

# Mind the Gap: Analyzing the impact of data gap in Millennium Development Goals' (MDGs) indicators on the progress towards MDGs

Arun Jacob

**United Nations Conference on Trade and Development (UNCTAD)**

*The views expressed are personal and need not reflect the views of the United Nations or its member states*

**Annual Bank Conference on Development Economics**

June 20, 2016

# Overview

- 1 Motivation
- 2 MDG Framework
- 3 Theoretical Framework
- 4 Empirical Model & Data
- 5 Key Results
- 6 Discussion
- 7 Appendix

# Motivation

- 8 MDGs, 21 targets and 60 indicators → 17 SDGs, 169 targets & 230 (approx) indicators
  - The need for harnessing the data revolution for SDGs.
  - US\$ 15 billion expected to be invested for data collection for SDGs.
- Taylor (2009) notes that there are relatively limited number of empirical studies on the impact of performance measurement.

***Important to gather evidence from the MDG experience on the significance of data collection and performance monitoring***

# MDG Framework

- The 'quantified' approach of MDG a major advantage and forced an unrelenting spotlight on the need for better data (Aryeetey 2012).
- Several Criticisms of the framework: Easterly (2009), Karver (2012) and Gauri (2012)
- Problem of Data Availability in MDG framework. Chen (2013) finds that nearly a third of MDG indicators lack data for more than half of the countries.
- Debate on how to measure MDG Performance : Fukuda (2010), Vandemoortele (2014), Clemens (2007), ODI (2010a),

# Theoretical Framework

- The core purpose of performance measurement is to enable use of information generated in decision making purposes and eventually to better performance (Cavalluzzo 2004, Taylor 2009, Hatry 2007, Sanger 2013)
- Wang (2000) identifies the following mechanism: performance measurement → budgeting → improved performance
- Behn (2003) states that the leaders of public agencies use performance measurement to (1) evaluate; (2) control; (3) budget; (4) motivate; (5) promote; (6) celebrate; (7) learn; and (8) improve.

# Theoretical Framework

- MDG indicators (the performance measures) were integrated to a large extent within national plans, budgets and strategies (AbouZahr 2007, UNDP 2010, Sanga 2011).
- Several challenges exist in the effective use of performance measurement : political, cultural factors (Newcomer 1997, Julnes 2001) & integration of performance measurement within the decision making system.

## Central Hypothesis

***H1: Efficient (inefficient) performance measurement system improves (reduces) the probability of performance success***

*'You can't manage what you don't measure'*

W. Edwards Deming & Peter Drucker

# Data and Key Variables

- Official MDG indicators data provided by the UN, WDI database of the World Bank.
- We use the 22 quantifiable indicators falling under 7 MDGs.

# Data and Key Variables

**Table: MDG Indicators Used in the Analysis**

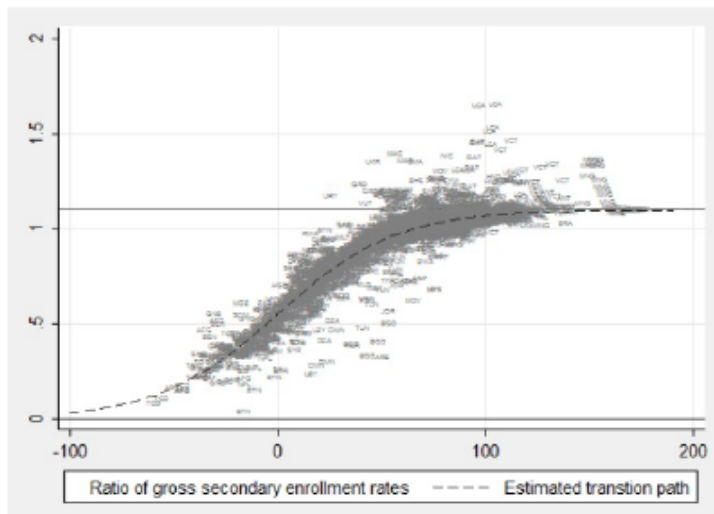
MDG	Indicators
Goal 1	1.1 Proportion of population below \$1.25 (PPP) per day
	1.2 Poverty gap ratio
	1.3 Prevalence of underweight children under-five years of age
	1.4 Proportion of population below minimum level of dietary energy consumption
	1.5 Employment-to-population ratio
	1.6 Proportion of employed people living below \$1.25 (PPP) per day
Goal 2	2.1 Net enrolment ratio in primary education
	2.2 Proportion of pupils starting grade 1 who reach last grade of primary
	2.3 Literacy rate of 15-24 year-olds, women and men
Goal 3	3.1 Gender Parity Index in primary level enrolment
	3.2 Gender Parity Index in secondary level enrolment
	3.3 Gender Parity Index in tertiary level enrolment
Goal 4	4.1 Under-five mortality rate
	4.2 Infant mortality rate
Goal 5	5.1 Maternal mortality ratio
	5.2 Antenatal care coverage (at least one visit)
	5.2 Antenatal care coverage (at least four visits)
	5.3 Unmet need for family planning
Goal 6	6.1 HIV prevalence among population aged 15-49 years
Goal 7	7.1 Proportion of land area covered by forest
	7.2 Proportion of population using an improved drinking water source
	7.3 Proportion of population using an improved sanitation facility



# Data and Variables

- Dependent Variable : Six ways to measure MDG progress
  - Relative Performance Gap : Normalized Gap between actual performance and MDG target in 2012.
  - Absolute Rate of Performance : Actual rate of performance achieved by country at each indicator level in 2012 (Easterly (2009))
  - (Performance Dummy = 1) if achieved rate of progress of indicators higher than required rate of progress to achieve MDGs (Leo 2010).
  - (Performance Dummy = 1) if Rate of Progress post MDG phase is greater than pre MDG Adoption (Fukuda-Parr et al. [2013])
  - Difference in Average Annual Rate of Reduction (AARR) between post and pre MDG (UNICEF)
  - Deviation from historical transition path (Klasen 2011)

Figure 4: S-shaped historical transition path of ratio of male to female gross enrolment ratio in secondary education 1970-2009, source: Klasen and Lange [2011]



# Data and Variables

- Key Independent Variables
  - Data gap variable: Number of missing data points for 2000 to 2012 for each indicator at country level (both absolute values and normalized values at indicator level)
  - Statistical Capacity, Infrastructure (electricity access), Public spending on social sectors (Education and Health Expenses as % of GDP), GNI per capita and growth rate, Foreign Aid, Gini, population, Proportion of Estimated or Modeled Data Points, Indicator dummies

# Empirical Model

$$MDG\_performance_{c,i} = \beta_1 Data\_gap_{c,i} + \beta_2 x_{c,i} + \beta_3 x_c + \beta_4 x_i + \epsilon_{c,i}$$

$x_{c,i}$  : country & indicator level covariates;  $x_c$  : country specific covariates  
 $x_i$  : indicator specific covariates

# Empirical Model

- Potential Endogeneity of  $Data\_gap_{c,i}$  due to omitted variables
  - Dealt with using IV 2sls estimation
  - pre-MDG phase data gap in each indicator used as an instrument for  $Data\_gap_{c,i}$  during the MDG phase.
  - the world average of each indicator level data gap used as an IV for robustness check
- Selection bias of the sample
  - Dealt with using the Heckman correction
  - Baseline data availability for each indicator (for 1990) used to define the exclusion restriction for the Heckman correction
- Final estimation combines the Heckman correction with IV 2sls

Table 11: Detailed Main Results: Combining Instrumental Variable Estimation and Heckman Correction

Performance Measure : MDG Performance Dummy = 1( <i>RateAchieved</i> ≥ <i>RateRequired</i> )				
	IV 2SLS First Stage	Heckman Selection First Stage	IV+ Heckman Probit	Marginal Effect
	Data Gap	Selection Dummy	MDP Performance Dummy	MDG Performance Dummy
<b>Data Gap</b>			-0.06*** (0.012)	-0.02*** (0.002)
<b>pre-MDG Data Gap</b>	1.09*** (.015)			
Baseline Data Missing		-2.21*** (.07)		
Statistical Capacity	-.010** (.005)	.002 (.004)	0.013** (0.006)	0.004*** (0.001)
Electricity Access (%)	.001 (.002)	-.003 (.001)	0.006** (0.002)	0.002** (0.0001)
Tax Revenue (% GDP)	.008 (.005)	.002 (.004)	0.007 (0.006)	0.002 (0.001)
Health Expenditure (% GDP)	.009 (.024)	-.015 (.019)	0.001 (0.029)	0.0004 (0.0009)
Public Education Expenditure (% of GDP)	.033 (.027)	-.007 (.021)	0.079** (0.032)	0.02** (0.01)
Net ODA	-.008 (.010)	.008 (.005)	0.001 (0.008)	0.0003 (0.002)
Population	.143 (.308)	-.595* (.339)	-0.176 (0.481)	-0.055 (0.151)
Gini Coefficient	-.008 (.007)	-.011 (.005)	-0.011 (0.007)	-0.003 (0.002)
GNI per capita	-.002 (.074)	-.034 (.054)	-0.076 (0.079)	-0.02 (0.02)
GNI per capita squared	-1.33 (5.56)	-1.44 (4.63)	8.55 (6.80)	2.69 (2.15)
GNI per capita growth	-.020 (.019)	0.01 (.013)	0.089*** (0.022)	0.028*** (0.007)
Constant	.784 (.567)	.98*** (.378)	-1.67*** (0.535)	
Indicator Dummies	yes	yes	yes	yes
Number of observations	723	723	723	723

Robust standard errors clustered at national level in parenthesis  
\*, \*\*, \*\*\* indicate 10%, 5% and 1 % level of significance respectively

Table 10: Summary of Main Results: Combining Instrumental Variable Estimation and Heckman Correction

Dependent Variables						
	Relative Performance Gap	Actual Rate of Performance post-2000	Dummy = $1(\text{Rate Achieved} \geq \text{Rate Required})$	Dummy = $1(\text{Rate of progress}_{\text{post-2000}} \geq \text{rate of progress}_{\text{pre-2000}})$	Dummy = $1(\text{AARR}_{\text{post-2000}} \geq \text{AARR}_{1990-2000})$	Performance Index (based on deviation from historical transition path)
Data Gap	.273** (.125)	-.147** (.096)	-.06** (.012)	-.857*** (.11)	-.064*** (.015)	-.244*** (.092)
IMR	.058 (.067)	1.63** (.93)				1.27** (.569)
N	887	587	723	771	369	158

Robust standard errors clustered at national level in parenthesis

\*, \*\*, \*\*\* indicate 10%, 5% and 1 % level of significance respectively

All specifications include the following covariates: electricity access, Tax Revenue (% GDP), Health Expenditure (% GDP), Public Education Expenditure (% of GDP), Net ODA , Population, Gini Coefficient, GNI per capita, GNI per capita squared, Indicator Dummies

In cases where the inverse mills ratio (IMR) is found to be statistically significant, bootstrapped standard errors are reported to account for the generated regressors

AARR is calculated only for six indicators and Performance Index is calculated only for three indicators, as explained in the data section

Table 13: Robustness Checks: Country Fixed Effects IV 2sls Estimations

First Stage Estimation						
	Relative Performance Gap	Actual Rate of Performance post-2000	Dummy = $1(\text{RateAchieved} \geq \text{RateRequired})$	Dummy = $1(\text{Rateofprogress}_{\text{post-2000}} \geq \text{rateofprogress}_{\text{pre-2000}})$	Dummy = $1(\text{AARR}_{\text{post-2000}} \geq \text{AARR}_{1990-2000})$	Performance Index (based on deviation from historical transition path)
pre-MDG Data Gap	0.43*** (0.029)	0.977*** (0.019)	1.1*** (0.014)	1.12*** (0.13)	1.22*** (0.014)	0.606*** (0.09)
R2	0.214	0.74	0.903	0.91	0.961	0.42
Second Stage Estimation						
Data Gap	0.4*** (0.1)	-0.286*** (0.05)	-0.013*** (0.003)	-0.005 (0.003)	-0.017*** (0.004)	-0.069*** (0.005)
N	887	958	723	771	369	158

Robust standard errors clustered at national level in parenthesis

\*, \*\*, \*\*\* indicate 10%, 5% and 1 % level of significance respectively

AARR is calculated only for six indicators and Performance Index is calculated only for three indicators, as explained in the data section



Table 14: Robustness Checks: Alternate IV estimation

First Stage Estimation						
	Relative Performance Gap	Actual Rate of Performance post-2000	Dummy = $1(\text{RateAchieved} \geq \text{RateRequired})$	Dummy = $1(\text{Rateofprogress}_{\text{post-2000}} \geq \text{rateofprogress}_{\text{pre-2000}})$	Dummy = $1(\text{AARR}_{\text{post-2000}} \geq \text{AARR}_{1990-2000})$	Performance Index (based on deviation from historical transition path)
Mean Data Gap	1.015*** (0.02)	0.994*** (0.029)	0.985*** (0.017)	1.12*** (0.13)	0.942*** (0.029)	0.665*** (0.078)
R <sup>2</sup>	0.704	0.84				0.469
Second Stage Estimation						
Data Gap	0.241** (0.122)	-0.08*** (0.023)	-0.041*** (0.014)	0.035 (0.012)	-0.073*** (0.018)	-0.045 (0.041)
N	887	587	723	771	369	158

The IV used in this table is the world average of the data gap of each indicator for the period 2000-2013

Robust standard errors clustered at national level in parenthesis

\*, \*\*, \*\*\* indicate 10%, 5% and 1 % level of significance respectively

AARR is calculated only for six indicators and Performance Index is calculated only for three indicators, as explained in the data section

Table 15: Robustness Checks: Lag Effects

Combining Instrumental Variable Estimation and Heckman Correction Dependent Variables						
	Relative Performance Gap	Actual Rate of Performance post-2000	Dummy = $1(\text{Rate Achieved} \geq \text{Rate Required})$	Dummy = $1(\text{Rate of progress}_{\text{post-2000}} \geq \text{rate of progress}_{\text{pre-2000}})$	Dummy = $1(\text{AARR}_{\text{post-2000}} \geq \text{AARR}_{1990-2000})$	Performance Index (based on deviation from historical transition path)
<i>DataGap<sub>lagged</sub></i>	.55** (.279)	-.291** (.097)	-.155** (.032)	-.952*** (.26)	-.145*** (.035)	-.612*** (.323)
IMR	.01 (.07)	1.34** (.57)				.461 (.363)
N	887	587	723	771	369	158

Robust standard errors clustered at national level in parenthesis

\*, \*\*, \*\*\* indicate 10%, 5% and 1 % level of significance respectively

All specifications include the following covariates: electricity access, Tax Revenue (% GDP), Health Expenditure (% GDP), Public Education Expenditure (% of GDP), Net ODA, Population, Gini Coefficient, GNI per capita, GNI per capita squared, Indicator Dummies

In cases where the inverse mills ratio (IMR) is found to be statistically significant, bootstrapped standard errors are reported to account for the generated regressors

AARR is calculated only for six indicators and Performance Index is calculated only for three indicators, as explained in the data section

# Discussion

- In all specifications, the Data gap variable negatively impacts MDG performance
- Throws light on the positive impact of the 'quantified' approach of MDG and the value of data
- Future research could be conducted to identify the exact channel of impact and mechanisms at play
  - Effect stronger in democracies?
  - Role of civil society
- Implication for SDGs → In addition to improving overall statistical capacity, we need to earmark resources for SDG performance monitoring
- Gather more lessons from the MDG experience

# Appendix

- Exploring the panel dimension of the data.
  - Country as the primary unit and MDG indicators as the second panel dimension.
  - Hausman test rejected FE and the Breusch and Pagan LM test rejected RE in favour of a pooled model.
  - In panel specifications, both in RE and FE, data gap negatively affects the likelihood of performance success and positively impacts the performance gaps.
  - So data used in the analyses is pooled cross-section.

Table 8: Preliminary Results : Probit Estimation Results

Dependent variable : MDG Performance Dummy (Rate Achieved vs Rate Required)							
	(i)	(ii)	(iii)	(iv)	(v)	(vi)	margins
Data Gap	-0.020** (0.007)	-0.024** (0.007)	-0.033** (0.008)	-0.036** (0.009)	-0.044** (0.010)	-0.044** (0.012)	-.016*** (.004)
Electricity Access		0.011** (0.002)	0.011** (0.002)	0.009** (0.002)	0.008** (0.002)	0.007** (0.003)	.002** (.001)
Tax Revenue (% GDP)			0.006 (0.005)	0.002 (0.006)	0.000 (0.007)	0.008 (0.007)	.003 (.002)
Health Expenditure (% GDP)			0.000 (0.020)	-0.003 (0.020)	0.005 (0.023)	0.003 (0.024)	.001 (.008)
Public Education Expenditure (% of GDP)			0.024 (0.028)	0.053* (0.029)	0.067** (0.029)	0.081** (0.028)	.03 *** (.01)
Net ODA			0.009 (0.009)	0.006 (0.009)	0.001 (0.010)	0.001 (0.010)	.000 (.003)
Population				-0.185 (0.254)	-0.135 (0.298)	-0.157 (0.342)	-.05 (.12)
Gini Coefficient				-0.023** (0.005)	-0.019** (0.006)	-0.011 (0.007)	-.004 (.002)
GNI per capita					0.014 (0.015)	-0.086 (0.080)	-.03 (.03)
GNI per capita squared					-0.272 (0.268)	8.758 (6.172)	3.29 (2.33)
GNI per capita growth					0.077** (0.024)	0.093** (0.025)	.03*** (.009)
Statistical capacity						0.014** (0.007)	.005** (.002)
Constant	-0.110* (0.057)	-0.936** (0.132)	-1.127** (0.227)	-0.070 (0.359)	-0.546 (0.375)	-1.591** (0.552)	
Indicator Dummies	yes	yes	yes	yes	yes	yes	yes

## Detailed Main Results: Combining Instrumental Variable Estimation and Heckman Correction

Performance Measure : Relative Performance Gap			
	Heckman Selection	IV2SLS First Stage	IV 2SLS
	Selection Dummy	Data Gap	Relative Performance Gap
<b>Data Gap</b>			0.27*** (0.124)
<b>pre-MDG Data Gap</b>		0.409*** (0.025)	
Baseline Data Missing	-1.676*** (0.067)	.642 (.961)	
Statisticalcapacity	0.002 (0.004)	-0.008** (0.002)	-0.01*** (0.004)
Electricity access	0.000 (0.002)	0.001 (0.001)	-0.01*** (0.002)
Tax Revenue (% GDP)	0.004 (0.004)	0.004** (0.002)	0.002 (0.005)
Health Expenditure (% GDP)	-0.022 (0.018)	-0.017 (0.010)	0.01 (0.023)
Education Expense (% GDP)	-0.018 (0.019)	0.022** (0.009)	-0.049* (0.027)
Net ODA total	0.005 (0.005)	0.001 (0.003)	0.004 (0.006)
Population	-0.839** (0.318)	0.439** (0.131)	-0.08 (0.286)
Gini Coefficient	-0.006 (0.005)	-0.001 (0.002)	-0.003 (0.004)
GNI p.c	-0.009 (0.047)	-0.032 (0.026)	0.002 (0.037)
Squared GNI p.c.	1.412 (3.948)	0.574 (2.049)	-2.07 (2.888)
GNI p.c growth rate	0.007 (0.012)	-0.016** (0.006)	-0.038*** (0.015)
Inverse Mills Ratio		-0.082** (0.029)	0.05 (0.066)
Constant	1.136** (0.334)	0.313* (0.181)	2.35*** (0.329)
Indicator Dummies	yes	yes	yes
R <sup>2</sup>		0.219	0.286
N	2090	887	887

Robust standard errors clustered at national level in parenthesis

\* \*\* \*\*\* indicate 10%, 5% and 1 % level of significance respectively

Table 6: Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Performance Dummy	1445	0.428374	0.495014	0	1
Performance Gap	1743	5.548497	15.10994	-50	100
Dummy( rate of progress post-2000 vs rate of progress pre-2000)	1517	.381	.4859	0	1
Dummy( AARR post-2000 vs AARR pre-2000)	656	.586	.492	0	1
Performance Index	303	.762	1.11	-13.55	4.19
Relative Performance Gap	1743	0.013553	0.97799	-5.28373	5.618526
Data gap post-MDG	5127	0.002542	0.998389	-5.83849	4.492894
Normalized Data gap pre-MDG	5127	0.001723	0.997306	-7.10828	4.492894
Data gap post-MDG	5127	8.160133	5.332443	0	13
Data gap pre-MDG	5127	6.842403	3.948228	0	10
Base Line Data Missing	5127	0.723425	0.447348	0	1
Statistical Capacity Index	3242	65.02608	16.66849	23.93939	94.44444
Electricity Access (%)	4555	73.36732	31.95602	0.8	100
Tax Revenue (% of GDP)	3701	18.63531	11.25909	0.020595	62.91546
Health Expenditure (% of GDP)	4116	6.457485	2.554189	2.068445	18.04439
Public Health Expenditure (% of GDP)	3763	4.641766	1.963763	0.6202	13.07351
Net ODA total (millions)	4577	5.570637	9.481865	-1.67737	74.64549
Total Population	4577	0.031436	0.124163	9.69E-06	1.309265
Gini coefficient	3090	39.90343	9.111138	23.095	64.3
GNI per capita (2005 constant prices)	4248	11.26153	17.91583	0.147813	126.5994
Squared GNI per capita (2005 constant prices)	4248	0.447724	1.461794	2.18E-05	16.02741