
<td>0.66</td>
<td>0.86</td>
<td>0.67</td>
</tr>
<tr>
<td>High</td>
<td>−0.651∗</td>
<td>−0.134</td>
<td>0.561∗</td>
<td>0.66</td>
<td>0.84</td>
<td>0.71</td>
</tr>
</tbody>
</table>

### Table 7: The table reports mean differences in GDP as a fraction of the average GDP in a given income group (the average GDP is calculated using all available WDI 2015 data), and the correlations between GDP levels in ICP 2011-based datasets. Asterisks (*) mark results that are significant at the 5 percent level in a paired t-test of the difference in levels. The number of country-year observations used for each test are reported in the parentheses below each estimate.

<table>
<thead>
<tr>
<th>Income group</th>
<th>WDI 15 vs. WDI 9.0</th>
<th>WDI 15 vs. UQICD 2.1</th>
<th>PWT 9.0 vs. UQICD 2.1</th>
<th>WDI 15 vs. WDI 9.0</th>
<th>WDI 15 vs. UQICD 2.1</th>
<th>PWT 9.0 vs. UQICD 2.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>9.24∗</td>
<td>24.9∗</td>
<td>12.5∗</td>
<td>0.97</td>
<td>0.98</td>
<td>0.97</td>
</tr>
<tr>
<td>Low</td>
<td>−0.775</td>
<td>19.6∗</td>
<td>22.5∗</td>
<td>0.81</td>
<td>0.65</td>
<td>0.55</td>
</tr>
<tr>
<td>Lower middle</td>
<td>11.1∗</td>
<td>28.7∗</td>
<td>15.2∗</td>
<td>0.85</td>
<td>0.93</td>
<td>0.83</td>
</tr>
<tr>
<td>Upper middle</td>
<td>12.7∗</td>
<td>26.2∗</td>
<td>6.82∗</td>
<td>0.83</td>
<td>0.94</td>
<td>0.83</td>
</tr>
<tr>
<td>High</td>
<td>9.60∗</td>
<td>23.1∗</td>
<td>10.1∗</td>
<td>0.92</td>
<td>0.95</td>
<td>0.94</td>
</tr>
</tbody>
</table>
### Mean Difference in GDP (as a percentage of average GDP of income group)

<table>
<thead>
<tr>
<th>Income group</th>
<th>WDI 13 vs. WDI 13</th>
<th>PWT 8.1 vs. PWT 7.1</th>
<th>WDI 13 vs. PWT 8.1</th>
<th>WDI 13 vs. PWT 7.1</th>
<th>PWT 8.1 vs. PWT 7.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>−1.57* (4509)</td>
<td>4.00* (4464)</td>
<td>6.03* (7148)</td>
<td>0.97</td>
<td>0.98</td>
</tr>
<tr>
<td>Low</td>
<td>−17.4* (766)</td>
<td>11.1* (763)</td>
<td>40.2* (1323)</td>
<td>0.85</td>
<td>0.94</td>
</tr>
<tr>
<td>Lower middle</td>
<td>−1.25 (1065)</td>
<td>4.63* (1051)</td>
<td>4.67* (1635)</td>
<td>0.64</td>
<td>0.91</td>
</tr>
<tr>
<td>Upper middle</td>
<td>2.40* (1218)</td>
<td>5.31* (1205)</td>
<td>−1.23* (1911)</td>
<td>0.82</td>
<td>0.84</td>
</tr>
<tr>
<td>High</td>
<td>3.19* (1460)</td>
<td>−1.28* (1445)</td>
<td>−6.76* (2279)</td>
<td>0.94</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Table 8: The table reports mean differences in GDP as a fraction of the average GDP in a given income group (the average GDP is calculated using all available WDI 2013 data), and the correlations between GDP levels in ICP 2005-based datasets. Asterisks (*) mark results that are significant at the 5 percent level in a paired t-test of the difference in levels. The number of country-year observations used for each test are reported in the parentheses below each estimate.

### Income Classification

<table>
<thead>
<tr>
<th>Income Classification</th>
<th>Obs. in income group</th>
<th>Complete obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>0.17</td>
<td>0.53</td>
</tr>
<tr>
<td>Lower Middle</td>
<td>0.23</td>
<td>0.52</td>
</tr>
<tr>
<td>Upper Middle</td>
<td>0.28</td>
<td>0.50</td>
</tr>
<tr>
<td>High</td>
<td>0.32</td>
<td>0.51</td>
</tr>
</tbody>
</table>

Table 9: The tabulations here are based on income groupings from WDI 2013. These fractions apply to the analysis of ICP 2005-based datasets.

### Income Classification

<table>
<thead>
<tr>
<th>Income Classification</th>
<th>Obs. in income group</th>
<th>Complete obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>0.14</td>
<td>0.27</td>
</tr>
<tr>
<td>Lower Middle</td>
<td>0.36</td>
<td>0.26</td>
</tr>
<tr>
<td>Upper Middle</td>
<td>0.24</td>
<td>0.27</td>
</tr>
<tr>
<td>High</td>
<td>0.26</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Table 10: The tabulations here are based on income groupings from WDI 2015. These fractions apply to the analysis of ICP 2011-based datasets.
Dependent variable: $\sigma_{lt}$

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>log($GDP$)</td>
<td>$-0.245^*$</td>
<td>$-0.136^*$</td>
<td>$-0.115^*$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td>$0.019^*$</td>
<td>$0.013^*$</td>
<td>$0.023^*$</td>
<td>$0.014^*$</td>
<td>$0.023^*$</td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Europe and Central Asia</td>
<td>$-0.074$</td>
<td>$0.061$</td>
<td>$1.34^*$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latin America and Caribbean</td>
<td>$0.661^*$</td>
<td>$0.200$</td>
<td>$0.245$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle East and North Africa</td>
<td>$2.70^*$</td>
<td>$1.94^*$</td>
<td>$2.05^*$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North America</td>
<td>$-1.05^*$</td>
<td>$-0.972^*$</td>
<td>$-0.987^*$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Asia</td>
<td>$0.114$</td>
<td>$-0.357$</td>
<td>$-0.342$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>$1.93^*$</td>
<td>$1.61^*$</td>
<td>$1.66^*$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Type^{ICP}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICP 2011 (household consumption)</td>
<td>$-1.69^*$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rolling</td>
<td>$-1.68^*$</td>
<td>$-1.70^*$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-benchmark</td>
<td>$-0.373$</td>
<td>$-0.209$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N$</td>
<td>4041</td>
<td>7089</td>
<td>4041</td>
<td>7089</td>
<td>4041</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.02</td>
<td>0.06</td>
<td>0.06</td>
<td>0.03</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Table 11: Estimates of the regression model from equation (3). The dependent variable is the standard deviation of growth rates across each of the three series for a given country-year observation (see equation (1)). Asterisks (*) mark significance at the 5% level. log($GDP$) is based on WDI 2015. In column 5 one of the indicators for ICP dropped due to multi-collinearity.
<table>
<thead>
<tr>
<th>Dependent variable: $\sigma_{it}$</th>
<th>[1]</th>
<th>[2]</th>
<th>[3]</th>
<th>[4]</th>
<th>[5]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log($GDP$)</td>
<td>−0.348*</td>
<td>−0.254*</td>
<td>−0.224*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td>−0.005</td>
<td>−0.024*</td>
<td>−0.004</td>
<td>−0.022*</td>
<td>0.003</td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Europe and Central Asia</td>
<td>−0.213</td>
<td>−0.045</td>
<td>1.68*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latin America and Caribbean</td>
<td>0.374*</td>
<td>−0.182</td>
<td>0.047</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle East and North Africa</td>
<td>2.21*</td>
<td>1.95*</td>
<td>2.01*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North America</td>
<td>−1.09*</td>
<td>−0.722</td>
<td>−0.182</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Asia</td>
<td>0.350</td>
<td>−0.101</td>
<td>−0.273</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>1.60*</td>
<td>1.18*</td>
<td>1.06*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Type^{ICP}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008 data available</td>
<td></td>
<td></td>
<td></td>
<td>−1.86*</td>
<td>−0.840*</td>
</tr>
<tr>
<td>Rolling</td>
<td></td>
<td></td>
<td>−1.79*</td>
<td>−2.38*</td>
<td></td>
</tr>
<tr>
<td>Non-benchmark</td>
<td></td>
<td></td>
<td>−0.644*</td>
<td>−0.511*</td>
<td></td>
</tr>
<tr>
<td>$N$</td>
<td>4489</td>
<td>7036</td>
<td>4489</td>
<td>7036</td>
<td>4489</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.04</td>
<td>0.06</td>
<td>0.08</td>
<td>0.05</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Table 12: Estimates of the regression model from equation (3). The dependent variable is the standard deviation of growth rates across each of the three series for a given country-year observation (see equation (1)). Asterisks (*) mark significance at the 5% level. log $GDP$ is based on WDI 2013.
<table>
<thead>
<tr>
<th>Country</th>
<th>Aid Received (2012)</th>
<th>Aid Uncertainty</th>
<th>Fractional Uncertainty %</th>
<th>Aid Uncertainty (Model-based)</th>
<th>Fractional Uncertainty % (Model-based)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethiopia</td>
<td>115.8</td>
<td>9.17</td>
<td>6.04</td>
<td>20.7</td>
<td>13.7</td>
</tr>
<tr>
<td>Ghana</td>
<td>111.9</td>
<td>5.36</td>
<td>4.79</td>
<td>31.1</td>
<td>27.8</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>155.6</td>
<td>3.00</td>
<td>1.92</td>
<td>49.6</td>
<td>31.9</td>
</tr>
<tr>
<td>Vietnam</td>
<td>2.17</td>
<td>2.57</td>
<td>119</td>
<td>36.6</td>
<td>1688</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>2.30</td>
<td>2.32</td>
<td>101</td>
<td>27.3</td>
<td>1187</td>
</tr>
<tr>
<td>Zambia</td>
<td>6.58</td>
<td>2.19</td>
<td>33.2</td>
<td>5.00</td>
<td>76.1</td>
</tr>
<tr>
<td>Malawi</td>
<td>108.6</td>
<td>2.16</td>
<td>1.99</td>
<td>40.0</td>
<td>36.8</td>
</tr>
<tr>
<td>Niger</td>
<td>21.1</td>
<td>1.42</td>
<td>6.74</td>
<td>6.22</td>
<td>29.5</td>
</tr>
<tr>
<td>Uzbekistan</td>
<td>0.012</td>
<td>1.40</td>
<td>11650</td>
<td>5.84</td>
<td>48699</td>
</tr>
<tr>
<td>Georgia</td>
<td>1.02</td>
<td>1.33</td>
<td>131</td>
<td>9.34</td>
<td>920</td>
</tr>
<tr>
<td>Burundi</td>
<td>98.4</td>
<td>1.31</td>
<td>1.33</td>
<td>43.8</td>
<td>44.5</td>
</tr>
<tr>
<td>Nepal</td>
<td>114.1</td>
<td>1.11</td>
<td>0.977</td>
<td>46.7</td>
<td>41.0</td>
</tr>
<tr>
<td>Rwanda</td>
<td>57.5</td>
<td>0.693</td>
<td>1.21</td>
<td>13.0</td>
<td>22.6</td>
</tr>
<tr>
<td>Madagascar</td>
<td>0.022</td>
<td>0.672</td>
<td>3059</td>
<td>4.78</td>
<td>21775</td>
</tr>
<tr>
<td>Cambodia</td>
<td>23.0</td>
<td>0.621</td>
<td>2.69</td>
<td>6.71</td>
<td>29.1</td>
</tr>
<tr>
<td>Cote d’Ivoire</td>
<td>102.4</td>
<td>0.501</td>
<td>0.489</td>
<td>46.3</td>
<td>45.3</td>
</tr>
<tr>
<td>Cameroon</td>
<td>2.34</td>
<td>0.621</td>
<td>19.8</td>
<td>4.18</td>
<td>178.4</td>
</tr>
<tr>
<td>Lao PDR</td>
<td>66.4</td>
<td>0.501</td>
<td>0.679</td>
<td>30.1</td>
<td>45.3</td>
</tr>
<tr>
<td>Moldova</td>
<td>1.60</td>
<td>0.465</td>
<td>27.9</td>
<td>2.98</td>
<td>186</td>
</tr>
<tr>
<td>Lesotho</td>
<td>2.67</td>
<td>0.450</td>
<td>15.0</td>
<td>1.25</td>
<td>46.8</td>
</tr>
<tr>
<td>Benin</td>
<td>40.4</td>
<td>0.447</td>
<td>0.993</td>
<td>11.2</td>
<td>27.6</td>
</tr>
<tr>
<td>Mongolia</td>
<td>3.16</td>
<td>0.402</td>
<td>11.8</td>
<td>0.406</td>
<td>12.9</td>
</tr>
<tr>
<td>Nicaragua</td>
<td>30.8</td>
<td>0.401</td>
<td>1.09</td>
<td>10.8</td>
<td>35.1</td>
</tr>
<tr>
<td>Gambia</td>
<td>30.3</td>
<td>0.374</td>
<td>1.10</td>
<td>13.7</td>
<td>45.3</td>
</tr>
<tr>
<td>Haiti</td>
<td>84.6</td>
<td>0.336</td>
<td>0.360</td>
<td>40.1</td>
<td>47.4</td>
</tr>
<tr>
<td>Guinea</td>
<td>41.5</td>
<td>0.335</td>
<td>0.727</td>
<td>17.9</td>
<td>43.1</td>
</tr>
<tr>
<td>Tajikistan</td>
<td>11.5</td>
<td>0.305</td>
<td>2.31</td>
<td>2.12</td>
<td>18.4</td>
</tr>
<tr>
<td>Mali</td>
<td>0.419</td>
<td>0.302</td>
<td>56.4</td>
<td>6.97</td>
<td>1665</td>
</tr>
<tr>
<td>Kyrgyz Republic</td>
<td>39.9</td>
<td>0.266</td>
<td>0.564</td>
<td>16.5</td>
<td>41.3</td>
</tr>
<tr>
<td>Bhutan</td>
<td>0.553</td>
<td>0.236</td>
<td>31.4</td>
<td>1.46</td>
<td>264.3</td>
</tr>
<tr>
<td>Liberia</td>
<td>18.4</td>
<td>0.225</td>
<td>0.853</td>
<td>6.47</td>
<td>35.1</td>
</tr>
<tr>
<td>Togo</td>
<td>23.2</td>
<td>0.173</td>
<td>0.609</td>
<td>8.88</td>
<td>38.3</td>
</tr>
<tr>
<td>Republic of Congo</td>
<td>24.8</td>
<td>0.157</td>
<td>0.569</td>
<td>11.7</td>
<td>47.0</td>
</tr>
<tr>
<td>Central African Republic</td>
<td>38.3</td>
<td>0.141</td>
<td>0.305</td>
<td>18.3</td>
<td>47.8</td>
</tr>
<tr>
<td>Comoros</td>
<td>2.97</td>
<td>0.141</td>
<td>2.48</td>
<td>1.22</td>
<td>41.2</td>
</tr>
</tbody>
</table>

Table 13: Estimates of aid uncertainty based on the prediction model given in (9). The first estimate of aid uncertainty (and the corresponding fractional uncertainty) is calculated using the standard deviation between the predicted aid amounts (based on alternate PPP GDP series) and the predicted aid amount using GNI per capita values in the model in (9). The model-based estimate of aid uncertainty (and the corresponding fractional uncertainty) is calculated using the standard deviation between the predicted aid amounts (based on alternate PPP GDP series) and actual IDA grant amounts from WDI 2015 tables. All numbers are in millions of U.S. dollars. Some IDA grantees from 2012 are not on this list because I dropped countries for which any single piece of information needed to calculate predicted aid amounts was missing.
Figure 1: Plot of actual values of log(GNI) per capita (current USD) versus predicted values of log(GNI) per capita, based on PPP GDP from WDI 2015, UQICD 2.1, and PWT 9.0. The red line denotes the $y = x$ line. If the model fit was perfect, all points would lie on this line. $R^2$ values for all three models were greater than 0.98.
References


In defense of neoclassical $Q$

Titus Galama*  Abdul A. Tariq†

Abstract

The neoclassical theory of firm investment, and the related Tobin’s $Q$-theory, provides the foundation for modern macroeconomics. Yet, studies find its ability to predict firm investment to be weak and that fundamentals, such as cash flow, perform as least as good as $Q$ or better. We investigate three hypotheses that may explain this. First, could Tobin’s average $Q$ suffer from measurement error? Using Compustat data, we employ an alternative measure based on the book value of capital. This measure explains an incremental 14.6 percentage points of the variation in investment, performing significantly better than the extant measure of $Q$, which explains 6.4 percentage points. Second, could the performance of cash flow be consistent with the neoclassical model? From theory we derive a measure of $Q$ based on cash flows that performs as well as extant $Q$. Thus, the performance of cash flows is not evidence against the neoclassical model. Third, could marginal $Q$, not average $Q$, determine investment, as theory suggests? From theory, we derive an empirical relation for marginal $Q$, but find its explanatory power to be weak. These results provide support for the neoclassical model of firm investment and, in particular, for Tobin’s average $Q$.

---

*Center for Economic and Social Research (USC)
†RAND Corporation and Pardee-RAND Graduate School
1 Introduction

The two dominant theories of firm investment are the neoclassical theory of Jorgenson (1963, 1967) and Tobin’s Q-theory (Tobin, 1969). In the neoclassical theory, firms maximize profits by using their capital in production, by optimally investing in capital and by optimally employing labor.\(^1\) Optimality in the neoclassical model entails investing if and only if marginal benefits to investment are higher than marginal costs. Tobin’s Q-theory provides an alternative viewpoint. Q is the ratio of the firm’s market value \(V\) to the replacement cost of its capital \(K\), \(Q = V/K\) (Tobin, 1969; Hayashi, 1982). In practice, since the replacement value of a company’s assets is hard to estimate, \(Q\) is commonly measured using the ratio of the market value of a company’s equity and liabilities to its book value, \(Q = V/V_{\text{book}}\). If the market value of the firm is higher than its replacement value (proxied by the firm’s book value), i.e. \(Q > 1\), the firm will purchase or rent additional capital or invest to acquire more capital because the value of capital is higher inside the firm, whereas a firm will reduce its capital if the opposite is true, i.e. \(Q < 1\). While this investment heuristic is intuitively appealing it is empirically problematic because Tobin’s \(Q\) is an average measure, whereas economic decisions are taken at the margin.

A relation between investment and marginal \(Q\) is obtained in neoclassical models of firm investment that incorporate the cost of adjusting the capital stock (e.g., Mussa, 1977; Abel, 1980; Hayashi, 1982). Hayashi (1982) established that Tobin’s average \(Q\) and neoclassical marginal \(Q\) are equivalent in the sense that Tobin’s average \(Q\) is identical to neoclassical marginal \(Q\) for a price-taking firm if the neoclassical production function and the capital production function are homogenous of degree one. Theory therefore predicts a strong association between investment opportunities, captured by marginal \(Q\), and firm investment.

Taking the equivalence of marginal \(Q\) and average \(Q\) as given, a series of papers has sought to validate \(Q\) theory, but the results suggest a smaller-than-expected role for \(Q\) (e.g., Chirinko, 1993; Gomes, 2001). Studies find that Tobin \(Q\)’s ability to predict firm investment is generally weak, with the dominant contribution to the \(R^2\)’s stemming from firm-fixed effects (e.g., Morck et al., 1990).

Lack of empirical validation would bring into question the theory itself. This is problematic as

\(^1\)Alternatively, firms purchase or rent capital. In a competitive equilibrium where prices equilibrate, the firm is indifferent to purchasing capital, renting capital or investing in capital. As a result, models that assume capital is rented or purchased are equivalent to a formulation where the firm invests in its capital.
the neoclassical model provides the foundation, not only for firm-investment theory, but also for the neoclassical (Solow, 1956; Swan, 1956; Cass, 1965; Koopmans et al., 1965) and modern endogenous (Romer, 1989; Lucas, 1988; Grossman and Helpman, 1991) models of economic growth, which in turn provide the foundation for modern macroeconomics. This may have real world consequences. Such models provide the foundation for real-business cycle theory and dynamic stochastic general equilibrium (DSGE) models, on which the world’s central banks rely to formulate and communicate monetary policy (Smets and Wouters, 2007; Edge et al., 2010; Sbordone et al., 2010).

A possible explanation for the perceived poor performance of the neoclassical model is that average $Q$ is not sufficiently accurately measured (e.g., Gomes 2001; Erickson and Whited 2000; Lewellen and Badrinath 1997). Tobin’s average $Q$ is measured using the ratio of the market value of the firm’s equity and liabilities to the corresponding book value, $Q = \frac{V}{V_{book}}$. In actuality, Tobin’s $Q$ equals $V/K$, where $K$ is the capital stock used in production. We construct a measure of Tobin’s average $Q$ using a more direct measure (more closely related to production) of the capital stock used in production $K$, namely net property, plant and equipment. We attribute the superior performance of this measure to two factors. First, theory suggests that the relevant capital stock is the stock used in production. The book value of property, plant and equipment is arguably a better measure of the capital used in production than the book value of the firm, which includes investments, advances and intangibles. Second, the book value of property, plant and equipment is a tangible measure that can be fairly accurately measured. The book value of the firm, on the other hand, contains besides property, plant and equipment several additional measures, including intangibles, which are not as easily ascertained. The book value of the firm is also more likely to suffer greater measurement error owing to the greater number of variables that are used in its construction.

Further, Tobin’s average $Q$ is a function of the market value of the firm $V$, which is measured using stock-market valuations. However, if the stock market is inefficient, driven, e.g., by human emotion (fads and fashions) (Shleifer, 2000; Morck et al., 1990), then stock markets may not reflect fundamentals (Stein, 2003), potentially explaining Tobin $Q$’s poor performance. Indeed, Morck et al. (1990) find that the stock market does not capture any information over and above fundamentals, such as cash flow and sales variables, and conclude that the stock market is essentially a sideshow in terms of its influence on investment. However, discounted cash flow models may
provide an alternative measure for the market value of the firm $V$. Using the discounted cash flow method, we construct a simple measure of Tobin’s $Q$ using cash flow and cash flow growth to estimate firm values $V$.

The first contribution of this paper is the use of these alternative empirical measures for Tobin’s average $Q$: a) employing a more direct measure of the capital stock $K$ and b) employing cash flow variables. Using the North American file of the Compustat panel dataset of firm financials over the period 1962 to 2014, comprising 188,527 firm-year observations, we find that the measure of Tobin’s average $Q = V/K$, based on the more direct measure of the capital stock, explains an additional 14.6 percentage points of the $R^2$’s (in a model that only includes firm fixed effects besides $Q$) whereas the canonical market to book value measure for Tobin’s average $Q = V/V_{book}$ explains an additional 6.4 percentage points of the $R^2$’s. The new measure performs substantially better than the extant measure of Tobin’s average $Q$, providing support for the notion that measurement error underlies the weak performance of the neoclassical theory. Further, Tobin’s average $Q$ explains a substantial amount of the variation, providing support in favor of the neoclassical model.

Using our measure of Tobin’s average $Q$ based on cash flow (and also profit) variables we find that it performs as well as the canonical measure $Q = V/V_{book}$, explaining an additional 5.7 percentage points of the $R^2$’s. These results are consistent with prior empirical findings that cash flow and profit variables can play at least as important a role as average $Q$ in explaining investment (see, e.g., Morck et al., 1990; Blanchard et al., 1993; Fazzari et al., 1988). The literature tends to interpret this finding as representing a failure of the neoclassical model. However, the cash flow measure was derived from the neoclassical model. In other words, the theory predicts a strong relationship between cash flow and cash flow growth (or alternatively with profit and profit growth). Rather than interpreting its strong performance as a failure of the neoclassical model, it is exactly what one would expect if cash flow (or profit) provides a better proxy for Tobin’s average $Q$ than the usual measure that is based on stock market valuations.

In addition, or alternatively, could it be that the assumptions required to achieve equivalence between average and marginal $Q$ are perhaps too restrictive? Could this explain the weak performance of Tobin’s $Q$? Indeed, Hayashi (1982) stressed that we should “feel uneasy about [the widespread use of average $Q$]”. However, Hayashi (1982) was unable to derive an expression for marginal $Q$ and noted that it “is not directly observable” (p. 218). This suggests that if we are to
evaluate the performance of the neoclassical theory of firm investment we need a way of measuring marginal $Q$.

The second contribution of this paper is the derivation of an empirical expression for marginal $Q$ without the need to impose the homogeneity restrictions of Hayashi (1982) on firm production and capital-production functions. Marginal $Q$ equals the marginal value (or shadow price) of the capital stock, i.e., it is the discounted future revenue derived from an additional unit of capital, $Q = \partial V / \partial K$ (e.g., Abel, 1980). It captures the effect that adding an additional unit of capital has on the discounted future profits of the firm.\(^2\) The expression is intuitive. If future opportunities are good, i.e. when investing in capital increases discounted future profits $\partial V / \partial K > 0$, then firms invest.

The expression for marginal $Q$ follows directly from Pontryagin’s maximization principle (we also provide an exact proof of the result). Under the assumption that capital depreciation is small we obtain a simple relation for marginal $Q$ that is observable using data on the market value of firms, cash-flow (or profit) variables, and the capital stock. The empirical relation we obtain is consistent with the expression for marginal $Q$ obtained by Abel and Blanchard (1986) (also employed by Chirinko and Schaller, 2001; Abel, 2015) and easier to implement empirically.

We test the performance of our empirical expression for marginal $Q$. We obtain a very large R-squared of 0.92. Most of the model’s performance, however, is due to firm fixed effects ($R^2$ of 0.83). Adding the measure of marginal $Q$ provides a 2.1 percentage-point increase in explanatory power, with an effect size that is both statistically and economically significant. The model predicts a 2.6% increase in investment for a 10% increase in marginal $Q$. Thus, somewhat surprisingly, we find that Tobin’s average $Q$ performs better than marginal $Q$. On the other hand, the overall model-fit is substantially higher for marginal $Q$ than it is for average $Q$ ($R^2$’s of 0.92 vs. 0.52, respectively).

Together, these findings provide support for the neoclassical theory of investment, and in particular for Tobin’s average $Q$, which explains a substantial amount (15 percent) of the variation in investment. The relatively poor performance of extant measures of $Q$ appears to be due to measurement error: using a measure of capital that is more closely tied to production, as theory suggests, in constructing Tobin’s average $Q$ significantly increases the explanatory power of $Q$.

---

\(^2\)For the remainder of the firm’s operations, i.e. from time $t$ onward till the end of the firm’s operations.
The structure of the paper is as follows. In section 2 we provide an overview of the literature that empirically evaluates the performance of Tobin’s average $Q$ and the neoclassical model. In section 3 we develop a theory of firm investment and derive an expression for marginal $Q$. In section 4 we obtain empirical expressions for Tobin’s average $Q$ and for marginal $Q$ that can be tested using data on using stock market valuations, measures of firm debt, cash flow variables, and measures of the capital stock. In section 5 we present the data and provide detailed information on the construction of relevant variables used in the analyses. In section 6 we present and interpret the empirical results. Section 7 summarizes and concludes.

2 Literature

Von Furstenberg (1977) is one of the very first papers to assess the relationship between Tobin’s average $Q$ and investment. He finds (see his Table 2) that an aggregate measure of average $Q$ is a statistically significant predictor of aggregate firm investment (t-stat between 5 and 8, depending on the model specification), but that it explains little of the variation in investment compared to other explanatory variables. He regresses investment on the ratio of the real gross stock of equipment to that of plant and equipment (GE/GPE) and on the capacity utilization rate (CU; the ratio of the actual to the potential output of firms). Adding average $Q$ raises the $R^2$ by only 1 percent (from 0.95 to 0.96; compare model 2.6 with 2.7). However, in a regression that excludes CU the performance of average $Q$ is similar to that of CU ($R^2$ of 0.96). Nevertheless, the coefficient estimate implies an elasticity of average $Q$ with respect to investment of around 0.7 (see p.404 of Von Furstenberg, 1977) – an economically meaningful number suggesting that investment is responsive to changes in Tobin’s average $Q$.

Oulton (1981) conducts a similar analysis of the explanatory power of average $Q$ using aggregate investment data on British industrial and commercial companies (see his Table 4). He obtains quite similar results. When average $Q$ is added to a model with capacity utilization as an explanatory variable, the model’s incremental power to explain investment is increased by only 1 percent ($R^2$). Tobin’s average $Q$ and the constant alone can explain 95 percent of the variation. Tobin’s average $Q$ is again found to be a significant determined of investment (t-stat of about 2 to 9, depending on the specification), and the elasticity of $Q$ with respect to investment is 0.70 (see p. 199 of Oulton, 1981), very similar to the value found by Von Furstenberg (1977) for the U.S.
Barro (1990) investigates the explanatory power of U.S. stock returns and average Q separately, on aggregate U.S. investment over a substantially longer time period, from 1891 to 1987 (with data gaps for the two world wars). Barro (1990) also uses aggregate investment data but runs first-difference models. He finds that the growth rate of the stock market price (his measure of the stock market return) outperforms average Q in predicting aggregate investment. The model that includes stock market returns alone obtains an $R^2$ of 0.67, whereas the model with Q alone has an $R^2$ of 0.31. When both variables are included, the $R^2$ is 0.69, suggesting that Q adds about 2 percent to a model with changes in the stock market price. Taken together, these results suggest a substantial role for stock market returns, but less so for average Q, in explaining variation in investment.

Blanchard et al. (1993) use aggregate U.S. firm data for the period 1900 to 1990 (excluding a 10 year window around the second world war). They run a horse race between investment (the ratio of investment to the capital stock to be precise) and average Q, profits, and, what they term, fundamentals (see their table I). For fundamentals they use a constructed measure of the present discounted value of the profit rate. Using difference models they find a limited role for average Q and a more substantial role for profits ($R^2$ of 0.35 for Q versus $R^2$ of 0.62 for profits). When controlling for profit, average Q is no longer significant (their table II). Further, a 1 percent increase in average Q leads to a 0.45 percent increase in investment, whereas a joint increase in fundamentals and average Q leads to a 2 percent increase in investment.

Turning to micro studies of firm investment (providing a more relevant comparison as our paper uses micro data), Fazzari et al. (1988) assess the impact of average Q on investment in a model with firm and year-specific fixed-effects that controls for cash flow. Their approach is most similar to ours. They follow the literature in using the investment to capital stock ratio as the dependent variable, as in equation (17). While the levels equation is their main specification, they also test models with lagged Q as an instrument for Q and estimate the model using first and second differences to address measurement error (see their Table 5). For models without cash flow, their $R^2$'s range between 0.11 and 0.23. Adding a cash flow variable increases the fit, and the $R^2$'s range between 0.19 and 0.46. These $R^2$'s are slightly lower than those in the aggregate investment literature, as is typical when comparing results from aggregate with results from micro data analyses. Their conclusions remain unchanged when using lagged values as instruments for
Q, as well as estimating models in first and second differences to address measurement error and serial correlation.

Blundell et al. (1992) use the same functional form as Fazzari et al. (1988) and this paper (17), but allow average \( Q \) to be endogenous and possibly correlated with firm-specific fixed-effects in a Generalized Method of Moments (GMM) framework. Using an unbalanced panel of U.K. companies over the period 1975 to 1986 they find average \( Q \) to be a statistically significant predictor of investment, even controlling for cash flow.

Morck et al. (1990) investigate the possibility that the stock market does not reflect fundamentals. To investigate this question they check the incremental explanatory power added by abnormal stock returns (capital asset pricing model \([\text{CAPM}]\) \( \alpha \)'s) to fundamentals (cash flow and sales growth). Stock returns alone are found to explain about 15 percent of the movement in investment. Cash flow and sales growth together explain nearly 21 percent of the variation in investment (see model 2.2 in their Table 2). Controlling for fundamentals (i.e. cash flow and sales growth) and stock returns, the stock market is found to explain an additional 3.8 percent of the variation beyond what is captured by fundamentals (compare with model 2.3). All three variables are statistically significant, but t-statistics are expected to be large with a sample size of several thousand firm-year observations (they use Compustat data as in this paper). Nevertheless, these results point to the importance of the stock market, especially considering that the effect sizes are not insignificant: a 10 percent abnormal stock return is associated with a 3.3 (see their model 2.3) percent increase in investment, holding cash flow and sales growth constant. Morck et al. (1990) conclude from their analysis that the stock market plays an important – if small – role in investment decisions.

Very few papers have attempted to test marginal \( Q \), as opposed to average \( Q \), directly. Abel and Blanchard (1986) construct marginal \( Q \) data that involves the computation of a series of derivatives of future profits w.r.t. future values of the capital stock. They find that, similar to studies in the literature that employ Tobin’s average \( Q \) (as cited above), their measure of marginal \( Q \) is a statistically significant predictor of firm investment but does not explain a large, serially correlated fraction of investment.

Unfortunately, in almost none of the above analyses it is transparent how much of the total variation is explained by average \( Q \) versus the constant or fixed effect or versus other variables that are included in the regressions. There are a few exceptions to this. Rhee and Rhee (1995) use
Compustat micro data of 297 companies over the period 1963 to 1988. To compare results with those based on macro data they first use aggregates of their data. They find that average $Q$ alone explains 24 percent ($R^2$) of investment (the ratio of investment to capital; see their table 2, model 1). This is comparable to the result of Barro (1990) and Blanchard et al. (1993) who obtained a non-negligible 31 and 35 percent, respectively (see discussion earlier).

Further, Rhee and Rhee (1995) in a fixed effects panel regression analysis of the micro data (their Table 3) find that average $Q$ explains 8 percent (model 1) of the variation in investment (incremental $R^2$) with each of three measures of fundamentals$^3$ performing worse than average $Q$ (models 2, 3 and 4 in their Table 3) and not adding any explanatory power over and above average $Q$ when added (their models 5, 6 and 7).

In conclusion, the literature appears to be somewhat dissatisfied with the role of Tobin’s average $Q$. Indeed Chirinko (1993) concludes that empirical studies of $Q$-theory provide low $R^2$’s and suffer from residual serial correlation. Few results are available on the incremental contribution of average $Q$. The micro study of Rhee and Rhee (1995) obtains an incremental $R^2$ of 8 percent and aggregate results suggest 24 percent (Rhee and Rhee, 1995), 31 percent (Barro, 1990) and 35 percent (Blanchard et al., 1993). In addition, several studies find that when controlling for fundamentals the role of average $Q$ is substantially diminished. In particular, the stock market and cash flow are found to be important determinants of investment.

So why does the neoclassical and related Tobin’s average $Q$-theory keep being used? Palley (2001) suggests: “A speculative answer is that “strong” $q$ [Tobin’s] and “neo-classical” $q$ have a logical consistency with the Arrow - Debreu general equilibrium model that is the benchmark of modern economics. Neo-classical $q$ is consistent with marginal productivity theory of income distribution, while strong $q$ has managerially controlled firms accumulating capital in accordance with the wishes of rational shareholders as communicated via the stock market.” In other words, much of modern finance, micro-theories of the firm, and macroeconomics relies on the empirical validity of these theories of investment.

$^3$Consisting of the present value of expected future dividends, a similar measure based on earnings, and a measure based on a so-called dividend smoothing approach.
3 A theory of firm investment

Consider a firm seeking to maximize the present value of its future profits

$$\max_{I(t),L(t),T} \int_0^T \pi(t)e^{-\int_0^s r(s)ds}dt + p_K(T)K(T)e^{-\int_0^T r(s)ds},$$

(1)

where \(r(s)\) is the rate of return on capital, and \(\pi(t)\) are profits. The endogenous end of the firm’s operations \(T\) is the time at which the firm stops producing and sells off its assets \(K(T)\) at the going price of capital \(p_K(T)\). Profits consist of the difference between revenues and the costs of inputs

$$\pi(t) = p_F(t)F[K(t), L(t), t] - w(t)L(t) - p_I(t)I(t),$$

(2)

where \(p_F(t)\) is the price of the firm’s output, \(F[.]\) represents the firm’s production process for the final good, \(K(t)\) is the capital stock, \(L(t)\) is labor, and \(I(t)\) is firm investment in the capital stock \(K(t)\). Investment has a price \(p_I(t)\) and labor is paid its wage rate \(w(t)\). The firm’s production process is strictly increasing in both of its arguments, but at a diminishing rate \((\partial F/\partial K > 0, \partial F/\partial L > 0, \partial^2 F/\partial K^2 < 0, \partial^2 F/\partial L^2 < 0)\).

The firm produces its own capital through investment

$$\frac{\partial K}{\partial t} = G[I(t), K(t), t] - \delta(t)K(t),$$

(3)

where \(G[.]\) represents the production process for capital \(K(t)\). It too is strictly increasing in both of its arguments at a diminishing rate \((\partial G/\partial I > 0, \partial G/\partial K > 0, \partial^2 G/\partial I^2 < 0, \partial^2 G/\partial K^2 < 0)\).

Thus, we have the following optimal control problem: the objective function (1) is maximized with respect to the control functions \(I(t), L(t), \) and the parameter \(T\), subject to the constraint (3), and the initial \(K(0) = K_0\) and end condition \(K(T) \geq 0\) (and free). The optimal planning horizon \(T\) is determined by the dynamic envelope theorem (see 7 below).

The Lagrangian (e.g., Seierstad and Sydsaeter, 1986; Caputo, 2005) of this problem is

$$\mathfrak{L}(t) = \pi(t)e^{-\int_0^t r(s)ds} + q_K(t)\frac{\partial K}{\partial t} + \lambda_I(t)I(t)e^{-\int_0^t r(s)ds},$$

(4)

where \(q_K(t)\) is the co-state variable associated with the dynamic equation (3) for capital \(K(t)\) and
$\lambda_I(t)$ is the Lagrange multiplier for the condition $I(t) \geq 0, \forall t$ ($\lambda_I(t) = 0$ for $I(t) \geq 0$ and $\lambda_I(t) > 0$ for $I(t) < 0$). The latter condition enforces that investment is non-negative $I(t) \geq 0, \forall t$, since in practice, it is hard for firms to disinvest or to sell off parts of their capital.

The co-state function $q_K(t)$ finds a natural economic interpretation in the following result from Pontryagin’s maximum principle (for a detailed proof see Appendix A, equation 34)

$$q_K(t) = \frac{\partial}{\partial K(t)} \left\{ \int_t^{T^*} \pi(*) e^{-\int_s^{T^*} r(x)dx} ds + p_K(T^*) K(T^*) e^{-\int_t^{T^*} r(x)dx} \right\},$$  \hspace{1cm} (5)

(e.g., Caputo, 2005) and where $T^*$ denotes the optimal timing of the end of operations and $\pi(*)$ denotes the maximized profit function (i.e., along the optimal paths for the controls, state functions, and for the optimal timing of the end of operations). Thus, $q_K(t)$ represents the marginal value of future profits over the remainder of the firm’s operations (i.e. from $t$ onwards till $T^*$) derived from additional capital $K(t)$. We refer to the co-state function $q_K(t)$ as the “marginal value of capital” (often also referred to as the shadow price of capital).

Further, we have the following transversality condition (see proof in Appendix A, equation 33)

$$q_K(T) = p_K(T) e^{-\int_0^T r(s)ds},$$  \hspace{1cm} (6)

i.e. the marginal value of capital $q_K(T)$ at the end of the firm’s operations is simply the discounted price of capital $p_K(T) e^{-\int_0^T r(s)ds}$. The condition for the optimal timing of the end of operation $T$ follows from the dynamic envelope theorem (see proof in Appendix A, equation 31)

$$\frac{\partial}{\partial T^*} \left\{ \int_t^{T^*} \pi(*) e^{-\int_s^{T^*} r(x)dx} ds + p_K(T^*) K(T^*) e^{-\int_t^{T^*} r(x)dx} \right\} = \frac{\partial}{\partial T^*} \int_0^{T^*} \Im(t) dt + \frac{\partial}{\partial T^*} \left[ p_K(T^*) K(T^*) e^{-\int_0^{T^*} r(x)dx} \right]$$

$$= \Im(T^*) + \frac{\partial}{\partial T^*} \left[ p_K(T^*) K(T^*) e^{-\int_0^{T^*} r(x)dx} \right] = 0,$$  \hspace{1cm} (7)

where $\Im(T^*) + (\partial / \partial T^*) \left[ p_K(T^*) K(T^*) e^{-\int_0^{T^*} r(x)dx} \right]$ is the marginal value of postponing the end of operations $T$ (see 7), capturing the effect of adding an increment of $T$ on the lifetime profits of the firm, and the time $T$ at which continuing operations no longer has value defines the optimal duration $T^*$. 
3.1 Some reflection

There are two things of note about our formulation. First, we have followed Hayashi (1982) in the set up of our model, except that, for simplicity of exposition, we do not incorporate taxes. Hayashi (1982) employs a more general production function for capital than is typically used (and so do we), in which not only investment but also capital generates future capital (see 3). This formulation is mathematically equivalent to so-called adjustment cost models (Hayashi, 1982).

Second, unlike Hayashi (1982) we allow for the possibility that the firm has a finite horizon \( T \). The firm chooses its horizon optimally and the special case where the horizon \( T \) is infinite (\( T \to \infty \)) is one potential solution to the problem formulated here. The benefits of allowing for a finite and endogenous duration of operations \( T \) are: a) a finite duration of operations captures an important reality that firms have finite length’s of life; and b) an endogenous duration of operations avoids the issue of firms ending operations at an arbitrary (exogenous) time devoid of economic meaning. The optimal duration of operations has a natural economic interpretation as the moment in time when there is no longer value in continuing operations (see 7 and the discussion in section 3.2).

3.2 First-order conditions

The first-order conditions for investment \( I(t) \) and labor \( L(t) \) follow from differentiating the Lagrangian (4) with respect to \( I(t) \) and \( L(t) \), respectively

\[
q_K(t) = \frac{p_I(t) - \lambda_I(t)}{\partial G[I]} e^{-\int_0^t r(s)ds}, \quad (8)
\]

\[
p_F(t) = \frac{w(t)}{\partial F[I]} e^{-\int_0^t r(s)ds}, \quad (9)
\]

where \( q_K(t) \) is the marginal benefit of capital investment (the marginal value of adding additional capital, see 5) and \( p_I(t) (\partial G[I]/\partial I)^{-1} e^{-\int_0^t r(s)ds} \) its marginal cost (see 8). For an interior solution (i.e. non-negative \( I(t) \geq 0 \) and vanishing \( \lambda_I(t) = 0 \)), the marginal cost of investment increases in the price of investment \( p_I(t) \) and in the level of investment due to diminishing returns to scale in investment \( I(t) \) of the capital-production process, \( \partial^2 G[I]/\partial I^2 < 0 \). Intuitively, at higher levels of investment \( I(t) \) the improvement in capital \( \partial K/\partial t \) (see 3) as a result of additional investment

\[\text{Footnote: Taxes can be incorporated following Hayashi (1982). Incorporating taxes adds a greater level of realism at the cost of greater complexity.}\]
δI(t) is smaller because the production curve G[.] is flatter at higher levels of investment due to concavity. As a result, it is more costly to invest at higher levels of investment. A corner solution occurs when I(t) = 0. As (8) shows, ∂G[.]/∂I becomes infinitely large as I(t) approaches zero, under the usual Inada conditions, and qK(t) approaches zero. Intuitively, when capital no longer has value the firm stops investing. A negative marginal value of capital qK(t), i.e. the optimal decision of the firm would be to hold less capital, would be associated with negative investment I(t). But in practice, it is hard for firms to disinvest or to sell off parts of their capital, captured in our assumption that investment is non-negative. The corner solution I(t) = 0 is guaranteed for λt = pI(t).

Likewise, pF(t) is the marginal benefit of labor and w(t) (∂F[.]/∂L)−1 its marginal cost (see 9). The marginal cost increases in the cost of labor w(t) and with labor L(t) due to diminishing returns to scale in labor L(t) of the final goods production process, ∂2F[.]/∂L2 < 0.

The marginal value of capital qK(t) evolves according to the dynamic co-state equation

\[-\frac{∂qK}{∂t} = pF(t)\frac{∂F[.]}{∂K}e^{−\int_0^t r(s)ds} + qK(t)\left[\frac{∂G[.]}{∂K} - \delta(t)\right],\]

(10)

where the rate at which the marginal benefit of capital qK(t) depreciates over a short interval of time, −∂qK/∂t, equals the sum of the direct benefits of capital and the contribution of capital to enhancing the value of the capital stock (see Dorfman, 1969, for an economic interpretation of dynamic co-state equations). Capital contributes to profits by raising output ∂F[.]/∂K > 0 (a production benefit). The additional output is valued at the discounted price of the final good pF(t)e^{−\int_0^t r(s)ds}. Capital also raises the efficiency of capital production ∂G[.]/∂K > 0 (another production benefit), valued at the marginal value of capital qK(t). The marginal value of capital appreciates with the rate of depreciation of capital δ(t) (a cost), also valued at the marginal value of capital qK(t).

The optimal timing of the end of operations T is determined by the optimality condition (7). We have

\[π(T)e^{−\int_0^T r(s)ds} + qK(T)\frac{∂K}{∂t}\bigg|_{t=T} + \frac{∂}{∂T} \left[pK(T)K(T)e^{−\int_0^T r(x)dx}\right] = 0,\]

(11)

where ∂K/∂t|_{t=T} denotes the derivative of capital with respect to time at t = T. When profits
π(t) remain positive, and the capital stock K(t) and scrap value of capital \( p_K(T)K(T)e^{-\int_0^T r(x)dx} \) continue to grow, an infinite duration of operations is possible. The optimal duration of operations is finite when the combination of these three terms balances, at which time there is no longer value in extending the duration of operations. This may occur when the firm is no longer capable of financing sufficient investment so that the capital stock declines (\( \partial K/\partial t < 0 \)), and/or when the price of output decreases and/or the price of inputs increases such that the firm starts operating at a loss (\( \partial \pi/\partial t < 0 \)), and/or when the scrap value of the firm’s capital is declining (\( \partial[p_K(t)K(t)e^{-\int_0^t r(x)dx}]/\partial t < 0 \)) so that it may be better to sell of the firm’s assets. These are all natural explanations as to why firms end their operations.

### 3.3 Marginal Q

Hayashi (1982) identified the marginal value of capital \( q_K(t) \) with marginal Q and showed that marginal Q is equivalent to Tobin’s average Q if both the production function for the final good and the production function for the capital stock are homogenous of degree one. However, Hayashi (1982) was unable to derive an expression for marginal Q. As a result, the literature has continued to rely on the concept of Tobin’s or average Q. Hayashi (1982) also noted that marginal Q “is not directly observable” (p. 218).

Pontryagin’s maximum principle provides a meaningful expression for co-state variables. In the model under consideration, the marginal value of capital \( q_K(t) \) is given by (5). As this equation shows, marginal Q, or \( q_K(t) \), is indeed the marginal value of the firm’s capital K(t) in terms of its contribution to the value of the firm \( V(t) \)

\[
q_K(t) = \frac{\partial V(t)}{\partial K(t)}, \tag{12}
\]

where

\[
V(t) \equiv \left\{ \int_t^{T^*} \pi(*) e^{-\int_t^s r(x)dx} ds + p_K(T^*)K(T^*)e^{-\int_t^{T^*} r(x)dx} \right\}. \tag{13}
\]

Thus \( V(t) \), the market value of the firm, is the net-present value of future profits plus the discounted scrap value of the firm’s capital. Note that \( V(t) \) is forward looking. Again, \( T^* \) denotes the optimal duration of operations and \( \pi(*) \) denotes the maximized profit function (i.e., along the optimal paths for the controls, state functions, and for the optimal duration of operations). Since \( V(t) \) is
maximized, it is no longer a function of the controls or state functions.\(^5\) Thus \(V(t)\) operates as a value function.

The relation (12) follows directly from Pontryagin’s maximum principle. It is not a new result (see, for example, Abel, 1980, p. 43). It is an important result for our purposes: we will use it to develop an empirically testable measure for marginal \(Q\). For this reason, and also because our theory differs from other models by including an endogenous scrap value of capital and an endogenous duration of operations, we provide proof of (12) in Appendix A.

4 Empirical Strategy

The analytical expression (12) for \(q_K(t)\) (or marginal \(Q\)) can, in principle, be constructed using stock market valuations, measures of firm debt, cash flow variables, and measures of the capital stock. Next we employ the expression to obtain relations that are suitable for empirical testing.

4.1 Empirical relations

We proceed by assuming a flexible functional form for the capital-production process,

\[
G[I(t), K(t), t] \equiv \mu_I(t)I(t)^\alpha K(t)^\beta.
\]

(14)

In particular, we assume diminishing returns to scale, i.e. \(0 < \alpha_I + \beta_I < 1\), rather than the usual homogeneity of degree one, \(\alpha_I + \beta_I = 1\). The functional \(\mu_I(t)\) captures potential changes in the productivity of capital investment (e.g., in the technology of investment) over time.

4.1.1 Tobin’s average \(Q\)

In Appendix B we replicate the proof of Hayashi (1982), that if both the production function for the final good and the production function for the capital stock are homogenous of degree one, then marginal \(Q\) equals Tobin’s average \(Q\), and both equal \(q_K(t) = Q(t) = V(t)/K(t)\). Thus, Tobin’s average \(Q\) is a special case of the more general expression (12) for marginal \(Q\), suggesting (consistent with Hayashi, 1982) that Tobin’s \(Q\) theory is a special case of the neoclassical theory.

\(^5\)This is a standard result of Pontryagin’s maximum principle (see, e.g., Caputo, 2005).
Replacing $\partial V(t)/\partial K(t)$ by $V(t)/K(t)$ in (12), using (8), (14) and $\alpha_I + \beta_I = 1$ (homogenous of degree one), we obtain

$$[I(t)/K(t)]^{1-\alpha_I} = \mu_I(t)\phi(t)Q(t)$$

$$= \mu_I(t)\phi(t)V(t)/K(t),$$

where

$$\phi(t) \equiv \alpha_I \frac{e^{-\int_0^T r(x)dx}}{p_I(t)}.$$ (16)

Taking logarithms on both sides, and discretizing, we can define an empirically testable specification for Tobin’s average $Q$

$$\ln \frac{I_{i,t}}{K_{i,t}} = \delta_{0,i} + \delta_1 \ln \left[ \frac{V_{i,t}}{K_{i,t}} \right],$$ (17)

where $\ln$ denotes the natural logarithm, $i$ denotes the firm, $t$ denotes time, and $\delta_{0,i}$ captures a firm-specific fixed effect to account for unobserved characteristics of the firm that are time invariant. To account for changes in technology and prices one might add year dummies to equation (17). Relation (17) is the canonical empirical relation for Tobin’s average $Q$, widely used in empirical analyses of firm investment (e.g., Blundell et al., 1992; Chirinko, 1993; Blanchard et al., 1993; Bond et al., 2000).

Last, as discussed earlier, the stock market may not reflect fundamentals (Stein, 2003). Using the discounted cash flow method we can construct an alternative measure of firm value $V$ and of marginal $Q$. Start with (13), assume a constant rate of return on capital $r$, assume a constant rate of growth $g$ of profits (cash-flow), $\pi(s) = \pi(t)e^{g(s-t)}$, assume $r > g$ and sufficiently large $T$. We then obtain

$$V(t) = \frac{\pi(t)}{r-g} + \left[ p_K(T)K(T) - \frac{\pi(t)}{r-g} e^{g(T-t)} \right] e^{-r(T-t)}$$

$$\approx \frac{\pi(t)}{r-g},$$ (18)
Thus, Tobin’s average $Q$ can be approximated by

$$Q(t) \approx \frac{\pi(t)}{K(r - g)}. \quad (19)$$

Inserting (19) into (17), we obtain

$$\ln \frac{I_{i,t}}{K_{i,t}} = \beta_{0,i} + \beta_1 \ln [CF_{i,t}] + \beta_2 \ln [r - CFG_{i,t}] + \beta_3 \ln K_{i,t}, \quad (20)$$

where $\ln$ denotes the natural logarithm, $i$ denotes the firm, $t$ denotes time, $\beta_{0,i}$ captures a firm-specific fixed effect to account for unobserved characteristics of the firm that are time invariant, $CF_{i,t}$ and $CFG_{i,t}$ denote, respectively, cash-flow and cash-flow growth (or profit and profit growth) of firm $i$ at time $t$ and $r$ is the rate of return on investment, taken to be constant. To account for changes in technology and prices one might add year dummies to (20).

### 4.1.2 Marginal $Q$

From the first-order condition for investment (8), and using (12), we obtain

$$I(t)^{1-\alpha} = \phi(t)\mu I(t)K(t)^{\beta} \frac{\partial V(t)}{\partial K(t)}. \quad (21)$$

Note that $V(t)$ is not a function of the control investment $I(t)$ since $V(t)$ represents the maximized value of the firm (see 13). Thus $V(t)$ is only a function of the various model parameters and functionals, but is no longer a function of the controls (e.g., Caputo, 2005). However, $V(t)$ is a function of the current state $K(t)$, since the current state $K(t)$ serves as the initial capital stock (the initial condition) for the problem starting at time $t$. Thus $V(t)$ and future values $V(\tau)$ ($\tau > t$) are not endogenous to investment $I(t)$, but the contemporaneous capital stock $K(t)$ and contemporaneous changes in the capital stock $\partial K/\partial t$ are.

To deal with the endogeneity of the capital stock $K(t)$ we replace $\partial V(t)/\partial K(t)$ by $\partial V(t)/\partial t \times [\partial K(t)/\partial t]^{-1}$ and substitute (3) for $\partial K(t)/\partial t$.\footnote{This approach is valid for as long as $\partial K(t)/\partial t \neq 0$. We will return to this point later.} Ignoring depreciation of capital ($\delta(t)$ small), we
obtain

\[ I(t) = \phi(t) \frac{\partial V(t)}{\partial t} \]

\[ = \phi(t) r(t) \left[ V(t) - \frac{\pi(t)}{r(t)} \right], \]

where we have used (13) and applied Leibniz’s integral rule to differentiate the integral in (13).

Expression (22) represents a relationship between investment and marginal \( Q \). Comparing the relation for marginal \( Q \) (22) with the relation for average \( Q \) (15) one finds that marginal \( Q \) and average \( Q \) are the same for \( \alpha_I = 0 \) and zero profits \( \pi(t) = 0 \). Also note that for a firm with a constant rate of profits \( \pi \) and a constant rate of return on capital \( r \), one would obtain \( V = \pi/r \) so that in such a scenario marginal \( Q \) and hence investment \( I(t) \) would be zero.

While the consensus in the literature is that marginal \( Q \) cannot be observed, Abel and Blanchard (1986) have constructed an empirical measure. Their expression 5b is equivalent to bringing the derivative w.r.t. \( K(t) \) in (12) and (13) inside the integral. The expression contains long products of derivatives of all future profits with respect to all future realizations of the capital stock. The benefit of our expression (22) is that it is significantly simpler and easier to implement.

Discretizing, and taking logarithms on both sides we obtain

\[ \ln I_{i,t} = \gamma_{0,i} + \gamma_1 \mathbb{I}(V_{i,t} - \pi_{i,t}/r > 0) \ln \left[ V_{i,t} - \frac{\pi_{i,t}}{r} \right], \]

where \( \ln \) denotes the natural logarithm, \( i \) denotes the firm, \( t \) denotes time, \( \gamma_{0,i} \) captures a firm-specific fixed effect to account for unobserved characteristics of the firm that are time invariant, \( V_{i,t} \) and \( \pi_{i,t} \) denote, respectively, the market value and profits of firm \( i \) at time \( t \) and \( r \) is the rate of return on investment, taken to be constant. \( \mathbb{I}(V_{i,t} - \pi_{i,t}/r > 0) \) is an indicator function. It is positive if marginal \( Q = V_{i,t} - \pi_{i,t}/r \) is non-negative and zero otherwise. This captures the feature of our model that investment is non-negative.\(^8\) To account for changes in technology and prices

\[ q_K(t) = \frac{\partial G(t)}{\partial K(t)} \int_t^{T^*} \frac{\partial \pi(s)}{\partial K(s)} e^{-\int_t^s r(x) + \delta(x) dx} ds + \text{other terms}. \]

Our result contains two additional terms that are absent from Abel and Blanchard (1986). The first term is due to variation in the optimal timing of the end of the firm’s operations \( T \), since \( T \) is endogenous in our model. The second term is the result of variation in the scrap value. A full derivation of the result can be found in Appendix C.

\(^8\) Note that for a model without capital depreciation this ensures that \( \partial K(t)/\partial t > 0 \) so that we can replace
one might add year dummies to (24).

For a more direct comparison between (24) and the canonical relation (17) we divide the non-loglinearized version of (24) by $K_{i,t}$ to obtain

$$\ln \frac{I_{i,t}}{K_{i,t}} = \gamma_{0,i} + \gamma_1 \mathbf{1}(V_{i,t}/K_{i,t} - \pi_{i,t}/rK_{i,t} > 0) \ln \left[ \frac{V_{i,t}}{K_{i,t}} - \frac{\pi_{i,t}}{r K_{i,t}} \right].$$

(25)

5 Data

We use the annual fundamentals file of Compustat North America, a panel dataset that provides the income statements, balance sheets, and statement of cash flows for over 47,000 publicly traded firms in North America from 1950 onwards (Compustat, 2004). To be included in the database, a firm must file distinct annual (form 10K) or quarterly (form 10Q) reports with the securities and exchange commission (SEC). Compustat prioritizes its database additions as follows: (1) Companies with an equity security included on the Standard and Poor’s (S&P) 500, S&P 400, S&P 600, S&P/Toronto Stock Exchange (TSX) Composite, or Russell 3000 indices; (2) Firms that are actively traded on the New York Stock Exchange (NYSE), National Association of Securities Dealers Automated Quotations (NASDAQ), American Stock Exchange (AMEX), TSX, or NYSE/Archipelago Exchange (Arca) exchanges; (3) “Selected high-profile, red-herring Initial Public Offering (IPO) and pro-forma filings”; and (4) Companies with active market activity in terms of prices, turnover or disclosure of business and financial information, that are requested by Compustat clients.

We start with the raw data for the period January 1st, 1950 to December 31st, 2014, and restrict the sample to non-financial firms (since these have non-missing data on capital expenditures, our measure of investment) as is standard in the literature (e.g., see Fazzari et al., 1988; Erickson and Whited, 2000). Following Polk and Sapienza (2009) we drop firm-years with negative or missing values of capital, negative or missing values of investment, and firms with price-to-book ratios that are missing or less than 0.01 or greater than 100, as these observations likely represent errors in the data. Similar to Barro (1990), in order to avoid currency-market fluctuations affecting our results, we estimate our models only on companies that report their financial statements in US dollars.

\[ \frac{\partial V(t)}{\partial t} \bigg/ \partial K(t) \bigg/ \partial t \bigg/ \partial K(t) \bigg/ \partial t \bigg]^{-1}. \] See footnote 7.

9In our case we include all industrial firms (manufacturing, retail, construction and other commercial operations other than financial services), by limiting Compustat Industry Format variable INDFMT to “INDL” (Industrial). We use the INDFMT variable for sample selection, rather than the North American Industry Classification System (NAICS) codes for manufacturing, so as to include as broad a universe of firms as possible.
Compustat reports consolidated financial accounts of any subsidiaries and the combined accounts are reported under the parent firm.

Our final sample covers the period December 31st, 1962 to December 31st, 2014 covering 18,602 distinct firms, yielding 188,525 firm-year observations. In Table 1 we report descriptive statistics for this unbalanced panel of firms. The average market valuation of firms in the sample is around $4.4 billion, and the average investment level is about $129 million.

5.1 Variables used in the analysis

We next describe in some detail the Compustat variables used in our analyses and the measures constructed.

**Market value of equity (MVE):** We use the Center for Research in Security Prices (CRSP) dataset and take the product of the average daily share prices over the month of December and the average number of shares outstanding to obtain the average December market value of equity (MVE) for each firm. For firms missing share prices on a given day, we use the daily bid/ask average for that day as is recommended in the CRSP manual. By averaging share prices over a month we obtain a less noisy measure of firm value.

**Investment (I):** We use as a measure of a firm’s investment the expenditures undertaken to add to the firm’s net property, plant, and equipment (Compustat variable CAPX, not including acquisitions).

**Capital (K):** For capital we use net property, plant and equipment, measured as the cost, minus accumulated depreciation, of tangible property used by the firm for revenue production (Compustat variable PPENT).

**Profits (π):** We define profit as a company’s income after expenses and taxes (Compustat variable IB: income before extraordinary items).

---

88

---

10 The raw file contains 418,787 firm year observations, i.e. about 55 percent of the sample is dropped.


12 This is a commonly used measure of investment in the literature on firm investment. See, e.g., Morck et al. (1990); Polk and Sapienza (2009).

13 This is a measure commonly used in the literature. For example see Polk and Sapienza (2009).
Cash flow (CF): Similar to Kaplan and Zingales (1997) we define cash flow as profits (Compustat variable IB) plus depreciation and amortization (Compustat variable DP). Depreciation and amortization measures non-cash charges for wear and tear on property, depletion charges, and allocation of the current portion of capitalized expenditures.

Price-to-book ratio: This is defined as the ratio of the market value of equity (MVE, see above) to the book value of common equity (Compustat variable CEQ).

Tobin’s average Q: Tobin’s Q is the ratio of the market value of the firm to the replacement value of its capital (Tobin, 1969). Calculating the market value of the firm, consisting of the value of its equity, preferred stock and debt, and calculating the replacement cost of capital is not straightforward (Lindenberg and Ross, 1981). The literature takes two approaches: the first approach uses the perpetual inventory method along with estimates of capital depreciation to derive the replacement value of capital (e.g., Lindenberg and Ross 1981; Salinger and Summers 1983; Fazzari et al. 1988; Blundell et al. 1992, among others). The second approach is more conservative in terms of data requirements and computational effort. Chung and Pruitt (1994) show that it is reasonable to use an approximation for Q. We follow this approach using the definition provided by Baker et al. (2003).\footnote{See their Table 2, footnote b.} In particular, we define Tobin’s Q (see 17) as:

\[
Q = \frac{MVE + AT - (CEQ + TXDB)}{AT}
\]  

The numerator is the market value of equity (MVE, as defined before) plus the total assets (Compustat variable AT)\footnote{This is current assets plus net property, plant, and equipment plus other noncurrent assets, including intangible assets, deferred items and investments and advances.} minus the book value of equity (CEQ + TXDB), which is defined as the sum of the book value of common equity (Compustat variable CEQ) and deferred taxes on the balance sheet (Compustat variable TXDB). The denominator is total assets AT, as defined before.

We define two alternative measures of Tobin’s average Q. The first is \(V/K\), i.e. the ratio of the market value of the firm \(V\) (MVE) to the book value of its capital stock \(K\) (Compustat variable PPENT, see above). The second measure of Tobin’s average Q is based on equation (20), and defined as the sum of the log of cash flow (CF) and the log of the difference between the rate of
return on capital \( r \) and cash flow growth (CFG; where growth is averaged over the period \( t-2 \) to \( t \)). Both \( \ln(CF) \) and \( \ln(r-CFG) \) are set to zero, if the argument is less than or equal to zero. This captures the feature of the model that investment is zero for negative marginal \( Q \) (investment is non-negative), as detailed in section 4.

**Marginal \( Q \):** Based on equation (24), we define a measure of marginal \( Q \) as the log of \( V-\pi/r \), where \( V \) is the market value of the firm (MVE), profits \( \pi \) are defined above and \( r \) is taken to be 0.10 (the rate of return on 30-year Treasury bills plus an 8% equity premium; see Bond et al. 2000). Likewise, \( \ln(V-\pi/r) \) is set to zero, whenever \( (V-\pi/r) \) is less than or equal to zero.

For a more direct comparison between (24) and the canonical relation (17) we divide (24) by \( K \) to obtain (25). Based on equation (25), we define a measure of marginal \( Q \) as the log of \( V/K-\pi/rK \). Likewise, \( \ln(V/K-\pi/rK) \) is set to zero, whenever \( (V/K-\pi/rK) \) is less than or equal to zero.

### 6 Regression results

Tables 3 to 12 present the results of our estimation. Because investment is a flow variable, which firms report at the end of a fiscal year, we consider investment over the period \( t \) to \( t+1 \), i.e. \( I_{t+1} \) to be reflective of the investment decision made at time \( t \) in response to expectations \( V_t \) (see 13) of future profits / cash flows and given the capital stock \( K_t \). Each table is organized in a similar way, as follows. For each specification, in the first row we report the \( R^2 \) of the specification with only firm-specific fixed-effects. These have been found to explain the bulk of the variation in investment (Rhee and Rhee, 1995; Morck et al., 1990). In the second row we add our measure of average \( Q \) or marginal \( Q \), report its coefficient and assess the change in \( R^2 \) to get at the incremental explanatory power of a given \( Q \) measure. Comparing incremental changes in \( R^2 \), rather than the overall \( R^2 \), enables us to compare the performance of different \( Q \) measures. In the third row, we include year dummies to control for time variation that is due to economic and technology trends that are shared by all firms. In the fourth row we add the one-period lag of the dependent variable to control for serial correlation (i.e. correlation of last period investment with current period investment). If the coefficient of our \( Q \) measure becomes insignificant as a result, this potentially indicates that its observed significance was simply the result of trending in the dependent variable. If the \( Q \) measure

---

16 This is consistent with the estimation strategy in the literature (in particular see Morck et al., 1990, who provide a justification).
remains significant, then $Q$ is more likely to have an independent explanatory function. We are less concerned about changes in effect size as a result of controlling for lagged dependent variables. Such diminishment is inevitable in time-series data with serial correlation (Achen, 2000). In the last row we include the one-period lag of the $Q$ measure. This is to see if investment responds over a longer time scale than one year.

6.1 Assessments of Tobin’s average $Q$ and marginal $Q$

In Table 2 we present correlations between the various measures of average $Q$ and our measures of marginal $Q$. In the first row, we have the standard measure of average $Q$ as used in the literature, given by the ratio of a firm’s market value to its book value ($V/V_{book}$; see equation 17). In the second row, we have our preferred measure of average $Q$, measured by the ratio of the firm’s market value to the book value of its capital stock ($V/K$; again equation 17). In the third row we have our measure of marginal $Q$ using profits, as derived in (24). The fourth row shows the same specification but with cash flow replacing profits. The fifth row shows specification (25), where marginal $Q$ is scaled by the book value of the firm to make it more comparable to the canonical measure of average $Q = V/V_{book}$. The sixth row shows the same specification, but scaled by the book value of the capital stock $K$. The seventh and eighth rows are similar to the fifth and sixth rows, but profits are replaced with cash flows.

A notable result is that average $Q$, as measured in the standard way as the market-to-book value of the firm $Q = V/V_{book}$, is fairly uncorrelated with average $Q$, as measured by the ratio of the market value of the firm to the book value of the capital stock $Q = V/K$ (correlation of about 10 percent). Next, the measures of marginal $Q$, $Q = V - \pi/r$ and $Q = V - CF/r$, are essentially uncorrelated with the measures of average $Q$, $Q = V/V_{book}$ and $Q = V/K$. This is suggestive evidence that average $Q$ and marginal $Q$ are not interchangeable. In the fifth to eight row, we present measures of marginal $Q$ that are divided by the relevant measure of the stock ($V_{book}$ or $K$) in order to make them more directly comparable to measures of Tobin’s $Q$, $Q = V/V_{book}$ and $Q = V/K$. As expected these measures of marginal $Q$ correlate strongly with their equivalent measures of Tobin’s $Q$.

In Table 3 we present estimation results for our measure of marginal $Q$ (see equation 24). The first row shows that firm-specific fixed-effects explain 83% of the variation in investment. Adding
the measure of marginal $Q$ provides a 2.1 percentage-point increase in explanatory power, with an effect size that is both statistically (t-statistic of 42.1) and economically significant. The model predicts a 2.5% increase in investment for a 10% increase in marginal $Q$. Adding year dummies increases the explanatory power by another 2.0 percentage points. The effect of marginal $Q$ remains significant when lagged values of the dependent and independent variable are added, however the effect size is substantially reduced.

Table 4 presents the same analysis but with cash flow replacing profits. Cash flow differs from profits in two ways: (1) by construction, cash flow is a more immediate reflection of the firm’s financial health because it measures the present financial condition of the firm, whereas profits can be declared on balance sheets for items that will materialize in the future; (2) cash flow is believed to be less manipulable than profits (Wolk et al., 2008). The results in Table 4 are substantially the same, with marginal $Q$ providing only a 0.9 percentage-point improvement in explanatory power after inclusion of firm fixed effects. Including the lagged dependent variable significantly diminishes the size of marginal $Q$, suggesting that trending is perhaps strong in cash flow, as compared to profits (where the magnitude of the reduction is smaller).

Testing the neoclassical theory of firm investment requires us to construct adequate measures of the value of the firm $V$. While the literature tends to use stock-market valuations, there are alternatives, such as the discounted cash flow method. Such alternative methods may provide superior measures of average and marginal $Q$ in light of the evidence that stock markets may not reflect fundamentals. In section 4 we therefore derived a simple alternative expression for Tobin’s average $Q$, based on the discounted cash flow method. This provides a relation between investment, cash flow and cash flow growth (20). In Table 5 we present the results of estimating (20). Cash flow alone provides a 4.4 percentage-point increase in explanatory power above firm fixed-effects, and the difference between the rate of return on capital $r$ and cash flow growth $CFG$ provides an additional 1.3 percentage-point increase in explanatory power, i.e. Tobin’s average $Q$, when based on cash flow variables, explains a total of 5.7 percentage points of the variation in investment. The full model explains 93 percent of the variation and predicts a 4.4% increase in investment for a 10% increase in cash flow. The effect of cash-flow growth on investment is significant, but smaller in magnitude.

The results using profits instead of cash flow are shown in Table 6. The incremental explanatory
power of profits is 3.0 percentage points, slightly lower than for cash flow. Further, the elasticity of investment with respect to profits is about two-thirds the size of the elasticity with respect to cash flow. This is consistent with the notion that cash flow has a more immediate relation to firm investment than profits because of the two aforementioned reasons.

These results are consistent with prior empirical findings that cash flow and profit variables can play at least as important a role as average \(Q\) in explaining investment (see, e.g., Morck et al., 1990; Blanchard et al., 1993; Fazzari et al., 1988). However, the literature interprets this as a failing of the neoclassical model because cash flow and profit variables are seen to “outperform” \(Q\). On the contrary, we find our results to support the neo-classical theory because we find a role for cash flow and profit variables that stems from the theory (as equation 20 shows). Earlier work has often included such variables in an ad hoc way without proper motivation (Chirinko, 1993; Erickson and Whited, 2000).\(^{17}\)

Tables 7 and 8 present investment models for two measures of Tobin’s average \(Q\): 1) a more direct measure using the ratio of the market value to the book value of capital \(V/K\), and 2) the canonical market-to-book value measure \(V/V_{book}\) that is typically employed in the literature. The models are based on equation (17), which is the canonical relation used in the literature to test the relationship between investment and average \(Q\) (e.g., Fazzari et al., 1988). While the effect sizes of the two measures are similar, the incremental explanatory power of \(V/K\) is 14.6 percentage points, compared with 6.4 percentage points for the canonical market-to-book measure \(V/V_{book}\). We attribute this large difference to the way capital is measured. This is discussed in greater detail in section 6.2.

Tables (9) to (12) show measures of marginal \(Q\) that are made more comparable to Tobin’s average \(Q\) by dividing both sides by the relevant measure of the capital stock, \(K\) or \(V_{book}\) (see equation 25). Note that the first term of marginal \(Q\) is average \(Q\). Once more the measures of marginal \(Q\) that use the more direct measure of average \(Q = V/K\) perform best (compare Tables 9 and 10 with Tables 11 and 12) and cash flow performs better than profits (compare Table 9 with Table 10). Still, even in the best performing specification (Table 10) marginal \(Q\) explains

\(^{17}\)In particular, in his review Chirinko (1993) states that “the theoretical basis for inserting variables representing finance constraints has been absent largely and, in light of the well-known theorem of Franco Modigliani and Merton Miller (1958), such a development was discouraged”. Erickson and Whited (2000) re-iterate this sentiment and say [the] “motivation for including cash flow in the regression is not based on a formal model, but rather on a loose analogy with the ‘excess sensitivity’ arguments in the consumption literature.”
incrementally only 1.5 percentage points more of the variation in investment than fixed effects alone.

6.2 Measurement error

We attribute the large difference in the performance of the more direct measure of average $Q = V/K$ versus the canonical measure of average $Q = V/V_{\text{book}}$ to the types of capital that enter each measure. For the $V/K$ measure, the denominator is the book value of the firm’s capital stock, which is the net property, plant and equipment that the firm uses in the production of revenue. On the other hand, the market-to-book measure of average $Q$ uses the book value of the total assets of the firm in its denominator. Broadly speaking, this consists of current assets, plus net property, plant and equipment, plus investment and advances at equity, plus the book value of intangible assets.$^{18}$ In other words, total assets contain more than just those items that a firm can use in its production process. We argue that the inclusion of these extra items weakens the relationship between $Q$ and investment, both conceptually and empirically.

Conceptually, the measure of capital used in $Q$ needs to reflect the nature of investment that is being investigated. The measure of investment that is typically employed in the literature, and also in this paper, is expenditures undertaken to add to the firm’s net property, plant, and equipment (e.g., Morck et al., 1990; Baker et al., 2003; Polk and Sapienza, 2009). Such expenditures are naturally most close related to net property, plant and equipment, our preferred measure of the capital stock $K$ in the calculation of $Q$.

Empirically, all variables are measured with some error and noise from the inclusion of additional items may dilute a pre-existing relationship. We offer some basic evidence in Table 13 to support this empirical point. In the first row of the table we estimate the specification (17), using $Q = V/K$. In each subsequent row, we add an additional component of the firm’s total assets to the denominator. The denominator in the sixth row represents the book value of total assets, i.e. the book value of the firm $V_{\text{book}}$. In other words, as we move down Table 13, we move from our preferred measure $Q = V/K$ to the canonical measure $Q = V/V_{\text{book}}$. The order in which items are added is based on our subjective judgment of their relative importance to the production process. We first add intangible assets, followed by current and “other” assets, followed by the book value of firm’s

$^{18}$In Appendix D we provide a detailed list of every item that comprises the book value of total assets, as reported in the Compustat manual (Compustat, 2004).
investments, and lastly cash.

Two trends emerge from Table 13. First, gradually including additional items in the denominator to eventually match the book value of the firm decreases the model’s explanatory power. Second, the regression coefficient on $V/K$ attenuates in size, and its standard error increases. The fact that the adjusted $R^2$’s decrease suggests that the additional items are less relevant in explaining the variation in firm investment. Moreover, coefficient attenuation is consistent with classical (i.e. uncorrelated) noise in the explanatory variable biasing its coefficient estimate downwards.\(^{19}\)

We are not the first to show that measurement error is a plausible cause of the poor performance of $Q$ in explaining variation in investment. Erickson and Whited (2000) develop a Generalized Method of Moments (GMM) estimator to correct for measurement error in their investment models. In contrast to Erickson and Whited (2000), we motivate our discussion by first considering conceptually how $Q$ ought to be measured, and then validate our hypothesis through a simple econometric exercise. Our conclusion is in line with the sophisticated treatment of measurement error undertaken by Erickson and Whited (2000): they find that average $Q$ plays a notable role in explaining investment. Our approach is more practical, in the sense that researchers can simply adopt the new measure $Q = V/K$ that we are advocating without the need to correct for measurement error using GMM estimators.\(^ {20}\)

### 6.3 Summary

We find that marginal $Q$ and average $Q$ are empirically distinct constructs, as evidenced by weak correlations between measures of each (Table 2). Average $Q$ provides substantially more incremental explanatory power (between 6% and 15%) than marginal $Q$ (between 1 – 2%). This may be because measures of marginal $Q$, measuring marginal changes in levels, are more volatile than measures of

\(^{19}\)If the noise is non-classical, then measurement error may not always attenuate coefficient estimates. In such situations, a slightly stronger assumption is required to ensure that attenuation bias increases with noise. In particular, suppose $X^*$ is the true variable which is measured with non-classical noise $u$ so that $\sigma_{X^*u} \neq 0$. It can be shown that in an OLS regression involving the observed value ($X^* + u$) of $X^*$, the coefficient estimate will be $\text{plim} \hat{\beta} = (\sigma_{X^*u}^2 + \sigma_{X^*u}^2)/[(\sigma_{X^*u}^2 + \sigma_{X^*u}^2) + (\sigma_{X^*u}^2 + \sigma_{u}^2)] \cdot \beta$. From this it is clear that coefficient attenuation will increase with noise if and only if $\sigma_{X^*u} + \sigma_{u}^2 > 0$. This is a mild assumption which is typically made in the literature in non-classical noise situations (see Black et al., 2000, for example).

\(^{20}\)Lewellen and Badrinath (1997) show that results of investment-Q models are sensitive to the precise way in which inventory valuation methods are used to construct replacement values of the firm (the denominator of the canonical empirical measure of $Q$). In our work, the issues pointed out by Lewellen and Badrinath (1997) are less relevant because Chung and Pruitt (1994) have shown that the book value of the firm is an adequate approximation for the replacement value of the firm. Nowadays, the literature has largely abandoned the inventory valuation method.
average $Q$, measuring levels. Further, lack of explanatory power may be a consequence of the investment dynamics, wherein we find that firm-specific fixed-effects explain most (83%) of the variation, leaving little room for any other variable to offer additional explanatory power. The fact that our measure of marginal $Q$ (Table 3) explains an additional 2% of the variation is notable when seen in this light.

We find that cash flow and profit variables provide adequate proxies for average $Q$ (explaining between 2 − 6% of the variation in investment over and above firm fixed-effects), countering the view in the literature that Tobin’s $Q$ and cash flow/profits are at odds. On the contrary, we provide theoretical and empirical justification that they are two sides of the same coin.21

Finally, we find that average $Q$ is alive and well. Its purported lack of explanatory power in earlier studies might just be due to measurement error. We verify this hypothesis using a simple empirical test and find evidence to suggest that indeed measurement error plays a role.

7 Discussion and Conclusion

There is a general dissatisfaction with the performance of the neoclassical theory of firm investment and the related Tobin’s average $Q$ theory. Studies find its ability to predict firm investment to be weak and that fundamentals, such as cash flow, perform as least as well as $Q$ or better. We investigate three hypotheses that may explain this.

First, could Tobin’s average $Q$ suffer from measurement error? We demonstrate, using a more direct measure of capital in the construction of Tobin’s average $Q = V/K$, that average $Q$ can explain an additional 15 percentage points of the variation in investment (with firm fixed effects explaining 32 percentage points). This represents a significant improvement over the performance of the canonical measure $V/V_{\text{book}}$. Thus, contrary to perception, the neoclassical theory performs rather well.

Second, could the performance of cash flow be consistent with the neoclassical model? Another factor that has led researchers to be dissatisfied with the performance of the neoclassical theory is the observation that cash flow and profit measures do as well, or even better, as average $Q$ in explaining firm investment. We argue that it is not necessary to rely on stock market valuations to construct measures of $Q$. Theory predicts strong associations between investment and cash flow.

(or profit) variables. Using a measure of average $Q$ that is based on discounted future cash flows we find it to perform as well as extant $Q$. Thus, finding cash flow measures to perform as well or better than extant $Q$ does not provide evidence against the neoclassical model since neoclassical theory predicts cash flow (or profit) to be an important explanatory variable. Strong performance of cash flow (or profit) variables is what one would expect if they provide a better proxy for Tobin’s average $Q$ than does the extant measure of $Q$.

Third, could marginal $Q$, not average $Q$, determine investment, as theory suggests? For marginal $Q$ we obtain a mixed result. Whereas Tobin’s average $Q$ performs better than marginal $Q$ in terms of it’s incremental explanatory power, the overall model fit for marginal $Q$ is better. This result is consistent with Abel and Blanchard (1986). Intuitively, marginal $Q$ is the better measure, as economic theory suggests decisions are made at the margin. It may simply be that marginal effects are harder to measure ($2^{\text{nd}}$ order) than average effects ($1^{\text{st}}$ order) or that the assumptions that needed to be made in order to arrive at a workable expression are too strong. Or perhaps marginal $Q$ simply does not provide a good model of investment.

In sum, the neoclassical theory, in particular average $Q$ performs quite well in explaining variation in investment. The performance of cash flow (and profit) variables is not to be held against the neoclassical model. In fact, it stems directly from the theory that such variables ought to explain investment. Last, marginal $Q$ does not perform as well as expected. Somewhat surprisingly empirical analyses appear to favor average $Q$ over marginal $Q$. 


<table>
<thead>
<tr>
<th>Variable ($\text{M}$)</th>
<th>Mean (S.D.)</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>N (Firm-year)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V$</td>
<td>4359.5 (48157.8)</td>
<td>169.0</td>
<td>0.070</td>
<td>3656290.0</td>
<td>181,595</td>
</tr>
<tr>
<td>$\Delta V$</td>
<td>299.3 (6674.2)</td>
<td>5.0</td>
<td>-794,397</td>
<td>719373.0</td>
<td>159,854</td>
</tr>
<tr>
<td>$V_{\text{book}}$</td>
<td>4125.7 (51395.9)</td>
<td>129.6</td>
<td>0.088</td>
<td>3771200.0</td>
<td>188,525</td>
</tr>
<tr>
<td>$I$</td>
<td>129.2 (909.1)</td>
<td>4.84</td>
<td>0</td>
<td>50233.4</td>
<td>188,525</td>
</tr>
<tr>
<td>$\Delta I$</td>
<td>8.9 (220.8)</td>
<td>0.114</td>
<td>-17993.0</td>
<td>10.489</td>
<td>167,091</td>
</tr>
<tr>
<td>$K$</td>
<td>746.3 (4879.7)</td>
<td>22.0</td>
<td>0</td>
<td>281677.0</td>
<td>188,527</td>
</tr>
<tr>
<td>$\Delta K$</td>
<td>54.1 (794.0)</td>
<td>0.757</td>
<td>-102739.0</td>
<td>82,400</td>
<td>187,091</td>
</tr>
<tr>
<td>$\pi$</td>
<td>103.9 (885.9)</td>
<td>3.24</td>
<td>-99,289</td>
<td>45220.0</td>
<td>186,637</td>
</tr>
<tr>
<td>$CF$</td>
<td>195.7 (1343.1)</td>
<td>7.04</td>
<td>-95,971</td>
<td>60768.0</td>
<td>181,595</td>
</tr>
<tr>
<td>Tobin’s $Q = V/V_{\text{book}}$</td>
<td>1.76 (2.04)</td>
<td>1.19</td>
<td>0.029</td>
<td>80.5</td>
<td>181,595</td>
</tr>
<tr>
<td>Tobin’s $Q = V/K$</td>
<td>54.2 (998.0)</td>
<td>5.79</td>
<td>0.045</td>
<td>189866.7</td>
<td>177,790</td>
</tr>
<tr>
<td>Marginal $Q = V - \pi/r$</td>
<td>3346.9 (44948.5)</td>
<td>123.4</td>
<td>-201409.3</td>
<td>3503796.0</td>
<td>181,593</td>
</tr>
<tr>
<td>Marginal $Q = V - CF/r$</td>
<td>2273.8 (41624.7)</td>
<td>55.2</td>
<td>-311875.9</td>
<td>3464706.0</td>
<td>179924</td>
</tr>
</tbody>
</table>

Table 1: Descriptive statistics. Reported values are panel averages in millions of USD, except for Tobin’s $Q$ which is dimensionless. The market value $V$ is defined as the market value of equity (MVE) plus total assets (AT) minus the book value of equity, which is the sum of the book value of common equity (CEQ) and the book value of deferred taxes (TXDB). Here Compustat variable names are provided in parentheses. More detail on these variables can be found in section 5. $\Delta V = V_t - V_{t-1}$ is the change in the market value $V$ between waves. The book value of the firm $V_{\text{book}}$ is defined as the total assets of the firm (AT). Investment $I$ is defined as capital expenditures on property, plant and equipment (CAPX). Firms with missing or negative values for $I$ are dropped. $\Delta I = I_t - I_{t-1}$ is the change in investment $I$ between waves. The capital stock $K$ is defined as the book value of property, plant and equipment (PPENT). Firms with missing or negative values of $K$ are dropped. $\Delta K = K_t - K_{t-1}$ is the change in the stock $K$ between waves. Profit $\pi$ is defined as income before extraordinary items (IB). Cash flow (CF) is defined as income before ordinary items (IB) minus depreciation and amortization (DP).
Table 2: Correlations between measures of marginal $Q$ and measures of Tobin’s average $Q$. Variable definitions follow those in section 5 and Table 1. Correlations are reported for all available observations for a given pair of variables.

<table>
<thead>
<tr>
<th></th>
<th>$\frac{V}{V_{book}}$</th>
<th>$\frac{V}{K}$</th>
<th>$V - \frac{\pi}{r}$</th>
<th>$V - \frac{CF}{r}$</th>
<th>$\frac{V}{V_{book}} - \frac{\pi}{rV_{book}}$</th>
<th>$\frac{V}{K} - \frac{\pi}{rK}$</th>
<th>$\frac{V}{V_{book}} - \frac{CF}{rV_{book}}$</th>
<th>$\frac{V}{K} - \frac{CF}{rK}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{V}{V_{book}}$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\frac{V}{K}$</td>
<td>0.0928</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$V - \frac{\pi}{r}$</td>
<td>-0.0075</td>
<td>0.0053</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$V - \frac{CF}{r}$</td>
<td>-0.0009</td>
<td>0.0057</td>
<td>0.9890</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\frac{V}{V_{book}} - \frac{\pi}{rV_{book}}$</td>
<td>0.6863</td>
<td>0.0584</td>
<td>-0.0072</td>
<td>-0.0005</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\frac{V}{K} - \frac{\pi}{rK}$</td>
<td>0.0522</td>
<td>0.5702</td>
<td>0.0025</td>
<td>0.0028</td>
<td>0.0792</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\frac{V}{V_{book}} - \frac{CF}{rV_{book}}$</td>
<td>0.6969</td>
<td>0.0653</td>
<td>-0.0028</td>
<td>0.0048</td>
<td>0.9919</td>
<td>0.0827</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$\frac{V}{K} - \frac{CF}{rK}$</td>
<td>0.0527</td>
<td>0.5713</td>
<td>0.0022</td>
<td>0.0028</td>
<td>0.0789</td>
<td>0.9994</td>
<td>0.0822</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3: Results for the market-value based measure of marginal $Q$ (equation 24). Variable definitions follow those in section 5 and Table 1. We set $r = 0.1$, following Bond et al. (2000), as the approximate rate of return on 30-year Treasury Bills ($\approx 2\%$) plus an 8% risk premium. We set $\ln(V_t - \pi_t/r) = 0$ if $(V_t - \pi_t/r) \leq 0$. See section 4 and equation (24) for detail. Standard errors are clustered at the firm level.
### Table 4: Results for the market-value based measure of marginal $Q$, using cash flow (CF) instead of profits (equation 24).

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Marginal $Q$: $\ln(V_t - CF_t/r)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln I_{t+1}$</td>
<td>$\ln(V_t - CF_t/r)$ [S.E.]</td>
</tr>
<tr>
<td>Firm F.E.’s only</td>
<td>Ommitted</td>
</tr>
<tr>
<td>+ $\ln(V_t - CF_t/r)$</td>
<td>0.132 [0.004]</td>
</tr>
<tr>
<td>+ year dummies</td>
<td>0.045 [0.003]</td>
</tr>
<tr>
<td>+ $\ln I_t$</td>
<td>0.0007 [0.0016]</td>
</tr>
<tr>
<td>+ $\ln(V_{t-1} - CF_{t-1}/r)$</td>
<td>-0.0035 [0.0015]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Adj-R²</th>
<th>F-Stat</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firm F.E.’s only</td>
<td>0.829</td>
<td>N/A</td>
<td>162,658</td>
</tr>
<tr>
<td>+ $\ln(V_t - CF_t/r)$</td>
<td>0.838</td>
<td>1045.1</td>
<td>155,760</td>
</tr>
<tr>
<td>+ year dummies</td>
<td>0.868</td>
<td>122.6</td>
<td>155,760</td>
</tr>
<tr>
<td>+ $\ln I_t$</td>
<td>0.915</td>
<td>567.4</td>
<td>138,425</td>
</tr>
<tr>
<td>+ $\ln(V_{t-1} - CF_{t-1}/r)$</td>
<td>0.915</td>
<td>555.1</td>
<td>137,398</td>
</tr>
</tbody>
</table>

We set $r = 0.1$, following Bond et al. (2000), as the approximate rate of return on 30-year Treasury Bills ($\approx 2\%$) plus an 8% risk premium. We set $\ln(V_t - CF_t/r) = 0$ if $(V_t - CF_t/r) \leq 0$. See section 4 and equation (24) for detail. Standard errors are clustered at the firm level.

### Table 5: Results for the discounted future cash flow-based measure of average $Q$ (equation 20).

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Average $Q$: $\ln(CF_t)$ and $\ln(r - CFG_t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln I_{t+1}$</td>
<td>$\ln(CF_t)$ [S.E.] $\ln(r - CFG_t)$ [S.E.]</td>
</tr>
<tr>
<td>Firm F.E.’s only</td>
<td>Ommitted</td>
</tr>
<tr>
<td>+ $\ln(CF_t)$</td>
<td>0.457 [0.006]</td>
</tr>
<tr>
<td>+ $\ln(r - CFG_t)$</td>
<td>0.443 [0.007]</td>
</tr>
<tr>
<td>+ year dummies</td>
<td>0.321 [0.006]</td>
</tr>
<tr>
<td>+ $\ln I_t$</td>
<td>0.177 [0.003]</td>
</tr>
<tr>
<td>+ $\ln(CF_{t-1})$</td>
<td>0.159 [0.003]</td>
</tr>
<tr>
<td>+ $\ln(r - CFG_{t-1})$</td>
<td>0.156 [0.003]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Adj-R²</th>
<th>F-Stat</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firm F.E.’s only</td>
<td>0.829</td>
<td>N/A</td>
<td>162,658</td>
</tr>
<tr>
<td>+ $\ln(CF_t)$</td>
<td>0.873</td>
<td>5954.3</td>
<td>161,546</td>
</tr>
<tr>
<td>+ $\ln(r - CFG_t)$</td>
<td>0.886</td>
<td>2174.4</td>
<td>129,608</td>
</tr>
<tr>
<td>+ year dummies</td>
<td>0.898</td>
<td>199.2</td>
<td>129,608</td>
</tr>
<tr>
<td>+ $\ln I_t$</td>
<td>0.925</td>
<td>732.7</td>
<td>128,321</td>
</tr>
<tr>
<td>+ $\ln(CF_{t-1})$</td>
<td>0.925</td>
<td>762.5</td>
<td>128,299</td>
</tr>
<tr>
<td>+ $\ln(r - CFG_{t-1})$</td>
<td>0.929</td>
<td>657.8</td>
<td>114,965</td>
</tr>
</tbody>
</table>

Cash flow growth ($CFG_t$) is average growth in $CF_t$ from year $t-2$ to $t$. We set $r = 0.1$, following Bond et al. (2000), as the approximate rate of return on 30-year Treasury Bills ($\approx 2\%$) plus an 8% risk premium. Both $\ln(CF_t)$ and $\ln(r - CFG_t)$ are set to zero if $CF_t \leq 0$ or $(r - CFG_t) \leq 0$, respectively. See section 4 and equation (20) for detail. Standard errors are clustered at the firm level.
### Table 6: Results for the discounted future profit-based measure of average $Q$ (equation 20). Variable definitions follow those in section 5 and Table 1. Profit growth ($\Delta \pi_t$) is average growth in profits from year $t − 2$ to $t$. We set $r = 0.1$, following Bond et al. (2000), as the approximate rate of return on 30-year Treasury Bills ($\approx 2\%$) plus an 8% risk premium. Both $\ln(\pi_t)$ and $\ln(r − \Delta \pi_t)$ are set to zero if $\pi_t ≤ 0$ or $(r − \Delta \pi_t) ≤ 0$, respectively. See section 4 and equation (20) for detail. Standard errors are clustered at the firm level.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Average $Q$: $\ln(\pi_t)$ and $\ln(r − \Delta \pi_t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln I_{t+1}$</td>
<td>$\ln(\pi_t)$ [S.E.]</td>
</tr>
<tr>
<td>Firm F.E.’s only</td>
<td>Omitted</td>
</tr>
<tr>
<td>+ $\ln(\pi_t)$</td>
<td>0.344 [0.005]</td>
</tr>
<tr>
<td>+ $\ln(r − \Delta \pi_t)$</td>
<td>0.325 [0.006]</td>
</tr>
<tr>
<td>+ year dummies</td>
<td>0.222 [0.004]</td>
</tr>
<tr>
<td>+ $\ln I_t$</td>
<td>0.123 [0.0026]</td>
</tr>
<tr>
<td>+ $\ln(\pi_{t-1})$</td>
<td>0.110 [0.0024]</td>
</tr>
<tr>
<td>+ $\ln(r − \Delta \pi_{t-1})$</td>
<td>0.109 [0.0025]</td>
</tr>
</tbody>
</table>

### Table 7: Results for the more direct measure of average $Q$ (equation 17), defined as the ratio of the market value of the firm $V_t$, to the book value of capital $K_t$. Variable definitions follow those in section 5 and Table 1. Standard errors are clustered at the firm level.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Average $Q$: $\ln(V_t/K_t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln \left( \frac{I_{t+1}}{K_t} \right)$</td>
<td>$\ln V_t/K_t$ [S.E.]</td>
</tr>
<tr>
<td>Firm F.E.’s only</td>
<td>Omitted</td>
</tr>
<tr>
<td>+ $\ln V_t/K_t$</td>
<td>0.650 [0.008]</td>
</tr>
<tr>
<td>+ year dummies</td>
<td>0.681 [0.008]</td>
</tr>
<tr>
<td>+ $\ln(I_t/K_{t-1})$</td>
<td>0.601 [0.008]</td>
</tr>
<tr>
<td>+ $\ln(V_{t-1}/K_{t-1})$</td>
<td>0.744 [0.009]</td>
</tr>
</tbody>
</table>
### Table 8: Results for the canonical measure of Tobin’s average $Q = \frac{V}{V_{book}}$ (equation 17). Variable definitions follow those in section 5 and Table 1. Standard errors are clustered at the firm level.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Average $Q$: $\ln \frac{V_t}{V_{t,book}}$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln \frac{I_{t+1} - \pi_t}{K_t}$</td>
<td>$\ln \frac{V_t}{V_{t,book}}$ [S.E.] Adj-R$^2$ F-Stat N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Firm F.E.’s only</td>
<td>Omitted</td>
<td>0.316</td>
<td>N/A</td>
<td>162,483</td>
</tr>
<tr>
<td>+ $\ln \frac{V_t}{V_{t,book}}$</td>
<td>0.673 [0.010]</td>
<td>0.380</td>
<td>4366.8</td>
<td>156,541</td>
</tr>
<tr>
<td>+ year dummies</td>
<td>0.666 [0.011]</td>
<td>0.404</td>
<td>139.3</td>
<td>156,541</td>
</tr>
<tr>
<td>+ $\ln (J_t/K_{t-1})$</td>
<td>0.512 [0.010]</td>
<td>0.446</td>
<td>197.9</td>
<td>139,061</td>
</tr>
<tr>
<td>+ $\ln (V_{t-1}/V_{t-1,book})$</td>
<td>0.561 [0.012]</td>
<td>0.446</td>
<td>192.7</td>
<td>138,125</td>
</tr>
</tbody>
</table>

### Table 9: Results for the measure of marginal $Q$ made more directly comparable to average $Q$ (equation 25) using the more direct measure of $Q = \frac{V_t}{K_t}$ and profits $\pi_t$. Variable definitions follow those in section 5 and Table 1. $\ln (\frac{V_t}{K_t} - \frac{\pi_t}{rK_t})$ is set to zero, if $(\frac{V_t}{K_t} - \frac{\pi_t}{rK_t}) \leq 0$. See section 4 and equation (25) for detail. Standard errors are clustered at the firm level.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Average $Q$: $\ln (\frac{V_t}{K_t} - \frac{\pi_t}{rK_t})$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln \frac{I_{t+1}}{K_t}$</td>
<td>$\ln (\frac{V_t}{K_t} - \frac{\pi_t}{rK_t})$ [S.E.] Adj-R$^2$ F-Stat N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Firm F.E.’s only</td>
<td>Omitted</td>
<td>0.316</td>
<td>N/A</td>
<td>162,483</td>
</tr>
<tr>
<td>+ $\ln \frac{V_t}{K_t} - \frac{\pi_t}{rK_t}$</td>
<td>0.132 [0.005]</td>
<td>0.327</td>
<td>611.7</td>
<td>156,539</td>
</tr>
<tr>
<td>+ year dummies</td>
<td>0.163 [0.005]</td>
<td>0.365</td>
<td>84.21</td>
<td>156,539</td>
</tr>
<tr>
<td>+ $\ln (J_t/K_{t-1})$</td>
<td>0.123 [0.005]</td>
<td>0.426</td>
<td>160.1</td>
<td>139,060</td>
</tr>
<tr>
<td>+ $\ln (V_{t-1}/K_{t-1} - \frac{\pi_{t-1}}{rK_{t-1}})$</td>
<td>0.107 [0.005]</td>
<td>0.426</td>
<td>156.0</td>
<td>138,122</td>
</tr>
</tbody>
</table>
### Table 10: Results for the measure of marginal Q made more directly comparable to average Q (equation 25) using the more direct measure of $Q = \frac{V_t}{K_t}$ and cash flow $CF_t$. Variable definitions follow those in section 5 and Table 1. $\ln \left( \frac{V_t}{K_t} - \frac{CF_t}{rK_t} \right)$ is set to zero if $(\frac{V_t}{K_t} - \frac{CF_t}{rK_t}) \leq 0$. See section 4 and equation (25). Standard errors are clustered at the firm level.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Average Q: $\ln(\frac{V_t}{K_t} - \frac{CF_t}{rK_t})$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\ln(\frac{V_t}{K_t} - \frac{CF_t}{rK_t})$ [S.E.]</td>
</tr>
<tr>
<td>Firm F.E.’s only</td>
<td>Omitted</td>
</tr>
<tr>
<td>+ ln $V_t/K_t - CF_t/rK_t$</td>
<td>0.128 [0.004]</td>
</tr>
<tr>
<td>+ year dummies</td>
<td>0.140 [0.004]</td>
</tr>
<tr>
<td>+ ln($I_t/K_{t-1}$)</td>
<td>0.098 [0.0037]</td>
</tr>
<tr>
<td>+ ln($V_{t-1}/K_{t-1} - CF_{t-1}/rK_{t-1}$)</td>
<td>0.0873 [0.0037]</td>
</tr>
</tbody>
</table>

### Table 11: Results for the measure of marginal Q made more directly comparable to average Q (equation 25) using the canonical measure of $Q = \frac{V_t}{V_{t,book}}$ and profits $\pi_t$. Variable definitions follow those in section 5 and Table 1. $\ln \left( \frac{V_t}{V_{t,book}} - \frac{\pi_t}{rV_{t,book}} \right)$ is set to zero if $(\frac{V_t}{V_{t,book}} - \frac{\pi_t}{rV_{t,book}}) \leq 0$. See section 4 and equation (25). Standard errors are clustered at the firm level.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Average Q: $\ln(\frac{V_t}{V_{t,book}} - \frac{\pi_t}{rV_{t,book}})$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\ln(\frac{V_t}{V_{t,book}} - \frac{\pi_t}{rV_{t,book}})$ [S.E.]</td>
</tr>
<tr>
<td>Firm F.E.’s only</td>
<td>Omitted</td>
</tr>
<tr>
<td>+ ln $V_t/V_{t,book} - \frac{\pi_t}{rV_{t,book}}$</td>
<td>-0.031 [0.005]</td>
</tr>
<tr>
<td>+ year dummies</td>
<td>-0.024 [0.005]</td>
</tr>
<tr>
<td>+ ln($I_t/K_{t-1}$)</td>
<td>-0.039 [0.004]</td>
</tr>
<tr>
<td>+ ln($V_{t-1}/V_{t-1,book} - \frac{\pi_{t-1}}{rV_{t-1,book}}$)</td>
<td>-0.047 [0.004]</td>
</tr>
</tbody>
</table>
### Table 12: Results for the measure of marginal $Q$ made more directly comparable to average $Q$ (equation 25) using the canonical measure of $Q = \frac{V_t}{V_{t,book}}$ and cash flow $CF_t$. Variable definitions follow those in section 5 and Table 1.

$\ln\left(\frac{V_t}{V_{t,book}} - \frac{CF_t}{rV_{t,book}}\right)$ is set to zero if $\left(\frac{V_t}{V_{t,book}} - \frac{CF_t}{rV_{t,book}}\right) \leq 0$. See section 4 and equation (25). Standard errors are clustered at the firm level.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Average $Q$: $\ln(\frac{V_t}{V_{t,book}} - \frac{CF_t}{rV_{t,book}})$</th>
<th>[S.E.]</th>
<th>Adj-R²</th>
<th>F-Stat</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln\left(\frac{I_{t+1}}{K_t}\right)$</td>
<td>$\ln(\frac{V_t}{V_{t,book}} - \frac{CF_t}{rV_{t,book}})$</td>
<td>0.316</td>
<td>N/A</td>
<td>162,483</td>
<td></td>
</tr>
<tr>
<td>Firm F.E.’s only</td>
<td>Omitted</td>
<td>0.315</td>
<td>101.0</td>
<td>155,600</td>
<td></td>
</tr>
<tr>
<td>+ ln $\frac{V_t}{V_{t,book}} - \frac{CF_t}{rV_{t,book}}$</td>
<td>0.033 [0.003]</td>
<td>0.348</td>
<td>70.5</td>
<td>155,600</td>
<td></td>
</tr>
<tr>
<td>+ year dummies</td>
<td>0.029 [0.003]</td>
<td>0.416</td>
<td>141.2</td>
<td>138,297</td>
<td></td>
</tr>
<tr>
<td>+ ln($\frac{I_t}{K_{t-1}}$)</td>
<td>0.0084 [0.0028]</td>
<td>0.415</td>
<td>138.8</td>
<td>137,275</td>
<td></td>
</tr>
<tr>
<td>+ ln($\frac{V_{t-1}}{V_{t-1,book}} - \frac{\pi_{t-1}}{rV_{t-1,book}}$)</td>
<td>0.00469 [0.0028]</td>
<td>0.415</td>
<td>138.8</td>
<td>137,275</td>
<td></td>
</tr>
</tbody>
</table>
Table 13: The basic model we estimate is given by equation (17), to which we add firm and year fixed-effects, as well as the lagged values of the dependent variable ($\frac{\ln V}{K}$), and the lagged values of the independent variable. Each row of the above table adds an additional variable which comprises the book value of total assets of the firm. The final row is equivalent to the book value of total assets. The variables are defined in further detail in Appendix D, but they are: INTAN (intangible assets), ACT (current assets), AO (other assets), IVAEQ (investments and advances - equity), and IVAO (investments and advance - other). There were comparatively fewer values that were negative versus values that were zero. We set both missing and negative values to zero.

<table>
<thead>
<tr>
<th>Definition of $K$ in $\ln V/K$</th>
<th>$\ln \frac{V}{K}$ [S.E.]</th>
<th>$R^2$</th>
<th>F-Stat</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K = \text{PPENT}$</td>
<td>0.744 [0.009]</td>
<td>0.5291</td>
<td>329.2</td>
<td>138125</td>
</tr>
<tr>
<td>$K = \text{PPENT} + \text{INTAN}$</td>
<td>0.565 [0.010]</td>
<td>0.4930</td>
<td>248.1</td>
<td>138125</td>
</tr>
<tr>
<td>$K = \text{PPENT} + \text{INTAN}$ + $\text{ACT}$</td>
<td>0.516 [0.012]</td>
<td>0.4483</td>
<td>190.9</td>
<td>138125</td>
</tr>
<tr>
<td>$K = \text{PPENT} + \text{INTAN}$ + $\text{ACT} + \text{AO}$</td>
<td>0.537 [0.012]</td>
<td>0.4475</td>
<td>188.8</td>
<td>138125</td>
</tr>
<tr>
<td>$K = \text{PPENT} + \text{INTAN}$ + $\text{ACT} + \text{AO} + \text{IVAEQ}$</td>
<td>0.531 [0.012]</td>
<td>0.4466</td>
<td>186.6</td>
<td>138125</td>
</tr>
<tr>
<td>$K = \text{PPENT} + \text{INTAN}$ + $\text{ACT} + \text{AO} + \text{IVAEQ} + \text{IVAO}$</td>
<td>0.525 [0.012]</td>
<td>0.4446</td>
<td>183.2</td>
<td>138125</td>
</tr>
</tbody>
</table>
References


A Equivalence of $q_K(t)$ and marginal $Q$: general case

Start with equation (13) for $t = 0$. Note that $V(0)$ is a function of model parameters $\beta$, the initial $K_0$ and the end state $K_T$ of the capital stock, and the initial $t = 0$ and end time $T$ (e.g., Caputo, 2005). Now take the derivative w.r.t. the initial capital stock $K_0$. Since $K(T)$ and $T$ are free we obtain

$$\frac{\partial V(0)}{\partial K_0} = \left. \frac{\partial V(0)}{\partial T} \right|_{T,K_T} + \left. \frac{\partial V(0)}{\partial T} \right|_{K_0,K_T} \frac{\partial T}{\partial K_0} + \left. \frac{\partial V(0)}{\partial T} \right|_{K_0,T} \frac{\partial K_T}{\partial K_0} \right|_{T,K_T},$$

(27)

where $\partial V(0)/\partial K_0|_{T,K_T}$ is the derivative of $V(0)$ w.r.t. $K_0$ for constant $T$, $K_T$; $\partial V(0)/\partial T|_{K_0,K_T}$ is the derivative of $V(0)$ w.r.t. $T$ for constant $K_0$, $K_T$; and $\partial V(0)/\partial K_T|_{K_0,T}$ is the derivative of $V(0)$ w.r.t. $K_T$ for constant $K_0$, $T$. Start with the first term

$$\left. \frac{\partial V(0)}{\partial K_0} \right|_{T,K_T} = \frac{\partial}{\partial K_0} \left. \left\{ \int_0^{T^*} \pi(*) e^{-\int_0^t r(x)dx} dt + p_K(T^*)K(T^*)e^{-\int_0^{T^*} r(x)dx} \right\} \right|_{T,K_T},$$

(28)

where $T^*$ denotes the optimal duration of operations and $\pi(*)$ denotes the maximized profit function, i.e., for the optimal solution, denoted by $q^*_K(t)$, $I^*(t)$, $L^*(t)$, and $T^*$. By definition the optimal solution satisfies (2), (3), (4), (6), (8), (9), (10), and (11).

First, reorganize (3) as follows $G[I(t), K(t), t] - \delta(t)K(t) - \partial K/\partial t = 0$, calculate the derivative w.r.t. $K_0$, multiply by the co-state function $q^*_K(t)$, and integrate from $t = 0$ to $T^*$. Now differentiate the condition $\lambda_I(t)I^*(t)e^{-\int_0^t r(x)dx} = 0$ w.r.t. $K_0$, note that $\partial \lambda_I(t)/\partial K_0 = \partial p_I(t)/\partial K_0 = 0$ (see section 3.2), and integrate from $t = 0$ to $T^*$. Calculate the derivae in (28) using (2) and add it to the previous two identities (note that these two previous identities are both identical to zero) to obtain

$$\left. \frac{\partial V(0)}{\partial K_0} \right|_{T,K_T} = \int_0^{T^*} \left( p_F(t) \left\{ \frac{\partial F}{\partial K^*} \left|_{T,K_T} \right. + \frac{\partial F}{\partial L^*} \left|_{T,K_T} \right. \right\} e^{-\int_0^t r(x)dx} dt - \int_0^{T^*} \left[ p_I(t) - \lambda_I(t) \right] \left|_{T,K_T} \right. \frac{\partial I^*}{\partial K_0} + w(t) \left|_{T,K_T} \right. \frac{\partial L^*}{\partial K_0} \right] e^{-\int_0^t r(x)dx} dt

+ \int_0^{T^*} q^*_K(t) \left|_{T,K_T} \right. \left( \frac{\partial G}{\partial K^*} \left|_{T,K_T} \right. + \frac{\partial G}{\partial I^*} \left|_{T,K_T} \right. \right) dt

- \int_0^{T^*} q^*_K(t) \left|_{T,K_T} \right. \left( \delta(t) \frac{\partial K^*}{\partial K_0} + \frac{\partial}{\partial t} \frac{\partial K^*}{\partial K_0} \right) dt.\)
Now use the first-order conditions for investment (8) and labor (9), and the dynamic co-state equation (10), to obtain

\[
\frac{\partial V(0)}{\partial K_0} \bigg|_{T,K_T} = - \int_0^{T^*} \left( \frac{\partial q_K^*}{\partial t} \frac{\partial K^*}{\partial K_0} + q_K^*(t) \frac{\partial K^*}{\partial t} \right) dt - q_K^*(0) \frac{\partial K^*(t)}{\partial K_0} \bigg|_{T,K_T,t=0} - q_K^*(T^*) \frac{\partial K^*(t)}{\partial K_0} \bigg|_{T,K_T,t=T^*} = q_K^*(0),
\]

(30)

where in the last step we have used \( \partial K^*(t)/\partial K_0|_{T,K_T,t=0} = 1 \) and \( \partial K^*(t)/\partial K_0|_{T,K_T,t=T^*} = 0 \), since \( K^*(t = 0) = K_0 \) and \( K_0 \) is constant, and \( K^*(t = T^*) \) is fixed at \( K_T \) for variation that takes \( K_T \) constant.

Similarly, differentiating w.r.t. \( T^* \) and taking identical steps, one obtains

\[
\frac{\partial V(0)}{\partial T^*} \bigg|_{K_0,K_T} = \int_0^{T^*} \left( p_F(t) \left( \frac{\partial F}{\partial K^*} \frac{\partial K^*}{\partial T^*} \bigg|_{K_0,K_T} + \frac{\partial F}{\partial L^*} \frac{\partial L^*}{\partial T^*} \bigg|_{K_0,K_T} \right) \right) e^{-\int_0^T r(x) dx} dt
- \int_0^{T^*} \left( p_I(t) - \lambda_I(t) \right) \left( \frac{\partial I^*}{\partial T^*} \bigg|_{K_0,K_T} + w(t) \frac{\partial L^*}{\partial T^*} \bigg|_{K_0,K_T} \right) e^{-\int_0^T r(x) dx} dt
+ \int_0^{T^*} q_K^*(t) \left( \frac{\partial G}{\partial K^*} \frac{\partial K^*}{\partial T^*} \bigg|_{K_0,K_T} + \frac{\partial G}{\partial L^*} \frac{\partial L^*}{\partial T^*} \bigg|_{K_0,K_T} \right) dt
- \int_0^{T^*} q_K^*(t) \left( \delta(t) \frac{\partial K^*}{\partial T^*} \bigg|_{K_0,K_T} + \frac{\partial G}{\partial T^*} \frac{\partial K^*}{\partial T^*} \bigg|_{K_0,K_T} \right) dt.
+ \pi^*(T^*) e^{-\int_0^{T^*} r(x) dx} + \frac{\partial}{\partial T^*} \left( p_K(T^*) K(T^*) e^{-\int_0^{T^*} r(x) dx} \right)
= - \int_0^{T^*} \left( \frac{\partial q_K^*}{\partial t} \frac{\partial K^*}{\partial T^*} \bigg|_{K_0,K_T} + q_K^*(t) \frac{\partial K^*}{\partial t} \frac{\partial T^*}{\partial T^*} \bigg|_{K_0,K_T} \right) dt
+ \pi^*(T^*) e^{-\int_0^{T^*} r(x) dx} + \frac{\partial}{\partial T^*} \left( p_K(T^*) K(T^*) e^{-\int_0^{T^*} r(x) dx} \right)
= q_K^*(0) \frac{\partial K^*(t)}{\partial T^*} \bigg|_{K_0,K_T,t=0} - q_K^*(T^*) \frac{\partial K^*(t)}{\partial T^*} \bigg|_{K_0,K_T,t=T^*}
+ \pi^*(T^*) e^{-\int_0^{T^*} r(x) dx} + \frac{\partial}{\partial T^*} \left( p_K(T^*) K(T^*) e^{-\int_0^{T^*} r(x) dx} \right)
= \pi^*(T^*) + \frac{\partial}{\partial T^*} \left[ p_K(T^*) K(T^*) e^{-\int_0^{T^*} r(x) dx} \right] = 0,
\]

(31)

where we have used \( \partial \lambda_I(t)/\partial T^* = \partial p_I(t)/\partial T^* = 0 \) (see section 3.2), the definition of the Lan-
grangian (4), and $\partial K^*(t)/\partial T^*|_{T,K_T,t=0} = 0$ since $K^*(t = 0) = K_0$ and $K_0$ is constant.

Further, we have used

$$0 = \frac{\partial K^*(T^*)}{\partial T^*} \Bigr|_{K_0,K_T} = \frac{\partial K^*(t)}{\partial t^*} \Bigr|_{K_0,K_T,t=T^*} + \frac{\partial K^*(t)}{\partial T^*} \Bigr|_{K_0,K_T,t=T^*.} \quad (32)$$

Condition (32) holds since $\partial K^*(T^*)/\partial T^*|_{T,K_T,t=T^*} = \partial K_T/\partial T^*|_{T,K_T} = 0$ (since $K_T$ is held constant) and because the optimal duration of operations $T^*$ serves two functions, the first being the role of time at $t = T^*$, and the second being the role of the parameter $T^*$ (the end point). These are the two terms on the right-hand side of (32). By definition the optimal duration of operations $T^*$ maximizes the objective function $V(0)$ so that the condition $\partial V(0)/\partial T^*|_{K_0,K_T} = 0$ holds (very last step of 31).

Similarly, differentiating w.r.t. $K_T^*$ and taking identical steps, one obtains

$$\frac{\partial V(0)}{\partial K_T^*} \Bigr|_{K_0,T} = -\int_0^{T^*} \left( \frac{\partial q^*_K}{\partial t} \frac{\partial K^*}{\partial K_T^*} \Bigr|_{K_0,T} + q^*_K(t) \frac{\partial K^*(t)}{\partial t} \frac{\partial K_T^*}{\partial K_T^*} \Bigr|_{K_0,T} \right) dt + p_K(T^*)e^{-\int_0^{T^*} r(x)dx} = q^*_K(0) \frac{\partial K^*(t)}{\partial K_T^*} \Bigr|_{K_0,T,t=0} - q^*_K(T^*) \frac{\partial K^*(t)}{\partial K_T^*} \Bigr|_{K_0,T,t=T^*} + p_K(T^*)e^{-\int_0^{T^*} r(x)dx} = p_K(T^*)e^{-\int_0^{T^*} r(x)dx} - q^*_K(T^*) = 0, \quad (33)$$

where we have used $\partial I(t)/\partial K_T^* = \partial p_I(t)/\partial K_T^* = 0$ (see section 3.2). By definition the optimal level of the end period stock of capital $K^*(T^*) = K_T^*$ maximizes the objective function $V(0)$ so that the condition $\partial V(0)/\partial K_T^*|_{K_0,T} = 0$ holds (very last step of 33).

Combine (27), (30), (31), and (33) to obtain

$$\frac{\partial V(0)}{\partial K_0} = \frac{\partial V(0)}{\partial K_0} \Bigr|_{T,K_T} = q_K(0). \quad (34)$$

Q.E.D.
B Equivalence of marginal $Q$ and average $Q$: special case

From (3) and (10) we obtain

$$
\frac{\partial [q_K(t)K(t)]}{\partial t} = \frac{\partial q_K(t)}{\partial t}K(t) + q_K(t)\frac{\partial K(t)}{\partial t}
= -p_F(t)\frac{\partial F[]}{\partial K}K(t)e^{-\int_0^t r(s)ds} - q_K(t)\left[\frac{\partial G[]}{\partial K} - \delta(t)\right]K(t)
+ q_K(t)\{G[I(t), K(t), t] - \delta(t)K(t)\}.
$$

(35)

Using the first-order conditions (8), (9) and, following Hayashi (1982), assume that both the production function for the final good $F[.]$ and the production function for capital $G[.]$ are homogenous functions of degree one,\(^{22}\) we obtain

$$
\frac{\partial [q_K(t)K(t)]}{\partial t} = \{p_I(t)I(t) + w(t)L(t) - p_F(t)F[.]\} \ e^{-\int_0^t r(s)ds}
= -\pi(t)e^{-\int_0^t r(s)ds},
$$

(36)

where, in the last step, we used (2). Now integrate the above differential equation from $t = 0$ to $t = T$ and impose the condition (6) to get

$$
q_K(0)K(0) = \int_0^T \pi(s)e^{-\int_0^s r(x)dx}ds + p_K(T)K(T)e^{-\int_0^T r(x)dx} \leftrightarrow
q_K(0) = \frac{V(0)}{K(0)}.
$$

(37)

We have essentially replicated the proof by Hayashi (1982) that marginal $Q$ equals Tobin’s average $Q$, if both the production function for the final good and the production function for the capital stock are homogenous of degree one.

Note that in deriving this result we have applied the first-order conditions (8), (9), the co-state equation (10), the initial condition $K(0)$ and the transversality condition (6). We have not yet applied the condition for the optimal duration of operation $T$ (7), but this does not affect the above result (it simply determines $T^*$) so that the result also holds for the optimal $T^*$. Since we have applied all optimality, initial, end, and transversality conditions to obtain the above result we can replace $T$ by its optimal value $T^*$ and $\pi(t)$ by the maximized profit function $\pi(*)$ (i.e., along

\(^{22}\)So that $F[K(t), L(t)] = \frac{\partial F[.]}{\partial K}K(t) + \frac{\partial F[.]}{\partial L}L(t)$ and $G[I(t), K(t)] = \frac{\partial G[.]}{\partial I}I(t) + \frac{\partial G[.]}{\partial K}K(t)$. 

113
Thus we obtain (37).

Note further, that one can formulate the problem starting at any time $t$ (not necessary starting at $t = 0$) and so the above results holds for all $t < T$. 
C Alternative expression for marginal Q

Combining (12) and (13) we have

$$q_K(t) = \frac{\partial}{\partial K(t)} \left\{ \int_t^{T^*} \pi(s)e^{-\int_t^s r(x)dx} ds + p_K(T^*)K(T^*)e^{-\int_t^{T^*} r(x)dx} \right\}$$

$$= \int_t^{T^*} \frac{\partial \pi(s)}{\partial K(s)} \frac{\partial K(s)}{\partial K(t)} e^{-\int_t^s r(x)dx} ds + p_K(T^*) \frac{\partial K(T^*)}{\partial K(t)} e^{-\int_t^{T^*} r(x)dx}$$

$$+ \left\{ \pi(T^*) e^{-\int_t^{T^*} r(x)dx} + \frac{\partial}{\partial T^*} \left[ p_K(T^*)K(T^*)e^{-\int_t^{T^*} r(x)dx} \right] \right\} \frac{\partial T^*}{\partial K(t)}, \forall s \in [t, T^*]$$ (38)

Integrate (3) and use the initial condition $K(s = 0) = K_0$ to obtain

$$K(s) = K_0 e^{-\int_0^s \delta(x)dx} + \int_0^s G[I(u), K(u), u] e^{-\int_u^s \delta(x)dx} du$$ (39)

Now take the derivative of $K(s)$ w.r.t. $K(t)$ to obtain

$$\frac{\partial K(s)}{\partial K(t)} = \frac{\partial G[I(t), K(t), t]}{\partial K(t)} e^{-\int_t^s \delta(x)dx}, \forall t \in [0, s]$$ (40)

Both (38) and (40) are valid for $0 \leq t \leq s \leq T^*$. Substitute (40) into (38) to obtain

$$q_K(t) = \frac{\partial G(t)}{\partial K(t)} \left\{ \int_t^{T^*} \frac{\partial \pi(s)}{\partial K(s)} e^{-\int_t^s [r(x) + \delta(x)]dx} ds + p_K(T^*)e^{-\int_t^{T^*} [r(x) + \delta(x)]dx} \right\}$$

$$+ \left\{ \pi(T^*) e^{-\int_t^{T^*} r(x)dx} + \frac{\partial}{\partial T^*} \left[ p_K(T^*)K(T^*)e^{-\int_t^{T^*} r(x)dx} \right] \right\} \frac{\partial T^*}{\partial K(t)}, \forall s \in [t, T^*]$$ (41)
D Measurement error

All variable names and definitions are taken directly from the Compustat online manual.

**Total Assets:** This item represents current assets plus net property, plant, and equipment plus other noncurrent assets, including intangible assets, deferred items and investments and advances. The item is the sum of:

- Current Assets - Total (ACT)
- Property, Plant and Equipment (Net) - Total (PPENT)
- Investment & Advances - Equity (IVAEQ)
- Investment & Advances - Other (IVAO)
- Intangible Assets - Total (INTAN)
- Assets - Other - Total (AO)

**Current Assets - Total (ACT):** This item represents cash and other assets that are expected to be realized in cash or used in the production of revenue within the next 12 months. It is the sum of

- Cash and Short-Term Investments (CHE) (CHSTI on RI)
- Current Assets - Other ? Total (ACO)
- Inventories - Total (INVT)
- Receivables - Total (RECT)

**Property, Plant and Equipment - Net (PPENT):** This item represents the cost, less accumulated depreciation, of tangible fixed property used in the production of revenue.

- Advances to vendors for plant expansion programs
- Capitalized leases
- Construction in progress and funds for construction, including funds held by trustees
• Display fixtures
• Equipment leased to others
• Improvements to leased or rental properties
• Intangibles, included on Schedule V by the company
• Leaseholds and leasehold improvements, unless presented as an intangible by the company
• Patterns
• Pollution abatements
• Property held for future use
• Tools and dies
• Unexpended proceeds of industrial revenue bonds
• Airline companies’ deposits and advances on flight equipment
• Banking companies’ and savings and loan companies’ office premises and equipment, net only
• Beverage producers’ bottles, kegs and cases
• Broadcasting companies’ broadcast rights
• Extractive industries’
• Exploration and development expenditures
• Investment in oil and gas properties at cost
• Mining concessions and undeveloped leases
• Patents and franchises on foreign property
• Prepaid mine development and stripping
• Seismic libraries
• Finance and insurance companies’ title plants
- Forestry and paper companies’ timberlands and timber rights
- Franchise rights and broadcast licenses
- Real estate companies’ and land developers’ land held for development and sale
- Shipping companies’ statutory reserve funds and allowances from the Maritime Administration for vessels traded in to be used for vessels under construction
- Computer software included in property, plant and equipment by the company

This item excludes:

- Computer software excluded from property, plant and equipment by the company, included in Assets ? Other
- Excess cost over value of property, included in Intangibles
- Idle land, included in Assets ? Other
- Goodwill, patents, and other intangibles, included in Intangibles
- Long-term inventory, included in Assets ? Other
- Non-real estate companies’
- Property purchased and held for investment, included in Investments and Advances ? Other
- Land held for resale, included in Investment and Advances ? Other
- Property not used in operations, included in Assets ? Other
- Property of discontinued operations, included in Assets ? Other
- Broadcasting companies
- Program rights, included in Deferred Charges
- Motion picture industries’ film distribution systems, included in Assets ? Other
**Investments & Advances - Equity (IVAEQ):** This item represents long-term investments and advances to unconsolidated subsidiaries and affiliates in which the parent company has significant control. It includes:

- All investments at equity
- Goodwill related to investments at equity
- Receivables from investments at equity

It excludes:

- All investments at cost, included in Investments and Advances - Other
- Joint ventures, when there is no indication the investment is at equity, included in Investments and Advances - Other
- Joint ventures not yet operating, included in Investments and Advances - Other

**Investments and Advances-Other (IVAO):** This item represents long-term receivables and other investments and advances including investments in unconsolidated companies in which there is no control. It includes:

- All investments carried at cost
- Direct financing leases when the company is the lessor
- Investments and advances to former subsidiaries
- Investments and advances to subsidiaries to be sold
- Joint ventures not yet operating
- Land held for resale, for companies whose primary business is not land development
- Leveraged leases when the company is the lessor
- Long-term receivables, including receivables from parent
- Marketable securities, unless restricted or held for collateral
• Investments in securities and mortgage loans
• Seat on or membership in a securities exchange
• Finance companies’ assets held strictly for investment purposes
• Publishing companies’ royalty advances to authors
• Real estate investment trust companies’
• Equity investments in real estate
• Mortgage loans on real estate
• Property acquired through foreclosure
• Extractive industries’ oil and gas

This item excludes:
• Advances to salesmen, included in Assets ? Other
• Equity in consolidated joint ventures when held for loan collateral, included in Assets ? Other
• Investments carried at equity, included in Investments and Advances ? Equity Method
• Investments in a company’s own securities, included in Assets ? Other
• Receivables from officers and directors, included in Assets ? Other
• Film production companies’ film costs, included in Property, Plant and Equipment
• Land development companies’ land held for development and sale
• Publishing companies’ royalty advances to authors, included in Deferred Charges

Intangible assets (INTAN): This item consists almost exclusively of the excess of cost over equity acquired in assets of purchased subsidiaries which are still unamortized or not eliminated by a direct charge to a capital account.
• Blueprints or building designs
• Copyrights
• Covenants not to compete
• Design costs
• Distribution rights and agreements
• Easements (i.e., gas rights, mineral rights, water rights)
• Engineering drawings
• Excess of cost or premium of acquisition (except on unconsolidated subsidiaries)
• Franchise and franchise fees
• Goodwill (except on unconsolidated subsidiaries)
• Import quotas
• Leases and lease acquisition costs when company is the lessee
• Leasehold expense when company is the lessee
• Licenses
• Operating rights
• Organizational expense
• Patents
• Subscription lists
• Trademarks and tradenames
• Transportation companies’ route acquisition costs
• Computer software patents
• Software or software costs (included in Assets - Other Excluding Deferred Charges)
It excludes:

- Contracts (included in Assets ? Other Excluding Deferred Charges)
- Deferred charges (included in Deferred Charges)
- Deferred financing costs (included in Deferred Charges)
- Film development costs (included in Property, Plant, and Equipment ? Net)
- Goodwill on unconsolidated subsidiaries (included in Investments and Advances ? Equity Method)
- Intangibles included in property, plant, and equipment by the company
- Preopening expenses (included in Deferred Charges)
- Prepaid Expenses (included in Deferred Charges)
- Start-up costs (included in Deferred Charges)
- Unamortized debt discount and expense (included in Deferred Charges)

**Total Assets - Other (AO):** This item represents noncurrent assets that cannot be classified as property, plant and equipment/ tangible fixed assets, investments and advances, or intangible assets. It includes:

- Acquisition costs
- Advances to sales staff
- Amounts due from directors, officers, employees, and principal holders of equity securities other than unconsolidated subsidiaries
- Assets of discontinued operations
- Banks and savings and loans’ acceptances and all other assets
- Broadcasters’ program rights, film productions, film rights
- Cash on deposit pursuant to loan agreements
• Cash surrender value of life insurance policies
• Claims in litigation
• Contracts
• Deferred charges
• Deferred financing costs
• Deferred policy costs
• Deferred taxes
• Deposits
• Equity in consolidated joint ventures when held for loan collateral
• Finance service companies’ deferred finance charges on installment obligations when presented as a deduction from receivables
• Idle land
• Investments in a company’s own securities
• Long-term inventory
• Long-term prepaid expenses
• Materials and supplies
• Minority interest in consolidated subsidiaries
• Motion picture companies’ film distribution systems
• Negative goodwill
• Pension and other special funds
• Pre-opening expenses or of discontinued operations
• Prepaid pension costs, if reported as a separate line item in long-term or noncurrent assets
• Property not used in operations or of discontinued operations

• Publishing and prepublication costs

• Publishing companies’ royalty advances to authors

• Purchased technology

• Restricted cash

• Start-up costs

• Stock issuance costs

• Timberlands other than those owned by forest and paper companies

• Tooling costs

• Treasury stock reported on the asset side of the balance sheet

It excludes:

• All items specifically labeled as intangibles by Standard & Poor’s definitions

• Computer software included in property, plant, and equipment on the balance sheet or on Schedules V and VI by the company

• Prepaid pension costs when included by the company in another item