Differences in Firm Growth across Countries: Does it Explain GDP Differences?*

M. Jahangir Alam†
HEC Montréal

October 31, 2017

Abstract

Using a sample of only three countries, Hsieh and Klenow (2014) find that firms grow at a slower rate in poorer countries than in richer countries. This paper asks whether their results can be generalized using uniquely well-suited comparable data from the World Bank Enterprise Surveys of 100 countries for 2006 to 2014. I confirm that firms in general grow at a slower rate in poorer than in richer countries. In addition, I establish that firm growth rate explains approximately 16 percent of GDP per capita variations.

Keywords: Employment, Firm growth rate, Patterns, Firm age, GDP

JEL classification: O11, O47

*I am grateful to Joanne Roberts and Trevor Tombe for their valuable comments. I would like to thank Curtis Eaton for helping me develop the theoretical model. I also thank Eugene Beaulieu, Isabel Cairo, Benoit Dostie, Danny Leung, and Lars Vilhuber for comments and discussions. For helpful comments, I thank seminar participants at the Canadian Economic Association meeting 2015, Toronto and University of Calgary, Department of Economics Seminar Series. Last but not least, I would like to thank the World Bank Group for access to the World Bank Enterprise Surveys.

†Department of Applied Economics, HEC Montréal, 3000, chemin de la Côte-Sainte-Catherine, Montréal (Québec), H3T 2A7, Canada (e-mail: jahangir.alam@hec.ca).

1 Introduction

In a study of three countries (the United States, Mexico, and India), Hsieh and Klenow (2014) established that firms grow at a slower rate in poorer countries than in richer countries, suggesting that firm growth rate may be related to GDP per capita differences across countries. However, the idea that firms grow more slowly in poorer countries than in richer ones has not been generalized over a broad selection of countries. In order to generalize this idea, I use firm-level[1] data from the World Bank Enterprise Surveys of 100 countries for 2006 to 2014. This database is uniquely well-suited for this research problem because these surveys use standardized instruments and a similar sampling technique across countries, thereby providing comparable datasets. I find that firms do indeed grow at a slower rate in poorer countries than in richer countries.

I measure the growth rate of firms in terms of the number of employees; hereafter, when I refer to firm growth, I mean the change in the number of employees at a firm. I divide the sample establishments into three age cohorts: less than 10 (age cohort 1), 10 to 19 (age cohort 2), and 20 years or more (age cohort 3). To calculate growth rates, I estimate three age cohort coefficients relative to the youngest age cohort by regressing the number of employees on three age cohort dummies. To verify the sample representativeness of the Enterprise Surveys, I compare my resulting age cohort coefficients for Mexico with those of Hsieh and Klenow (2014). I find that my coefficients for Mexico do not significantly differ from their results.

My results show that the age cohort coefficients are higher in older establishments than in younger establishments, meaning that establishments on average grow as they age. Additionally, the age cohort coefficients differ by country, which means that the growth rate of establishments differs by country. To determine the pattern of this difference, I calculate the correlations between the coefficients of older establishments (cohorts 2 and 3) and GDP per capita, and find that they are positive. I thus conclude that firms in poorer countries grow at a slower rate than those in richer countries.

[1] I use the terms "establishment" and "firm" or "plant" interchangeably in this paper. However, note that the Enterprise Surveys are based on establishment-level. Establishments are defined by a specific physical location; the establishment identifier remains the same, even when the establishment changes ownership. A firm may be composed of one or more establishments.

To see whether the firm growth rate is correlated to GDP per capita (the order of my variable in the regression therefore does not matter), I regress GDP per capita on the age cohort coefficients for cohorts 2 and 3. The resulting coefficients are statistically significant. Moreover, the coefficient for cohort 3 explains approximately 16 percent of GDP per capita variations. In order to provide a meaningful interpretation of this association, I sketch a simple theoretical model showing that the firm growth rate measures the degree of competitiveness of the country's institutional environment: the more competitive that environment in terms of resource allocation, the greater the proportion of inefficient firms exiting the market is. By promoting resource reallocation from the exiting to surviving firms, a highly competitive institutional environment results in higher GDP.

The general idea of the model is this: in every period, N new firms of size 1 (like the age cohort) are created. At the end of every period, a proportion $0 < \lambda < 1$ of the firms in a cohort disappear, and these firms are the least productive in the cohort. A proportion $0 < \delta < 1$ of the employees of the firms that disappear are randomly allocated to the surviving firms. A proportion $1 - \delta$ of these employees are randomly allocated to a firm in the cohort of new firms that will enter the market in the next period. Parameters λ and δ characterize the economy of a particular country, and differ from country to country. Specifically, λ determines how quickly the less productive firms disappear, and δ determines where the employees of the failed firms end up, either in another firm in the same cohort or in a firm in the next cohort entering the market. These parameters measure the extent to which countries' institutions are pro or anti competitive. These parameters determine the steady state GDP of the economy, and in every period $t, (t \in (, 2, 3, ..., \infty))$ for the cohort that entered in period 1, the size, number, and productivity of firms. A process like this produces data in which the countrys GDP is correlated with the relative sizes of the firms in the country. This measures the extent to which differences in the competitive environment across countries explain differences in GDP per capita.

This paper contributes to the empirical literature on firm dynamics, which mainly focuses on firms in the United States. In the United States, manufacturing firms tend to start small and grow substantially as they age (Atkeson and Kehoe, 2005; Haltiwanger et al., 2013). However, firm dynamics in developing countries differ from those in the United States. Hsieh and Klenow (2014) show that, in terms of employment levels, 40-year-old manufacturing plants in the United States are seven times larger than plants less than five years old. In Mexico, by contrast, 25-year-old plants are only a little more than twice the size of younger plants, and employment levels remain unchanged after plant age 25. In India, older plants are no larger than younger plants. Iacovone et al. (2014) show that African firms, at any age, tend to be 20-24 percent smaller than firms in other regions of the world. By studying manufacturing firms in India, Indonesia, and Mexico, Hsieh and Olken (2014) highlight that these countries have a very large number of small firms, but mid-sized and large firms are rare. Thus, the literature on firm dynamics in developing countries is limited due to the lack of comparable data across countries.

The existing literature has some explanations for why older establishments in developing countries are relatively small. Levy (2008) explains that payroll taxes in Mexico are enforced on more stringently large plants. Hsieh and Klenow (2014) find that establishments accumulate less organization capital, which is establishment-specific intangible capital, in India and Mexico than in the United States. Foster et al. (2015) show that establishment growth is largely driven by the rising demand for the plant's products as it ages. Using a field experiment on large Indian textile firms, Bloom et al. (2013) suggest that firms do not grow over the life cycle due to contract enforcement problems, which make it costly to hire the necessary skilled managers.

This paper proceeds as follows. Section 2 provides a detailed description of the Enterprise Survey data, and the steps taken to clean the datasets. Section 3 describes the method to estimate the age cohort coefficients. Section 4 reports the results and discussion. Section 5 provides a sketch of a theoretical model. Finally, this paper concludes in Section 6.

2 Data and Facts

To determine whether firms in poor countries grow at a slower rate than those in rich countries, I use datasets from the World Bank Enterprise Surveys.[2] The Enterprise Surveys, conducted by private contractors on behalf of the World Bank, contains data on 125,000 firms in 139 countries. The sampling unit of the Enterprise Surveys is the establishment. Registered private establishments with 5 or more employees were interviewed; 100% government or state ownership establishments are not included.

The main advantage of the Enterprise Surveys is that it uses standardized instruments and a similar sampling technique across countries, which minimizes measurement error and provides comparable data. In order to ensure the proper representativeness of sample establishments, these surveys use three levels of stratification: two-digit industry, establishment size, and geographical region. The overall sample size in these surveys within an economy depends on the sample size for each level of stratification. Moreover, the stratification relates to the size of the economy measured by gross national income. In general, 150 establishments are surveyed in small economies, 360 establishments in medium-sized economies, and for large economies, 1200 - 1800 establishments are surveyed. A potential problem of the Enterprise Surveys is that the dataset represents only establishments willing to participate in the survey. Due to attrition and reluctance to participate in the survey, additional establishments are surveyed to reach the original target sample size per stratum. In addition, some establishments do not want to disclose all of their financial information (i.e., financial statements). This leads to missing data in the Enterprise Surveys.

I use only the Enterprise Surveys that follow the global methodology. Using the raw country-year datasets from the Enterprise Surveys, I construct a dataset for manufacturing establishments. By comparing each dataset with its respective questionnaire, I match codes and units for each variable. I drop observations if there is missing or unknown data pertaining to sales or the number of full-time employees, and exclude observations with a value of zero for sales or number of full-time employees and drop observations if the sampling weight is

[2] Enterprise Surveys (http://www.enterprisesurveys.org), The World Bank, Downloaded 23 October 2014. Firm-level surveys were conducted since the 1990s by different units within the World Bank. Since 2005-06, most data collection efforts were centralized within the Enterprise Analysis Unit.

missing.

To calculate the number of employees per establishment, I include both full-time permanent and temporary or seasonal employees; however, I measure a full-time permanent employee as equivalent to two full-time seasonal or temporary employees. To calculate the age of establishments, I subtract the year the establishment began operations from the fiscal year of data collection and then add one. If the establishment is missing the year it began operating, then I use the formal registration year. I drop observations if the calculated age of establishments is less than or equal to zero or greater than 100. I divide the sample establishments into three age cohorts: less than 10 years (age cohort 1), 10 to 19 years (age cohort 2), and 20 years or more (age cohort 3). Due to insufficient observations in each of the four-digit industries, I use sales to calculate the weight of an establishment instead of value-added.

To clean the dataset further, I use the number of employees to identify outliers. I use the STATA robust regression method[3] with tune (6.1) that produces either zero or missing weight for outliers. I apply this method to country-year datasets. I drop observations that have either zero or missing weight and exclude a country from my analysis if the total number of establishments is less than ten for a given age cohort.

Table 1 reports the mean employees and number of observations after cleaning, dropping outliers, and excluding countries. I use 138 country-year pair datasets (some countries have multiple years of data) in 100 countries with around 37,000 observations (Table 2). The cleaned dataset contains the highest number of countries for 2013 (42 countries) and the maximum number of establishments in 2009 (around 6,800 establishments). The Enterprise Surveys collect data in some countries more than once. The cleaned dataset contains two-year datasets from 56 countries with around 24,000 establishments (Appendix Table A1). In fact, only two countries have three years of data.

Appendix Table A2 shows the age distribution of establishments by income group. I use the World Bank's income classifications for 2014: high-income, upper middle-income, lower middle-income, and low-income. The mean age of establishments in the high- and low-

[3]Robust regression is an alternative to least squares regression when data has outlier observations. It is also used to detect outliers.

Table 1: Number of employees and observations by size and age of establishment

Size	Age cohort	Before cleaning		After cleaning	
		# of Employees	Observation	# of Employees	Observation
Small	1	10	6,718	10	6,718
Small	2	10	6,541	10	6,541
Small	3	10	4,526	10	4,496
Medium	1	47	4,155	47	3,966
Medium	2	51	5,790	51	5,582
Medium	3	51	5,481	50	5,258
Large	1	439	1,937	165	895
Large	2	440	3,451	183	1,455
Large	3	499	4,727	152	1,672

Source: Enterprise Surveys, 23 October 2014

Note: I use the definition of establishment size by the Enterprise Surveys; establishment sizes are 5 - 19 (small), 20 - 90 (medium), and 100 or more employees (large). If the number of employees in an establishment is less than five, I categorize them into the small establishment group. The age cohorts include less than 10 years (age cohort 1), 10 to 19 years (age cohort 2), and 20 years or more (age cohort 3). I apply the STATA robust regression with tune (6.1) to drop outliers.

Table 2: Countries and establishments over year (#)

Year	# of Countries	# of Establishments
2006	23	6,418
2007	12	5,512
2008	7	1,396
2009	39	6,814
2010	27	5,634
2011	5	1,262
2012	3	2,507
2013	42	6,768
2014	2	272
Total	100	37,000

Source: Enterprise Surveys, 23 October 2014

Note: I use the Enterprise Surveys' datasets from 2006 to 2014. The cleaned dataset contains the highest number of countries in 2013 and the maximum number of establishments in 2009.

income groups are 16 and 14, respectively. The mean age of establishments in middle-income countries lies between the high- and low-income countries. This age distribution suggests that the mean age increased if a country became richer.

Appendix Figure A1 presents the employment distribution by income group. This figure shows that the employment levels grow over establishment age for all income groups. It also indicates that the growth in employment levels differs by income group.

3 Estimation of Age Cohort Coefficients

In this section, I focus on three objectives. First, I show how I estimate the age cohort coefficients. Second, I show that the age cohort coefficients do not differ by year for a given country. This ensures that I can pool country-year datasets. Finally, I compare my result for the age cohort coefficients with those of Hsieh and Klenow (2014). This confirms that the datasets are representative.

As I mentioned, I divide the sample establishments into three age cohorts: less than 10 years (age cohort 1), 10 to 19 years (age cohort 2), and 20 years or more (age cohort 3). To estimate the age cohort coefficients, I regress the employment levels on three age cohort dummies. I use the following regression equation to estimate the age cohort coefficients and their confidence intervals:

$$E_i = \sum_{a=1}^{3} \beta_a A_{ai} + \epsilon_i \qquad (1)$$

where A_{ai} is the age cohort dummy for a given age (a) and establishment (i); E_i is the rescaled employment level of establishments (discussed later); β_a is the mean number of employees for each age cohort relative to the youngest age cohort; and ϵ_i is the disturbance term, which is the unobserved random component of the regression equation.

To show the growth rates of firms over their life cycles, I rescale employment levels relative to the mean employment of the youngest cohort. After rescaling, the coefficient for the youngest cohort is one, and the coefficients for older cohorts can then be interpreted in terms of relative firm growth rates. Since the Enterprise Surveys apply a stratified sampling, I use the sampling weight to estimate equation (1). In addition, since different industries have different contributions to productivity growth, I multiply the sampling weight by the sales weight to capture patterns within industries.

To test whether the age cohort coefficients differ by year for a given country, I apply the interpretation of confidence intervals. I use 138 country-year pair datasets to estimate the age cohort coefficients, which I present in Appendix Table A3.

Table 3: Age cohort coefficients (yearly data)

		Age < 10			**Age 10-19**			**Age ≥ 20**		
Country	Year		Conf. Interval			Conf. Interval			Conf. Interval	
		Coef.	Lower	Upper	Coef.	Lower	Upper	Coef.	Lower	Upper
Mexico	2006	1	0.81	1.19	1.24	1.04	1.44	1.93	1.64	2.22
Mexico	2010	1	0.63	1.37	1.66	1.34	1.98	1.96	1.51	2.42
Nicaragua	2006	1	0.68	1.32	0.80	0.47	1.14	0.57	0.29	0.85
Nicaragua	2010	1	0.76	1.24	3.37	1.67	5.06	2.25	1.31	3.20
Russian Federation	2009	1	0.64	1.36	1.34	0.98	1.70	1.73	1.07	2.39
Russian Federation	2012	1	0.83	1.17	1.10	0.82	1.37	3.53	1.93	5.12

Source: Enterprise Surveys, 23 October 2014

Note: In this table, I use individual datasets for a given country-year pairs.

Since it would be onerous to discuss all country-year pairs, I consider three countries as an example (3). The coefficients of age cohort 2 for Mexico in 2006 and 2010 are 1.24 and 1.66, respectively. Since the lower value of the confidence interval in 2010 is smaller than the higher value of the confidence interval in 2006, the coefficient of age cohort 2 does not significantly differ by year. In addition, the coefficient of age cohort 3 for Mexico 2010 is 1.96 while the coefficient of age cohort 3 for Mexico 2006 is 1.93. Since the lower value of the confidence interval in 2010 is smaller than the higher value of the confidence interval in 2006, I can again conclude that the coefficient of age cohort 3 does not significantly differ by year. Similarly, the coefficients of age cohorts 2 and 3 for the Russian Federation in 2012 do not significantly differ from the coefficients for these age cohorts in 2009. Although the coefficient of age cohort 2 for Nicaragua in 2010 significantly differs from the same age cohort for Nicaragua in 2006, and the coefficient of age cohort 3 for Nicaragua in 2010 significantly differs from that in 2006, in general, the age cohort coefficients do not significantly differ by year. I can therefore pool the datasets for a given country.

To verify the representativeness of the Enterprise Surveys, I compare my results for the age cohort coefficients for Mexico with those of Hsieh and Klenow (2014).[4] In Figure 1, I plot the age cohorts on the horizontal axis and the age cohort coefficients on the vertical axis. On the graph, the line shows the age cohort coefficients, and the shaded areas represent their confidence intervals. The figure shows some inconsistencies in the point estimates, which could be due to either sample selection bias in the Enterprise Surveys or insufficient observations. Due to the Enterprise Surveys over-sampling of large establishments, which are typically older, there are fewer observations in the youngest cohort. However, the estimated age cohort coefficients for Mexico do not statistically differ from those of Hsieh and Klenow (2014). Since their results fall within the confidence intervals of my results, I can conclude that the age cohort coefficients are consistent with their results for Mexico.

[4] By calculating the mean of each age cohort, where the youngest cohort is normalized to one, I derive three age cohorts means from their nine age cohort coefficients. I compare these three age cohorts with the results of the employment distribution using the Enterprise Surveys dataset.

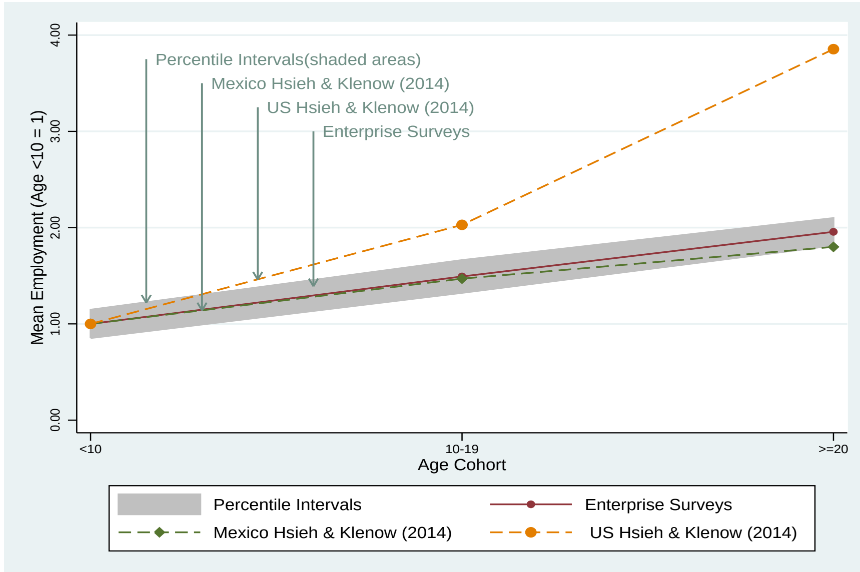

Figure 1: Age cohort coefficients for Mexico

Source: Enterprise Surveys, 23 October 2014

Note: I calculate age cohort coefficients for Mexico and the U.S. using Hsieh Klenow's (2014) figure.

In order to further verify the representativeness of the Enterprise Surveys, I compare my results for the age cohort coefficients for Mexico with those for the United States in Hsieh and Klenow (2014) since the Enterprise Surveys do not provide data for the United States. I also plot their results for the United States in Figure 1. Since their results for the United States fall outside the confidence intervals for Mexico, I can conclude that the estimated age cohort coefficients statistically differ from their results for the United States. Thus, the estimated age cohort coefficients are consistent with the results in Hsieh and Klenow (2014). Since the Enterprise Surveys are representative, I can estimate the age cohort coefficients for a larger set of countries.

4 Results and Discussion

In this section, first, I test two hypotheses. The first hypothesis is that firms in general grow as they age, which I test using the older age cohort coefficients (for example, age cohort 3) that significantly differ from younger age cohort coefficients (for example, age cohort 1). The second hypothesis is that firms grow at different rates in different countries, which I test by examining whether the age cohort coefficients significantly differ by country. Second, I determine the patterns of the age cohort coefficients for a broad collection of countries. Finally, to illustrate the implications of the age cohort coefficients, I regress both TFP and GDP per capita on the older age cohort coefficients. As mentioned, I pool the datasets for a given country to estimate the age cohort coefficients. Table 4 presents the results.

To investigate whether the age cohort coefficients for a given country significantly differ from each other, I first consider Mexico. The coefficient of age cohort 2 is 1.49, which implies that these establishments have 1.49 times more employees than establishments in age cohort 1. Since the lower value of the confidence intervals for age cohort 2 is higher than the higher value of confidence intervals for age cohort 1, the coefficient of age cohort 2 significantly differs from age cohort 1. Similarly, establishments in age cohort 3 have approximately twice the number of employees than establishments in age cohort 1. Applying the confidence interval interpretation, the coefficient of age cohort 3 significantly differs from those of age cohorts 1 and 2. By comparing age cohort coefficients, I can infer that establishments grow as they age in Mexico. Although establishments in some countries do not grow as they age (for example, Georgia), and in some countries, they do not consistently grow as they age (for example, Senegal), I can conclude that establishments in general grow as they age.

To examine whether the age cohort coefficients differ by country, I consider four countries: the Czech Republic (high-income), Mexico (upper-middle income), Senegal (lower-middle income), and Mali (low-income). Applying confidence interval interpretation, the age cohort coefficients 2 and 3 for the Czech Republic are significantly higher than the age cohort coefficients 2 and 3 for Mali. This means that the establishment growth in terms of the number of employees is greater in the Czech Republic than in Mali. Similarly, the age cohort coefficients for Mexico significantly differ from those for Senegal. In some countries, there are three possible scenarios. First, the coefficient of age cohort 2 is similar between two countries, but the coefficient of age cohort 3 significantly differs (Senegal and Mali, for example). Second, the coefficient of age cohort 3 is similar, but the coefficient of age cohort 2 significantly differs. Third, the coefficient for both age cohorts 2 and 3 are similar. Although there are some exceptions in a few groups of countries, I can infer that establishments grow at different rates in different countries.

Table 4: Age cohort coefficients (pooled data)

Country	Age < 10			Age 10-19			Age ≥ 20		
		Conf. Interval			Conf. Interval			Conf. Interval	
	Coef.	Lower	Upper	Coef.	Lower	Upper	Coef.	Lower	Upper
Afghanistan	1	0.82	1.18	0.85	0.63	1.08	2.53	2.25	2.82
Angola	1	0.92	1.08	0.95	0.82	1.08	1.41	1.23	1.60
Argentina	1	0.84	1.16	1.01	0.87	1.15	1.24	1.16	1.33
Armenia	1	0.74	1.26	1.00	0.83	1.17	1.18	0.88	1.48
Azerbaijan	1	0.59	1.41	2.14	1.65	2.63	2.49	1.84	3.14
Bangladesh	1	0.92	1.08	1.13	1.06	1.21	0.98	0.88	1.08
Barbados	1	0.21	1.79	1.66	0.89	2.42	2.72	1.95	3.49
Belarus	1	0.67	1.33	1.29	1.03	1.54	1.99	1.52	2.46
Belize	1	0.40	1.60	1.08	0.69	1.47	1.60	1.20	2.00
Benin	1	0.82	1.18	0.50	0.29	0.71	0.70	0.53	0.87
Bhutan	1	0.77	1.23	1.00	0.47	1.53	0.70	0.23	1.17
Bolivia	1	0.77	1.23	0.93	0.74	1.12	1.37	1.19	1.55
Bosnia and Herzegovina	1	0.76	1.24	0.68	0.54	0.83	1.34	1.13	1.54
Botswana	1	0.68	1.32	1.55	1.20	1.89	3.04	2.56	3.52
Brazil	1	0.86	1.14	1.13	1.01	1.26	2.06	1.97	2.15
Bulgaria	1	0.86	1.14	1.79	1.65	1.92	1.71	1.39	2.02
Burkina Faso	1	0.61	1.39	0.81	0.59	1.03	1.51	0.75	2.28
Cabo Verde	1	-0.60	2.60	0.40	0.29	0.51	2.15	1.96	2.34
Cambodia	1	0.76	1.24	1.08	0.73	1.44	1.37	0.53	2.20
Cameroon	1	0.57	1.43	0.69	0.38	1.00	1.07	0.88	1.26
Chad	1	0.49	1.51	1.75	1.45	2.05	1.67	1.15	2.19
Chile	1	0.83	1.17	0.93	0.82	1.03	0.90	0.83	0.96
China	1	0.90	1.10	1.11	1.03	1.19	1.20	1.03	1.38
Colombia	1	0.88	1.12	1.51	1.38	1.65	1.52	1.38	1.65
Congo, Dem. Rep.	1	0.90	1.10	0.86	0.75	0.97	1.69	1.49	1.89
Costa Rica	1	0.44	1.56	0.83	0.41	1.26	2.32	1.99	2.66
Croatia	1	0.78	1.22	1.02	0.88	1.15	1.53	1.27	1.79
Czech Republic	1	0.40	1.60	2.12	1.83	2.42	2.69	2.21	3.16
Ivory Coast	1	0.82	1.18	1.56	1.28	1.84	1.18	0.72	1.65
Djibouti	1	0.79	1.21	1.18	0.82	1.55	0.37	0.09	0.65
Dominican Republic	1	0.35	1.65	2.22	1.59	2.86	1.25	0.80	1.70
Ecuador	1	0.66	1.34	1.37	1.09	1.65	1.72	1.49	1.96
El Salvador	1	0.79	1.21	1.55	1.30	1.80	2.45	2.16	2.74
Eritrea	1	0.75	1.25	0.70	0.53	0.87	0.95	0.59	1.31
Estonia	1	0.51	1.49	1.39	1.06	1.72	2.82	2.06	3.57

Continued on next page

Table 4 – Age cohort coefficients (pooled data) (continued)

Country	Age < 10			Age 10-19			Age ≥ 20		
		Conf. Interval			Conf. Interval			Conf. Interval	
	Coef.	Lower	Upper	Coef.	Lower	Upper	Coef.	Lower	Upper
Ethiopia	1	0.77	1.23	1.22	0.97	1.47	1.16	0.75	1.57
Georgia	1	0.84	1.16	1.33	1.07	1.58	1.05	0.54	1.57
Ghana	1	0.86	1.14	1.31	1.21	1.42	1.37	1.24	1.51
Guatemala	1	0.78	1.22	0.95	0.68	1.23	2.12	1.90	2.33
Honduras	1	0.57	1.43	0.60	0.25	0.94	1.92	1.53	2.32
Hungary	1	0.51	1.49	1.04	0.76	1.32	1.16	0.49	1.83
Indonesia	1	0.88	1.12	1.25	1.13	1.37	1.21	0.99	1.42
Iraq	1	0.91	1.09	1.12	1.05	1.19	0.62	0.47	0.77
Israel	1	0.58	1.42	0.67	0.37	0.98	0.92	0.71	1.13
Jamaica	1	0.05	1.95	1.29	0.61	1.97	3.34	2.84	3.85
Jordan	1	0.64	1.36	1.44	1.13	1.76	2.33	1.98	2.68
Kazakhstan	1	0.83	1.17	0.94	0.73	1.15	1.95	1.33	2.56
Kenya	1	0.80	1.20	0.78	0.63	0.94	0.78	0.68	0.87
Kosovo	1	0.80	1.20	1.13	0.93	1.33	1.18	0.93	1.43
Kyrgyz Republic	1	0.82	1.18	1.24	1.05	1.42	1.06	0.70	1.42
Lao PDR	1	0.70	1.30	1.58	1.27	1.89	0.70	0.31	1.09
Latvia	1	0.67	1.33	1.39	1.14	1.63	1.96	1.14	2.78
Lebanon	1	0.53	1.47	2.15	1.71	2.58	1.79	1.50	2.08
Lithuania	1	0.52	1.48	1.59	1.15	2.04	2.36	1.89	2.84
Macedonia, FYR	1	0.78	1.22	1.22	0.98	1.45	1.30	0.89	1.71
Madagascar	1	0.81	1.19	1.28	1.12	1.43	0.85	0.60	1.09
Malawi	1	0.66	1.34	0.71	0.41	1.01	1.32	0.92	1.71
Mali	1	0.94	1.06	1.14	1.04	1.24	0.80	0.63	0.97
Mauritius	1	0.64	1.36	1.46	1.00	1.92	3.56	2.98	4.14
Mexico	1	0.85	1.15	1.49	1.32	1.66	1.96	1.81	2.10
Mongolia	1	0.82	1.18	0.87	0.69	1.05	3.15	2.58	3.73
Mozambique	1	0.98	1.02	0.22	0.14	0.29	0.42	0.33	0.50
Myanmar	1	0.81	1.19	0.97	0.80	1.13	1.12	0.82	1.42
Namibia	1	0.66	1.34	1.94	1.52	2.35	1.82	1.32	2.31
Nepal	1	0.84	1.16	1.10	0.98	1.23	1.56	1.31	1.82
Nicaragua	1	0.88	1.12	0.87	0.78	0.97	0.60	0.48	0.73
Niger	1	0.54	1.46	1.64	1.20	2.08	0.82	0.40	1.24
Nigeria	1	0.94	1.06	0.92	0.85	1.00	1.21	1.06	1.35
Pakistan	1	0.86	1.14	1.33	1.22	1.44	1.16	1.05	1.26
Panama	1	0.17	1.83	1.35	0.67	2.03	3.08	2.70	3.45
Paraguay	1	0.68	1.32	1.02	0.85	1.19	1.44	1.26	1.63

Continued on next page

Table 4 – Age cohort coefficients (pooled data) (continued)

Country	Age < 10			Age 10-19			Age ≥ 20		
		Conf. Interval			Conf. Interval			Conf. Interval	
	Coef.	Lower	Upper	Coef.	Lower	Upper	Coef.	Lower	Upper
Peru	1	0.86	1.14	1.07	0.94	1.20	0.94	0.80	1.08
Philippines	1	0.82	1.18	1.64	1.49	1.79	2.15	1.97	2.32
Poland	1	0.51	1.49	1.94	1.73	2.16	0.96	0.57	1.34
Romania	1	0.79	1.21	0.75	0.63	0.88	0.87	0.03	1.71
Russian Federation	1	0.91	1.09	1.18	1.08	1.27	2.39	2.26	2.52
Rwanda	1	0.75	1.25	1.80	1.26	2.35	2.30	1.92	2.67
Senegal	1	0.92	1.08	1.13	1.01	1.24	1.47	1.33	1.62
Serbia	1	0.29	1.71	2.95	2.17	3.74	2.96	2.06	3.86
Slovak Republic	1	0.35	1.65	1.07	0.63	1.50	3.21	2.17	4.25
Slovenia	1	0.47	1.53	1.12	0.86	1.39	1.50	1.11	1.88
South Africa	1	0.83	1.17	1.66	1.47	1.86	2.85	2.63	3.08
Sri Lanka	1	0.66	1.34	1.23	0.99	1.48	1.28	1.04	1.52
St. Lucia	1	0.70	1.30	1.04	0.69	1.39	1.13	0.77	1.48
Swaziland	1	0.58	1.42	2.09	1.55	2.62	1.32	0.48	2.16
Tajikistan	1	0.62	1.38	1.63	1.23	2.03	5.87	5.20	6.53
Tanzania	1	0.85	1.15	1.43	1.29	1.57	1.29	1.06	1.53
Tonga	1	0.65	1.35	2.03	1.02	3.05	2.11	1.80	2.42
Trinidad and Tobago	1	0.35	1.65	2.43	1.51	3.35	2.88	2.21	3.56
Turkey	1	0.85	1.15	1.14	1.02	1.27	1.79	1.65	1.92
Uganda	1	0.94	1.06	1.17	1.03	1.30	1.58	1.41	1.74
Ukraine	1	0.89	1.11	0.98	0.89	1.07	1.64	1.41	1.88
Uruguay	1	0.75	1.25	0.95	0.68	1.22	1.32	1.18	1.46
Uzbekistan	1	0.77	1.23	1.31	0.90	1.71	3.14	2.44	3.85
Venezuela, RB	1	0.57	1.43	3.02	2.40	3.65	1.62	1.10	2.14
Vietnam	1	0.88	1.12	1.53	1.34	1.73	1.69	1.45	1.93
West Bank and Gaza	1	0.74	1.26	0.65	0.49	0.81	0.70	0.54	0.86
Yemen, Rep.	1	0.85	1.15	1.05	0.87	1.23	1.27	1.14	1.40
Zambia	1	0.85	1.15	1.18	1.01	1.35	2.31	2.11	2.50
Zimbabwe	1	0.55	1.45	0.93	0.73	1.14	0.97	0.85	1.09

Source: Enterprise Surveys, 23 October 2014

Note: Due to rescaling, the coefficient of the youngest cohort (less than 10 years) is one. To explain the results, I use Mexico as an example. The coefficient of age cohort 2 (age 10-19) is 1.49, which implies that these establishments have 1.49 times more employees than establishments in age cohort 1 (less than 10 years). Since the lower value of the confidence intervals for age cohort 2 is higher than the highest value of the confidence intervals for age cohort 1, the coefficient of age cohort 2 significantly differs than that for age cohort 1.

To determine the pattern of this difference, I plot the age cohort coefficients (cohorts 2 and 3) and real GDP per capita. To calculate the real GDP per capita, I use the World Bank's Consumer Price Index (CPI) and GDP per capita at current prices in U.S. dollars.[5] I drop Myanmar (2013) for missing GDP per capita. In addition, I drop Eritrea (2008), Uzbekistan (2008 and 2012), Venezuela, RB (2006), and the West Bank and Gaza (2013) for missing the CPI, and Tajikistan due to the outlier coefficient. After dropping these countries, I use 95 countries for the scatter plots.

Figure 2 shows the scatter plot between the coefficient of age cohort 3 and real GDP per capita (see Appendix Figure A2 for the association between the age cohort 2 and real GDP per capita). On the horizontal axis, I use log GDP per capita. On the vertical axis, I use the growth rate of establishments instead of the age cohort 3 coefficient. Since the coefficient of age cohort 1 is one, I add a reference line at zero to depict that the coefficients of age cohort 3 differ from age cohort 1. In addition, I plot a fitted line to explain the association between the coefficient of age cohort 3 and GDP per capita. In a few countries, the growth rate of establishments is less than zero. However, the growth rate of establishments for most countries is positive. This positive growth rate suggests that the size of establishments grow as they age. Figure 2 also shows that the association between the growth rate of establishments and GDP per capita is positive, indicating that establishments in poorer countries grow at a slower rate than those in richer countries do.

To see whether the firm growth rate is correlated to GDP per capita (the order of the variables in the regression therefore does not matter), I regress both TFP and GDP per capita on the age cohort coefficients for cohorts 2 and 3:

$$Y_c = \beta_0 + \beta_1 size2_c + \epsilon_c \tag{2}$$

$$Y_c = \beta_0 + \beta_1 size3_c + \epsilon_c \tag{3}$$

where Y_c is aggregate TFP in country c (discussed later) or GDP per capita in country c;

[5] I use GDP per capita (current U.S. dollars) data from the World Bank instead of the Penn World Table (PWT), where GDP data for sixteen sample countries are missing.

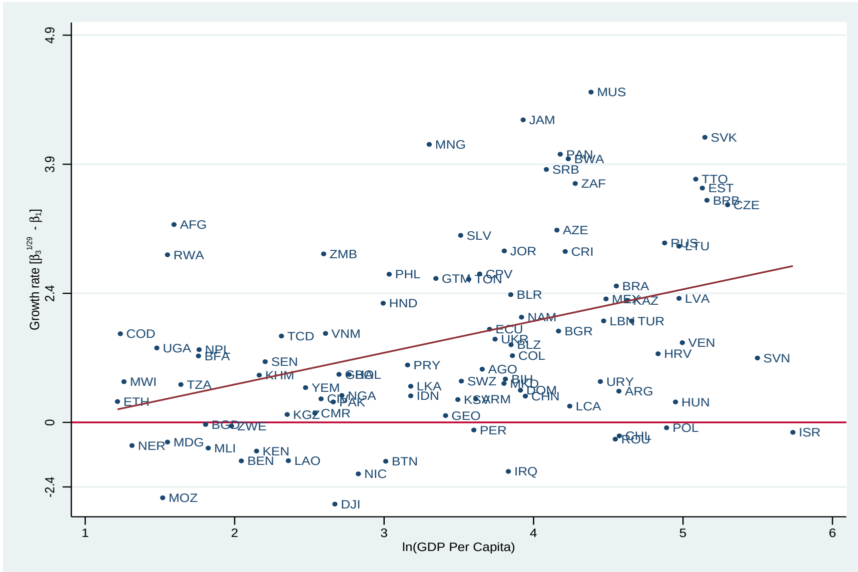

Figure 2: Correlation between the coefficient of age cohort 3 and GDP per capita
Source: Enterprise Surveys, 23 October 2014 and World Development Indicators, 2014

Note: Since the coefficient of age cohort 1 is one, I add a reference line at zero to depict the difference of the coefficients compared to age cohort 1. In addition, I add a fitted line to explain the association between the age cohort coefficients and GDP per capita.

$size2_c$ is the size of establishments for age cohort 2; $size3_c$ is the size of establishments for age cohort 3; and ϵ_c is the disturbance term, which is the unobserved random component of the regression equation.

I use TFP from the Penn World Table (PWT) 8.1 (Feenstra et al., 2015).[6] Since the PWT 8.1 only covers up to 2011, I calculate average TFP using the welfare-relevant TFP levels at current PPPs (USA=1) during 2006 to 2011. I use TFP only for 60 countries due to missing data in the PWT. Table 5 shows the regression results for TFP. The coefficients in the TFP regressions are statistically significant. The R-square of the TFP regression for age cohorts 2 and 3 are 0.05 and 0.09, respectively. These results imply that older age cohort coefficients explain more TFP variations.

Table 5 also shows the regression results for GDP per capita. The coefficients in the GDP per capita regressions are statistically significant. The R-square of GDP per capita regressions for age cohorts 2 and 3 are 0.08 and 0.16, respectively. These results also imply that the older age cohort coefficients explain more GDP variations. The R-square of the

[6] Available for download at www.ggdc.net/pwt

Table 5: Implications of age cohort coefficients

Variables	TFP	TFP	GDP per capita	GDP per capita
Age cohort 2	0.08*		0.68***	
	(1.70)		(2.85)	
Age cohort 3		0.08**		0.65***
		(2.33)		(4.20)
Constant	0.45***	0.42***	2.56***	2.36***
	(6.41)	(6.73)	(7.80)	(8.52)
N	60	60	95	95
R^2	0.05	0.09	0.08	0.16

Source: GDP per capita, World Bank 2014 and TFP, Penn World Table 8.1

Notes: t statistics in parentheses ($^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$). In the TFP regressions, the number of observations is 60 due to missing TFP data in the TFP in PWT. I calculate TFP using data during for 2006 to 2011.

TFP regressions are smaller than the R-square of the GDP per capita regressions. This could be because there are fewer observations in the TFP regressions. This R square value of 0.16 means that the coefficient of age cohort 3 explains approximately 16 percent of GDP per capita variations. Thus, I conclude that the growth rate of establishments explains approximately 16 percent of GDP per capita variations.

5 Theoretical Model

In order to provide a meaningful interpretation of the association between firm growth rates and GDP per capita, I sketch a simple theoretical model showing that firm growth rate measures the degree of competitiveness of the country's institutional environment: the more competitive an environment is in terms of resource allocation, the greater the proportion of inefficient firms that exit the market. By promoting resource reallocation from these exiting to surviving firms, a highly competitiveness institutional environment results in higher GDP.

The general idea of the model is this: a process that eliminates inefficient firms and reallocates their former resources to more efficient firms promotes productivity growth. This process that eliminates inefficient firms differs by country, reflecting the extent to which countries' institutions are pro- or anti-competitive. A process like this produces data in which country productivity is correlated with the relative sizes of firms in the country in periods 2, 3, and so on, after normalization for the relative size of firms in period 1.

In every period, N new firms of size 1 are created. The productivity of these firms is uniformly distributed on interval $[L, H]$, where $H > L \geq 0$. A firm's productivity is a fixed attribute of the firm. The main objective of this model is to follow this cohort of firms through time. At the end of every period, a proportion $0 < \lambda < 1$ of the firms in a cohort disappear, and these are the least productive firms in the cohort. A proportion $0 < \delta < 1$ of the employees of the firms that disappear are randomly allocated to the surviving firms. This ensures that in the next period, all of the surviving firms are of the same size. A proportion $1 - \delta$ of these employees are randomly allocated to a firm in the cohort of new firms that will enter the market in the next period.

This model has five parameters: N, L, H, λ and δ. Parameter N is a scale parameter. Parameters λ and δ characterize the economy of a particular country, where the former determines how quickly the less productive firms disappear and the latter determines where the employees of the failed firms end up, either in another firm in the same cohort or in a firm in the next entry cohort. These parameters determine the steady state GDP of the economy and, in every period $t, (t \in (, 2, 3, ..., \infty))$ for the cohort that entered in period 1, the size, number, and productivity of firms. In Table 6, I present these results for a cohort of firms that enter in period 1.

Table 6: Model specification

Variables	$t = 1$	$t > 1$
n_t	N	$N(1-\lambda)^{t-1}$
s_t	1	$(\frac{1-\lambda+\delta\lambda}{1-\lambda})^{t-1}$
b_t	L	$b_{t-1} + \lambda(H - b_{t-1})$
e_t	$\frac{H+L}{2}$	$e_{t-1} + \frac{\lambda(H-b_{t-1})}{2}$
c_t	$(1-\delta)\lambda N$	$(1-\delta)\lambda N(1-\lambda+\delta\lambda)^{t-1}$

where n_t is the number of firms in the cohort in period t, s_t is the size of firms in period t, b_t is the lower bound on productivity in period t, e_t is the mean productivity in period t, and

c_t is the number of employees sent to the entry pool in period t. Note that the sum from 1 to ∞ of c_t is N, as it must be for the model to make sense.[7]

To express the steady state GDP as a function of the parameters of the model, I first need the closed form solution for b_t and e_t. To find their solutions, I use the conjecture method (see the derivation in Appendix 7.2). The following is the solution for b_t and e_t:

$$b_t = (1-\lambda)^{t-1}L + \frac{1+(1-\lambda)-(1-\lambda)^{t-1}-(1-\lambda)^t}{1-(1-\lambda)^2}\lambda H \quad (4)$$

$$e_t = \frac{H+L}{2} + \frac{\lambda(H-L)}{2}\left[\frac{1+(1-\lambda)-(1-\lambda)^{t-1}-(1-\lambda)^t}{1-(1-\lambda)^2}\right] \quad (5)$$

In period t, this cohort's contribution to GDP is $n_t s_t e_t$. The economy in steady state is composed of an infinite number of cohorts: the one that just entered, the one that entered 1 period before, the one that entered 2 periods before, and so on. That being the case, the steady state GDP is simply the sum from 1 to ∞ of $n_t s_t e_t$:

$$\begin{aligned} GDP &= \sum_{t=1}^{\infty} n_t s_t e_t \\ &= \frac{N(H+L)}{2}\frac{1}{1-a} + \frac{\lambda N(H-L)}{2}\frac{a}{(1-a)(1-ab)} \end{aligned} \quad (6)$$

where, $a = \frac{(1-\lambda)^2}{1-\lambda+\delta\lambda}$ and $b = 1-\lambda$.

To test my models prediction for the association between firm growth rates and GDP per capita, I calibrate the model by creating 100 dummy countries and assuming the value of the five parameters (Table 7). Since parameters λ and δ characterize the economy of a particular country, I assume λ to be uniformly distributed on interval [0.1, 0.15] and δ to be uniformly distributed on interval [0.45, 0.80]. The other assumed parameters are $N = 1000$,

[7]I can show the sum from 1 to ∞ of c_t is N:

$$\begin{aligned} \sum_{t=1}^{\infty} c_t &= (1-\delta)\lambda N + (1-\delta)\lambda N(1-\lambda+\delta\lambda) + (1-\delta)\lambda N(1-\lambda+\delta\lambda)^2 + \cdots \\ &= (1-\delta)\lambda N \frac{1}{1-(1-\lambda+\delta\lambda)} \\ &= N \end{aligned}$$

$H = 50$, and $L = 10$. Using these parameter values, I simulate the model 999 times to estimate steady state firm size and GDP, and then calculate the average of these simulated steady state values. This model predicts that firms grow at a slower rate in poorer countries than in richer countries.

Table 7: Parameter values

Parameter	Definition	Value or Target
N	Number of new firms	1000
L	Lower limit of productivity distribution	10
H	Upper limit of productivity distribution	50
λ	Exit rates	Uniformly distributed on interval [0.1, 0.15]
δ	Allocation of resources to surviving firms	Uniformly distributed on interval [0.45, 0.80]

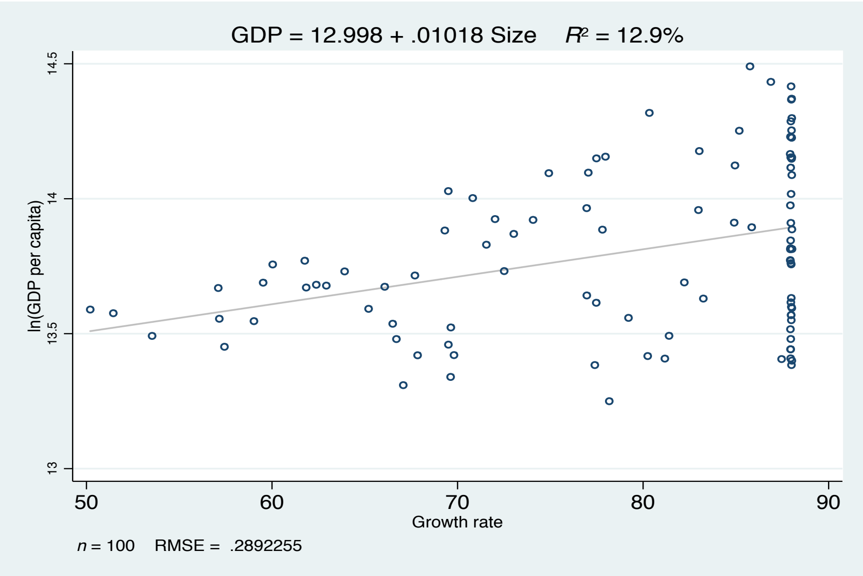

Figure 3: Model predicted association between firm growth and GDP

6 Conclusions

To generalize an established fact that firms grow at a slower rate in poorer countries than in richer countries based on a study of three countries, I use the World Bank Enterprise Surveys of 100 countries covering 2006 to 2014. This database is uniquely well-suited to this research problem because it provides comparable datasets across countries. I confirm that firms do indeed grow at a slower rate in poorer countries than in richer countries. To see whether the firm growth rate is related to GDP per capita across countries, I regress GDP per capita on the age cohort coefficients for cohorts 2 and 3. I find that the resulting coefficients are statistically significant. Moreover, I find that the coefficient for cohort 3 explains approximately 16 percent of GDP per capita variations.

This paper is an important first step in documenting the relationship between the size of establishments and their age. Future research into the reasons for the observed empirical patterns can be promising, especially in identifying the presence of inefficiency and to design well-targeted policy interventions. Additionally, another extension would be to identify the causal effect of the size of establishments on GDP per capita.

References

Atkeson, A. and P. J. Kehoe (2005). Modeling and Measuring Organization Capital. *Journal of Political Economy 113*(5), pp. 1026–1053.

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013). Does Management Matter? Evidence from India. *Quarterly Journal of Economics 128*(1), 1–51.

Feenstra, R. C., R. Inklaar, and M. P. Timmer (2015). The Next Generation of the Penn World Table. *American Economic Review 105*(10), 3150–82.

Foster, L., J. Haltiwanger, and C. Syverson (2015, May). The Slow Growth of New Plants: Learning about Demand? Working paper, University of Chicago.

Haltiwanger, J., R. S. Jarmin, and J. Miranda (2013, May). Who Creates Jobs? Small versus Large versus Young. *Review of Economics and Statistics 95*(2), 347 – 361.

Hsieh, C.-T. and P. J. Klenow (2014, August). The Life Cycle of Plants in India and Mexico. *Quarterly Journal of Economics 129*(3), 1035–1084.

Hsieh, C.-T. and B. A. Olken (2014, Summer). The Missing "Missing Middle". *Journal of Economic Perspectives 28*(3), 89–108.

Iacovone, L., V. Ramachandran, and M. Schmidt (2014, March). Stunted Growth: Why Don't African Firms Create More Jobs? Working Papers 353, Center for Global Development.

Levy, S. (2008). *Good Intentions, Bad Outcomes.* Washington, DC: Brookings Institution Press.

7 Appendix

7.1 Data and age cohort coefficients

Table A1: Frequency of data collection (#)

Frequency	# of Countries	# of Establishments
Once	42	11,479
Twice	56	24,011
Thrice	2	1,093

Source: Enterprise Surveys, 23 October 2014

Note: The Enterprise Surveys collect data in different years for a given country. The cleaned dataset contains two year datasets from 53 countries with around 28,000 establishments that could include multiple surveys. Only three countries have three years data for around 1,600 establishments.

Table A2: Age distribution of establishments by income group

Income Group	Mean Age	Standard Deviation	Observation	Min	Max
High income	16	12	4762	1	100
Upper middle income	16	12	14450	1	100
Lower middle income	15	11	10554	1	100
Low income	14	11	6817	1	100

Source: Enterprise Surveys, 23 October 2014

Note: To calculate the age of an establishment, I subtract the year the establishment began operations from the year of data collection. I use the formal registration year of an establishment if the year first operating year is missing. I delete observation for establishments of more than 100 years of age.

Table A3: Age cohort coefficients (yearly data)

Country	Year	Age < 10			Age 10-19			Age ≥ 20		
			Conf. Interval			Conf. Interval			Conf. Interval	
		Coef.	Lower	Upper	Coef.	Lower	Upper	Coef.	Lower	Upper
Angola	2006	1	0.93	1.07	1.14	0.91	1.37	1.06	0.74	1.38
Argentina	2006	1	0.55	1.45	1.00	0.71	1.28	1.42	1.15	1.68
Argentina	2010	1	0.53	1.47	1.03	0.77	1.30	1.06	0.69	1.43
Armenia	2009	1	0.27	1.73	1.48	1.06	1.89	0.68	0.47	0.89
Armenia	2013	1	0.33	1.67	0.68	0.37	0.99	1.48	0.86	2.10
Azerbaijan	2009	1	0.23	1.77	3.45	1.85	5.06	2.85	1.32	4.39
Azerbaijan	2013	1	0.76	1.24	1.38	0.96	1.80	0.92	0.54	1.31
Bangladesh	2007	1	0.78	1.22	1.12	0.90	1.33	0.79	0.55	1.03
Bangladesh	2013	1	0.73	1.27	1.21	1.00	1.43	1.58	1.32	1.84
Barbados	2010	1	0.60	1.40	1.66	0.85	2.46	2.72	1.54	3.90
Belarus	2008	1	0.58	1.42	0.47	0.19	0.76	1.41	1.02	1.80
Belarus	2013	1	0.40	1.60	2.55	1.79	3.31	2.84	1.83	3.84
Belize	2010	1	0.56	1.44	1.08	0.75	1.42	1.60	1.08	2.12
Benin	2009	1	0.54	1.46	0.50	0.32	0.68	0.70	0.54	0.87
Bhutan	2009	1	0.61	1.39	1.00	0.41	1.59	0.70	0.41	0.99
Bolivia	2006	1	0.65	1.35	0.89	0.63	1.15	1.24	0.94	1.55
Bolivia	2010	1	0.13	1.87	1.59	0.66	2.52	1.91	0.97	2.85
Bosnia and Herzegovina	2009	1	0.71	1.29	0.58	0.39	0.77	1.64	1.06	2.22
Bosnia and Herzegovina	2013	1	0.43	1.57	0.74	0.53	0.95	1.13	0.76	1.50
Botswana	2006	1	0.61	1.39	1.82	1.08	2.55	3.66	2.37	4.94
Botswana	2010	1	0.40	1.60	1.22	0.99	1.45	1.55	0.91	2.18
Brazil	2009	1	0.55	1.45	1.13	0.87	1.40	2.06	1.41	2.70
Bulgaria	2007	1	0.62	1.38	1.16	0.96	1.37	1.80	0.73	2.87
Bulgaria	2013	1	0.44	1.56	2.01	1.50	2.53	1.52	0.75	2.30
Burkina Faso	2009	1	0.76	1.24	0.81	0.17	1.45	1.51	0.49	2.54
Cabo Verde	2009	1	0.49	1.51	0.40	-0.01	0.80	2.15	2.11	2.19
Cambodia	2013	1	0.60	1.40	1.08	0.74	1.43	1.37	0.50	2.23
Cameroon	2009	1	0.56	1.44	0.69	0.40	0.98	1.07	0.54	1.60
Chad	2009	1	0.59	1.41	1.75	1.39	2.11	1.67	0.60	2.74
Chile	2006	1	0.70	1.30	0.90	0.66	1.15	0.85	0.72	0.98
Chile	2010	1	0.68	1.32	0.94	0.68	1.21	0.92	0.74	1.10
China	2012	1	0.86	1.14	1.11	0.98	1.23	1.20	0.89	1.52
Colombia	2006	1	0.78	1.22	1.66	1.34	1.98	1.41	1.14	1.69
Colombia	2010	1	0.50	1.50	1.44	0.92	1.95	1.61	0.88	2.35
Congo, Dem. Rep.	2006	1	0.84	1.16	1.17	0.90	1.44	1.05	0.82	1.29
Congo, Dem. Rep.	2013	1	0.84	1.16	1.19	0.88	1.50	1.64	1.23	2.05
Costa Rica	2010	1	0.18	1.82	0.83	0.46	1.21	2.32	1.23	3.42
Croatia	2007	1	0.62	1.38	0.86	0.70	1.02	4.18	3.33	5.04
Croatia	2013	1	0.52	1.48	1.36	0.72	2.00	1.09	0.78	1.39
Czech Republic	2013	1	0.77	1.23	1.99	1.30	2.69	2.66	1.98	3.35
Ivory Coast	2009	1	0.74	1.26	1.56	0.91	2.21	1.18	0.44	1.93
Djibouti	2013	1	0.63	1.37	1.18	0.92	1.45	0.37	0.09	0.65
Dominican Republic	2010	1	0.32	1.68	2.22	1.48	2.97	1.25	0.68	1.81
Ecuador	2006	1	0.43	1.57	1.56	0.43	2.69	2.33	1.65	3.02
Ecuador	2010	1	0.20	1.80	1.02	0.15	1.88	1.10	0.74	1.46
El Salvador	2006	1	0.78	1.22	1.55	1.05	2.04	2.42	1.30	3.53
El Salvador	2010	1	-0.05	2.05	2.08	0.89	3.27	3.49	2.33	4.65
Eritrea	2009	1	0.65	1.35	0.70	0.55	0.85	0.95	0.56	1.34
Estonia	2009	1	0.55	1.45	1.31	0.68	1.93	3.06	1.56	4.56

Continued on next page

Table A3 – Age cohort coefficients (yearly data) (continued)

Country	Year	Age < 10			Age 10-19			Age ≥ 20		
			Conf. Interval			Conf. Interval			Conf. Interval	
		Coef.	Lower	Upper	Coef.	Lower	Upper	Coef.	Lower	Upper
Estonia	2013	1	0.67	1.33	1.52	0.65	2.39	1.98	0.69	3.28
Ethiopia	2011	1	0.59	1.41	1.22	0.68	1.76	1.16	0.55	1.78
Ghana	2007	1	0.74	1.26	1.31	1.12	1.50	1.38	1.04	1.73
Ghana	2013	1	0.76	1.24	1.37	1.00	1.75	1.25	0.84	1.66
Guatemala	2006	1	0.70	1.30	0.95	0.62	1.29	2.13	0.96	3.30
Guatemala	2010	1	0.40	1.60	0.88	0.38	1.39	1.20	0.85	1.54
Honduras	2006	1	0.12	1.88	0.47	0.19	0.76	1.20	0.22	2.18
Indonesia	2009	1	0.89	1.11	1.25	0.90	1.61	1.21	0.78	1.64
Iraq	2011	1	0.90	1.10	1.12	0.94	1.31	0.62	0.36	0.88
Israel	2013	1	0.74	1.26	0.67	0.47	0.88	0.92	0.34	1.50
Jamaica	2010	1	0.50	1.50	1.29	0.84	1.74	3.34	2.04	4.65
Jordan	2013	1	0.75	1.25	1.44	1.11	1.78	2.33	1.60	3.06
Kazakhstan	2013	1	0.48	1.52	0.93	0.52	1.34	1.29	0.71	1.87
Kenya	2007	1	0.66	1.34	1.34	1.03	1.65	1.66	1.36	1.97
Kenya	2013	1	0.35	1.65	0.48	0.29	0.67	0.51	0.36	0.66
Kosovo	2009	1	0.71	1.29	1.42	1.05	1.78	1.06	0.58	1.54
Kosovo	2013	1	0.69	1.31	0.90	0.55	1.26	1.24	0.43	2.04
Lao PDR	2009	1	0.42	1.58	1.84	1.01	2.68	0.49	0.12	0.86
Lao PDR	2012	1	0.35	1.65	1.02	0.39	1.65	1.31	0.53	2.08
Latvia	2013	1	0.58	1.42	1.40	0.88	1.93	1.91	1.11	2.71
Lebanon	2013	1	0.77	1.23	2.15	0.39	3.90	1.79	1.31	2.26
Lithuania	2013	1	0.43	1.57	2.63	1.41	3.86	2.36	-0.37	5.09
Macedonia, FYR	2009	1	0.49	1.51	1.14	0.62	1.66	1.78	1.21	2.35
Macedonia, FYR	2013	1	0.51	1.49	1.29	0.49	2.08	1.10	0.60	1.59
Madagascar	2009	1	0.56	1.44	0.99	0.70	1.28	0.70	0.41	1.00
Madagascar	2013	1	-0.17	2.17	1.65	0.65	2.64	1.50	0.96	2.04
Malawi	2009	1	0.56	1.44	0.71	0.34	1.09	1.32	0.90	1.73
Mali	2007	1	0.90	1.10	1.17	0.97	1.37	0.98	0.79	1.17
Mali	2010	1	0.64	1.36	0.89	0.15	1.63	0.34	0.32	0.36
Mauritius	2009	1	0.29	1.71	1.46	0.60	2.32	3.56	2.80	4.32
Mexico	2006	1	0.81	1.19	1.24	1.04	1.44	1.93	1.64	2.22
Mexico	2010	1	0.63	1.37	1.66	1.34	1.98	1.96	1.51	2.42
Mongolia	2009	1	0.71	1.29	1.36	0.96	1.76	3.18	1.43	4.93
Mozambique	2007	1	0.95	1.05	0.22	0.18	0.25	0.42	0.32	0.51
Myanmar	2014	1	0.74	1.26	0.97	0.69	1.24	1.12	0.32	1.92
Namibia	2006	1	0.65	1.35	1.94	1.32	2.55	1.82	0.67	2.96
Nepal	2009	1	0.77	1.23	1.16	1.01	1.31	1.38	0.79	1.97
Nepal	2013	1	0.65	1.35	1.03	0.62	1.45	1.63	1.18	2.08
Nicaragua	2006	1	0.68	1.32	0.80	0.47	1.14	0.57	0.29	0.85
Nicaragua	2010	1	0.76	1.24	3.37	1.67	5.06	2.25	1.31	3.20
Niger	2009	1	0.64	1.36	1.64	1.25	2.03	0.82	0.18	1.46
Nigeria	2007	1	0.93	1.07	0.92	0.84	1.00	1.21	0.93	1.49
Pakistan	2007	1	0.72	1.28	1.33	1.03	1.62	1.16	0.98	1.34
Panama	2006	1	0.77	1.23	1.79	1.25	2.33	3.02	1.42	4.61
Paraguay	2006	1	0.76	1.24	1.47	1.11	1.84	1.67	1.13	2.21
Paraguay	2010	1	0.51	1.49	0.92	0.09	1.74	1.40	0.85	1.95
Peru	2006	1	0.33	1.67	0.50	0.35	0.65	0.83	0.62	1.04
Peru	2010	1	0.70	1.30	1.35	0.75	1.95	1.00	0.73	1.27
Philippines	2009	1	0.84	1.16	1.64	0.57	2.71	2.15	1.48	2.81
Poland	2009	1	0.13	1.87	0.92	0.53	1.31	0.96	0.44	1.47

Continued on next page

Table A3 – Age cohort coefficients (yearly data) (continued)

Country	Year	Age < 10			Age 10-19			Age ≥ 20		
			Conf. Interval			Conf. Interval			Conf. Interval	
		Coef.	Lower	Upper	Coef.	Lower	Upper	Coef.	Lower	Upper
Poland	2013	1	0.42	1.58	3.13	3.13	3.13	1.40	0.70	2.10
Romania	2013	1	-0.16	2.16	0.52	0.06	0.99	0.71	0.06	1.35
Russian Federation	2009	1	0.64	1.36	1.34	0.98	1.70	1.73	1.07	2.39
Russian Federation	2012	1	0.83	1.17	1.10	0.82	1.37	3.53	1.93	5.12
Rwanda	2011	1	0.75	1.25	0.99	0.46	1.52	1.87	1.23	2.50
Senegal	2007	1	0.89	1.11	1.13	0.97	1.29	1.47	1.21	1.73
Serbia	2009	1	0.42	1.58	4.07	2.49	5.65	6.10	2.31	9.89
Serbia	2013	1	0.61	1.39	0.82	0.41	1.23	0.82	0.39	1.24
Slovak Republic	2013	1	0.30	1.70	0.80	0.54	1.06	0.93	0.33	1.54
Slovenia	2009	1	0.26	1.74	0.65	0.33	0.96	2.72	1.54	3.90
Slovenia	2013	1	0.72	1.28	2.15	1.49	2.81	1.15	0.59	1.72
South Africa	2007	1	0.83	1.17	1.66	1.37	1.96	2.85	2.49	3.21
Sri Lanka	2011	1	0.50	1.50	1.23	0.74	1.73	1.28	0.77	1.79
St. Lucia	2010	1	0.36	1.64	1.04	0.80	1.28	1.13	0.38	1.87
Swaziland	2006	1	0.67	1.33	2.09	0.99	3.18	1.32	0.08	2.55
Tajikistan	2008	1	0.51	1.49	1.15	0.88	1.42	2.92	2.24	3.61
Tanzania	2006	1	0.71	1.29	1.56	0.99	2.14	1.46	0.89	2.03
Tanzania	2013	1	0.80	1.20	1.32	0.87	1.76	1.09	0.83	1.35
Tonga	2009	1	0.60	1.40	2.03	1.27	2.80	2.11	1.42	2.80
Trinidad and Tobago	2010	1	0.27	1.73	2.43	1.15	3.70	2.88	2.09	3.68
Turkey	2008	1	0.57	1.43	1.32	0.97	1.67	2.24	1.72	2.75
Turkey	2013	1	0.64	1.36	0.91	0.73	1.09	1.11	0.78	1.43
Uganda	2006	1	0.78	1.22	1.10	0.92	1.27	1.66	1.20	2.13
Uganda	2013	1	0.81	1.19	1.36	0.78	1.95	1.51	0.59	2.43
Ukraine	2008	1	0.77	1.23	0.66	0.46	0.86	1.90	1.51	2.28
Ukraine	2013	1	0.65	1.35	1.59	1.34	1.83	1.27	1.10	1.44
Uruguay	2006	1	0.73	1.27	1.13	0.70	1.56	1.10	0.88	1.32
Uruguay	2010	1	0.65	1.35	0.75	0.39	1.12	1.99	1.49	2.49
Uzbekistan	2008	1	0.65	1.35	0.94	0.28	1.60	3.35	1.98	4.73
Uzbekistan	2013	1	0.37	1.63	2.49	1.30	3.68	1.87	-0.15	3.89
Venezuela, RB	2006	1	0.77	1.23	1.12	0.78	1.47	1.27	0.90	1.64
Venezuela, RB	2010	1	0.62	1.38	7.10	-0.63	14.84	6.31	6.05	6.57
Vietnam	2009	1	0.85	1.15	1.53	1.03	2.04	1.69	1.14	2.23
West Bank and Gaza	2013	1	0.22	1.78	0.65	0.43	0.87	0.70	0.44	0.96
Yemen, Rep.	2010	1	0.84	1.16	1.05	0.87	1.24	1.13	0.95	1.32
Yemen, Rep.	2013	1	0.15	1.85	0.85	0.51	1.19	2.82	2.04	3.60
Zambia	2007	1	0.73	1.27	1.15	0.89	1.41	2.71	2.20	3.23
Zambia	2013	1	0.77	1.23	1.21	0.87	1.55	1.46	1.02	1.90
Zimbabwe	2011	1	0.60	1.40	0.93	0.69	1.18	0.97	0.82	1.13

Source: Enterprise Surveys, 23 October 2014

Note: To explain that the employment coefficients significantly differ by year, I use Mexico as an example. The coefficient of age cohort 2 for Mexico in 2010 is 1.66 and 1.24 for 2006. The coefficient of age cohort 2 does not significantly differ by year because the lower value of the confidence interval in 2010 is smaller than the higher value of confidence interval in 2006. Similarly, the coefficient of age cohort 3 does not significantly differ by year.

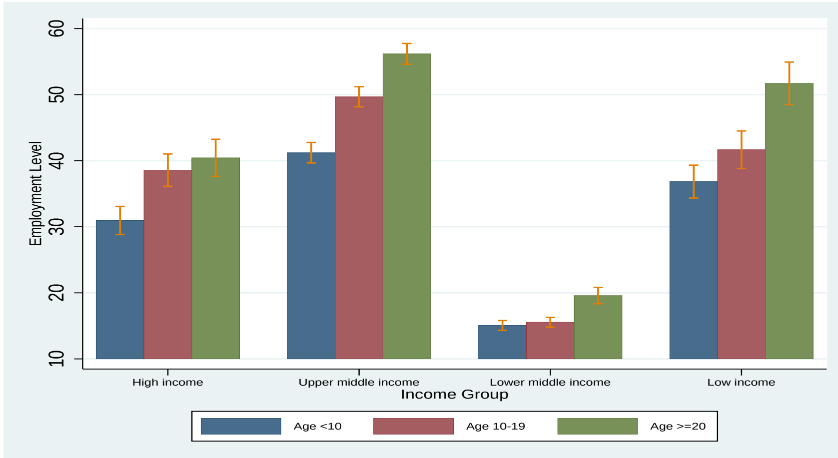

Figure A1: Employment distribution by income group

Source: Enterprise Surveys, 23 October 2014 and World Bank, 2014

Note: To calculate the number of employees at the establishments, I include both full-time permanent and temporary or seasonal employees. I consider a full-time permanent employee as equal to two full-time seasonal or temporary employees. For income groups, I use the World Bank's income classifications set on 1 July 2014.

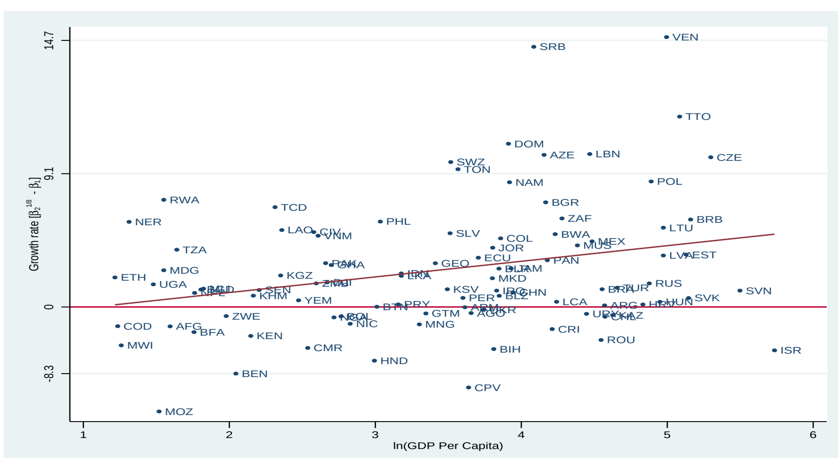

Figure A2: Correlation between the coefficient of age cohort 2 and GDP per capita

Source: Enterprise Surveys, 23 October 2014 and World Development Indicators, 2014

Note: Since the coefficient of age cohort 1 is one, I add a reference line at zero to depict that the coefficients differ from age cohort 1. In addition, I add a fitted line to explain the association between the age cohort coefficients and GDP per capita.

7.2 Derivation

To express the steady state GDP as a function of the parameters (N, L, H, λ and δ) of the model, I need the closed form solution for b_t and e_t (see Table 6), which I find using the conjecture method. The following is the solution for b_t and e_t:

At $t = 1$, I use the value of b_1 from the model specification (initial condition):

$$b_1 = L \tag{7}$$

At $t = 2$, I can write b_2 form from the model specification. Using equation 7, I can derive the following after simplifying:

$$\begin{aligned} b_2 &= (1-\lambda)b_1 + \lambda H \\ &= (1-\lambda)L + \lambda H \end{aligned} \tag{8}$$

At $t = 3$, I can write b_3 form from the model specification. Using equation 8 and after simplifying, I can derive the following:

$$\begin{aligned} b_3 &= (1-\lambda)b_2 + \lambda H \\ &= (1-\lambda)\left[(1-\lambda)L + \lambda H\right] + \lambda H \\ &= (1-\lambda)^2 L + (1-\lambda)\lambda H + \lambda H \end{aligned} \tag{9}$$

At $t = 4$, I can write b_4 form from the model specification. Using equation 9, I can derive the following after simplifying:

$$\begin{aligned} b_4 &= (1-\lambda)b_3 + \lambda H \\ &= (1-\lambda)\left[(1-\lambda)^2 L + (1-\lambda)\lambda H + \lambda H\right] + \lambda H \\ &= (1-\lambda)^3 L + (1-\lambda)^2\lambda H + (1-\lambda)\lambda H + \lambda H \end{aligned} \tag{10}$$

Since the expressions for b_1, b_2, b_3, and b_4 reveal a developing pattern, I can conjecture the solution of b_t:

$$\begin{aligned} b_t &= (1-\lambda)^{t-1}L + (1-\lambda)^{t-2}\lambda H + \cdots + (1-\lambda)^2\lambda H + (1-\lambda)\lambda H + \lambda H \\ &= (1-\lambda)^{t-1}L + \lambda H\left[(1-\lambda)^{t-2} + \cdots + (1-\lambda)^2 + (1-\lambda) + 1\right] \\ &= (1-\lambda)^{t-1}L + \frac{1+(1-\lambda)-(1-\lambda)^{t-1}-(1-\lambda)^t}{1-(1-\lambda)^2}\lambda H \qquad (11) \end{aligned}$$

Similarly, I solve for e_t. At $t = 1$, I use the value of e_1 directly from the model specification (initial condition at $t = 1$):

$$e_1 = \frac{H+L}{2} \qquad (12)$$

At $t = 2$, I can write e_2 from the model specification. Using equations 12 and 7, I can derive the following after simplifying:

$$\begin{aligned} e_2 &= e_1 + \frac{\lambda(H-b_1)}{2} \\ &= \frac{H+L}{2} + \frac{\lambda(H-L)}{2} \qquad (13) \end{aligned}$$

At $t = 3$, I can write e_3 form from the model specification. Using equations 13 and 8, and after simplifying, I can derive the following:

$$\begin{aligned} e_3 &= e_2 + \frac{\lambda(H-b_2)}{2} \\ &= \frac{H+L}{2} + \frac{\lambda(H-L)}{2} + \frac{\lambda(H-(1-\lambda)L-\lambda H)}{2} \\ &= \frac{H+L}{2} + \frac{\lambda(H-L)}{2}\left[1+(1-\lambda)\right] \qquad (14) \end{aligned}$$

At $t = 4$, I can write e_4 from the model specification. Using equations 14 and 9, I can derive the following after simplifying:

$$
\begin{aligned}
e_4 &= e_3 + \frac{\lambda(H - b_3)}{2} \\
&= \frac{H+L}{2} + \frac{\lambda(H-L)}{2}\left[1 + (1-\lambda)\right] + \frac{\lambda(H - (1-\lambda)^2 L - (1-\lambda)\lambda H - \lambda H)}{2} \\
&= \frac{H+L}{2} + \frac{\lambda(H-L)}{2}\left[1 + (1-\lambda) + (1-\lambda)^2\right] \qquad (15)
\end{aligned}
$$

Since the expressions of e_1, e_2, e_3, and e_4 follow a developing pattern, I can conjecture the solution of e_t:

$$
\begin{aligned}
e_t &= \frac{H+L}{2} + \frac{\lambda(H-L)}{2}\left[1 + (1-\lambda) + (1-\lambda)^2 + \cdots + (1-\lambda)^{t-2}\right] \\
&= \frac{H+L}{2} + \frac{\lambda(H-L)}{2}\left[\frac{1 + (1-\lambda) - (1-\lambda)^{t-1} - (1-\lambda)^t}{1 - (1-\lambda)^2}\right] \qquad (16)
\end{aligned}
$$

In period t, this cohort's contribution to GDP is $n_t s_t e_t$. The economy in steady state is composed of an infinite number of cohorts: the one that just entered, the one that entered 1 period before, the one that entered 2 periods before, and so on. That being the case, the steady state GDP is simply the sum from 1 to ∞ of $n_t s_t e_t$:

$$
\begin{aligned}
GDP &= \sum_{t=1}^{\infty} n_t s_t e_t \\
&= n_1 s_1 e_1 + n_2 s_2 e_2 + n_3 s_3 e_3 + n_4 s_4 e_4 + \cdots \\
&= N1\frac{H+L}{2} + N(1-\lambda)\left(\frac{1-\lambda}{1-\lambda+\delta\lambda}\right)\left(\frac{H+L}{2} + \frac{\lambda(H-L)}{2}\right) \\
&+ N(1-\lambda)^2\left(\frac{1-\lambda}{1-\lambda+\delta\lambda}\right)^2\left(\frac{H+L}{2} + \frac{\lambda(H-L)}{2}\left[1+(1-\lambda)\right]\right) \\
&+ N(1-\lambda)^3\left(\frac{1-\lambda}{1-\lambda+\delta\lambda}\right)^3\left(\frac{H+L}{2} + \frac{\lambda(H-L)}{2}\left[1+(1-\lambda)+(1-\lambda)^2\right]\right) + \cdots \\
&= \frac{N(H+L)}{2}\left[1 + a + a^2 + a^3 + \cdots\right] \\
&+ \frac{\lambda N(H-L)}{2}\left[a(1 + a(1+b) + a^2(1+b+b^2) + \cdots)\right] \qquad (17)
\end{aligned}
$$

where, $a = \frac{(1-\lambda)^2}{1-\lambda+\delta\lambda}$ and $b = 1-\lambda$. To find the solution, I use the geometric series.[8] The solution for steady state GDP as follows:

$$GDP = \frac{N(H+L)}{2}\frac{1}{1-a} + \frac{\lambda N(H-L)}{2}\frac{a}{(1-a)(1-ab)} \tag{18}$$

8

$$1 + a + a^2 + \cdots = \sum_{n=0}^{\infty} a^n = \frac{1}{1-a}, \quad |a| < 1$$

$$\begin{aligned} 1 + a(1+b) + a^2(1+b+b^2) + \cdots &= \sum_{n=0}^{\infty} a^n \sum_{m=0}^{n} b^m \\ &= \sum_{n=0}^{\infty} a^n \frac{1-b^{n+1}}{1-b} \\ &= \frac{1}{1-b}\left(\sum_{n=0}^{\infty} a^n - b\sum_{n=0}^{\infty}(ab)^n\right) \\ &= \frac{1}{1-b}\left(\frac{1}{1-a} - \frac{b}{1-ab}\right), \quad |a| < 1 \;\; \& \;\; |ab| < 1 \\ &= \frac{1}{(1-a)(1-ab)} \end{aligned}$$