On the Relationship between Quality Assurance and Productivity in Software Companies

Carlos Henrique C. Duarte*
BNDES, Av. Chile 100, Rio de Janeiro, RJ, 20001-970, Brazil

ABSTRACT

Quality assurance methods based on software process improvement models have been regarded as a main source of variability in software productivity. In this paper, we investigate the relationship between labor productivity and quality assurance levels, using a data set containing more than 500 Brazilian software firms. We perform statistical analyses relating labor productivity, as measured through the annual gross revenue per worker ratio, to quality levels, whose maturity was examined in appraisals performed from 2006 to 2012 according to two distinct software process improvement models (CMMI and MPS.BR). As a preparatory step to our findings, we investigate the relationship between these models. We show that CMMI and MPS.BR appraised maturity levels are correlated, but we find no statistical evidence that the implemented quality assurance methods are related to higher labor productivity or productivity growth.

Categories and Subject Descriptors

D.2.9 [**Software Engineering**]: Management—*Productivity, Software Process Models, Software Quality Assurance.*

General Terms

Economics.

Keywords

CMMI, MPS.BR, Software Engineering Economics.

1. INTRODUCTION

Quality assurance methods based on the implementation of software process improvement models have always been regarded by the Software Engineering community as one of the main sources of (negative and positive) variability in software productivity. Many argue that such models have an impending cost [21], whereas many others identify in their use a way of complying with software development patterns and standards [15], producing economic value [12] and guaranteeing corporate performance improvement [11].

The concerns with software productivity seem to be reasonable, since more and more software systems are demanded, but the resources available for their production cannot scale quickly. For instance, if salaries could rise following demand increases, a shortage of qualified human resources would eventually appear. On the other hand, productivity may be positively influenced by disruptive software development methodologies, such as the use of lean methods or automated development tools. There is an economic interest in software productivity that justifies its measurement, which in turn allows us to draw comparisons and propose public policies or programmes aimed at the software industry.

The measurement of software productivity is a recurrent theme in the literature, but it is *per se* not easy to perform. Boehm [2], in particular, defines software productivity as the ratio between the outputs and the inputs of a process: inputs comprise labor, equipment and third party components, while outputs consist in source code, specifications or other software artifacts. It appears to be a consensus that such productivity components can be computed estimating the corresponding monetary values and that labor has the greater cost of all. In this way, it is possible to address software productivity using an economic approach, as proposed in the present paper.

An obvious way to attempt to deal with the cost of labor issue and improve software productivity is to adopt quality assurance methods ensuring that process outputs have higher perceived values and/or that required amounts of labor are reduced. There are many such methods, but we consider particularly appealing the adoption of process improvement models that allow software companies to organize their software process in structured and guided ways. Two such models are the Capability Maturity Model Integration (CMMI), developed by the Software Engineering Institute [18], and the MPS.BR, an acronym of Melhoria do Processo de Software Brasileiro (in portuguese), a joint effort of Brazilian industry, government and research institutions coordinated by the Softex Society [16]. Both were originally proposed to establish quality standards for software development and have helped hundreds of companies to implement the respective reference models and improve the corresponding software processes with external feedback provided by appraisal assessors.

In this paper, we argue that there is a need to define context and focus in studies relating quality assurance to software productivity. We address this subject focusing

*mail: carlos.duarte@computer.org, cduarte@bndes.gov.br; web: http://chcduarte.webs.com.

CESI '14. June 2, 2014. Hyderabad, India.

on software companies and their environment, where they coexist with other stakeholders, such as customers, employers, shareholders, standardization institutions and others, since these appear to be the actors of economic interest in this context. We perform observational studies using a data set currently containing revenues, employment and appraisals of more than 500 Brazilian software firms, which has been collected and analyzed by the author in qualitative [5, 6] and quantitative [8, 7] studies since the 1990s. This time, we adopt statistical methods to draw our conclusions.

We believe that the relationship between adopting quality assurance methods based on software process improvement models and obtaining higher levels of productivity in software companies has not been sufficiently studied. In many cases, the benefits and prejudices of adopting one such a method seem to be overstated, as well as the advantages and disadvantages of using a specific model in relation to others. Therefore, in this paper, we test the following hypotheses:

(HYP1) CMMI and MPS.BR appraised maturity levels are correlated;

(HYP2) On average, labor productivity in companies with appraised quality levels is higher than in other companies;

(HYP3) On average, labor productivity growth in companies with appraised quality levels is higher than in other companies.

We are not aware of other studies performed with similar context, focus and methodology. Krishnan and others [15] recognize the importance of establishing a relationship between quality and productivity in the development of software products. They study this subject using statistical methods with a focus on product quality and labor productivity, formulating distinct functions for each of these notions in terms of independent variables such as program sizes and personnel capability. In general, the literature on Software Engineering Economics treats quality and productivity matters mixed with more technical subjects, such as software artifact metrics, artifact and process complexity and management [1, 12]. We believe, however, that the corporate outwards view adopted here provides us with a simpler framework for studying the relationship between quality and productivity in software companies.

The remainder of the paper is organized as follows: Section 2 describes our research universe, which consists in software company productivity measures (2.1) and software process improvement methods (2.2); Section 3 presents our data sets and research methodology; Section 4 contains our data analyses and major findings concerning the correlation between CMMI and MPS.BR appraised maturity levels and the relationship of quality assurance to labor productivity in software companies; and Section 5 presents a systematic assessment of these results. We conclude the paper commenting on our results and suggesting further research.

2. RESEARCH UNIVERSE

2.1 Software Company Productivity

When productivity measurement is a subject of study in Economy, production or productivity functions are formulated explicitly relating, in a time relying manner, dependent variables (called outputs) to independent ones (called inputs). They can be applied to distinct focus objects, such as individuals, business units, companies, economic sectors and even national accounts. In this way, such functions are considered adequate for measuring individual effectivity and assessing comparative performance.

Griliches [10] proposes an elaborated formulation of production functions based on the classical multiplicative model of time series, in which outputs are presented as dependent variables of capital and labor inputs, apart from other more intangible factors, such as technology changes and accumulated investments in R&D. Even the independent variables that reflect tangible resources, such as investments in capital goods, are considered difficult to treat, due to the effect of economic processes such as depreciation and inflation happening in each analyzed time period. On the other hand, variables that capture intangible aspects help explain why sometimes increases in capital and labor inputs are not reflected in productivity growth, for instance due to technological lags or debts. The recognition of the importance of such intangible factors further motivates our study on the relationship between quality assurance and productivity.

The tradition in Software Engineering Economics, in turn, is to present productivity functions as follows [2]:

$$productivity_o(t) = \frac{output_o(t)}{input_o(t)}$$

We have added above references to time frames t and focus objects o inspired by the economic formulation of productivity, since these are needed in dealing with indexed families of time series.

Since the cost of labor is paramount in research and development intensive companies, the typical software productivity functions adopt the cost of labor, worked hours or employed personnel as inputs and the number of produced lines of code, components or function points as outputs. As we have already argued, the mixture of such economic and technical independent variables seems to be counter-intuitive, since they belong to different contexts, hindering our efforts to characterize productivity as a dependent variable, in the way suggested in [9].

At this point, it is clear that we wish to mix the two aforementioned approaches in studying the productivity of software companies. That is, we wish to perform contextualized temporal analyses of productivity, based on non technical variables, which are intuitive to policy makers and the whole software industry, not only due to their familiarity with the adopted independent variables but also because of a simple productivity function formulation. Moreover, we prefer to adopt variables for which publicly available observations are at hand, since these may be subject to third party validation. Finally, due to the comparative nature of our study, between individuals or groups of companies in specific time periods, we do not need to care about currency, inflation, depreciation nor with any other financial aspect that equally affects the performance of the whole population or sample under investigation. All these requirements guide us in the choice of independent variables and in the formulation of a specific productivity function.

Here we propose the adoption of the total number of company workers as the measure of input (involving trainees, employers and possibly some shareholders, among others). This is an objective metric that allows us to compare the production capability of distinct software companies. As a measure of output, we propose the use of annual gross revenues. In this way, we adopt in the present paper the gross revenue per worker ratio as a measure of productivity. This is usually called labor productivity in the literature.

Any difficulty that may arise from using the gross revenue per worker ratio can be overcome by the specific data collection, classification and filtering methodology described in Section 3. To begin with, revenues are considered in analyses only with third party validation and companies are stratified into segments taking into account their main source of revenue, avoiding that firms with different businesses be directly compared. On the other hand, companies of the same segment that maintain different cost structures, such as software service companies that may or may not have highly outsourced operations (as studied in [19]) can be directly compared, but this appears to be fair since the effectiveness of the outsourcing structure is captured by the productivity measure. For a detailed discussion on the advantages and disadvantages of adopting specific productivity functions, the reader is referred to the extensive study of the OECD on this subject reported in [17].

2.2 Software Process Improvement Models

Quality assurance methods based on the implementation of software process improvement models aim to lower the number of potential defects in software artifacts; to improve the proposition and execution of tasks in a timely fashion and on budget; to provide high end user satisfaction and good software warranty, among others. In order to achieve these goals, such models are provided with guidelines that facilitate their implementation, external appraisal and maintenance.

The CMMI model is provided with guidelines that frame software process maturity in five levels, each one defined in terms of key process areas representing issues that a company or business unit should address so as to be considered mature in that respect [18]. The first level, denoted by the number 1 in the proper number scale, corresponds to incipient maturity, found in entities with *ad hoc* development processes. The second level corresponds to a situation in which processes are managed and repeatable. The third level is related to the formal definition of software processes and the use of organizational learning in process improvement. The next level focuses on software process control and monitoring. The final level requires the use of quantitative data for guaranteeing continuous process improvement.

The MPS.BR model is more fine grain than the CMMI model and suggests that software companies begin their software process improvement attempts earlier in time, due to cost and risk aspects [16]. After the incipient maturity level (undefined in the respective letter scale), a partially managed level is defined (denoted by letter G), in which the implementation of requirement and project management attributes is expected. Measurement, quality assurance, portfolio, configuration and acquisition management are attributes expected just in the next level (F). The attributes of defined software process are gradually required as a company or business unit progresses from level E to C. Levels B and A correspond almost precisely to the CMMI 4 and 5 levels respectively.

Entities that comply with these models are required to perform assessments to verify their maturity in conducting software processes. Guidelines are provided to serve not only as references to evaluations but also as checklists for process maturity attribute maintenance. The managing institution responsible for each model maintains a database of accredited

Table 1: Indexes for maturity level normalization.

INDEX	1	2	3	4	5	6	7
MPS.BR Level	G	F	E	D	C	B	A
CMMI Level		2		3		4	5

institutions which are allowed to perform evaluations in strict accordance to model guidelines. In the end of an assessment or appraisal, the evaluated entity receives access to the produced assessment records and public recognition that it complies with a certain maturity level.

Both models have had their definitions improved and scope expanded over the years. Today, there are in the CMMI specific maturity models for software development processes, software acquisition and service provision [14], while the MPS.BR distinguishes just service from process models. Any company may implement more than one specific model.

Although the relationship between CMMI and MPS.BR is a subject of study in this paper, we find it convenient to present at this point the adopted assignment of index numbers to maturity levels, so that they could be used in the normalization of data sets. We present in Table 1 the adopted indexes and stress that the association of a CMMI level to the same index of a MPS.BR level does not necessarily mean that such levels have a direct and strict equivalence, since the indexes are used here just as a discrete representation of such categorical data.

3. RESEARCH DATA & METHODOLOGY

We use in our research a data set containing the revenues and employment of more than 1.100 Brazilian information and communication technology companies, which have been collected and analyzed by the author since the 1990s. For each company, a time series was created accounting for observations of each of these variables in time.

We have added two other families of time series to our original data set, respectively containing data on external appraisals and maturity levels of Brazilian companies according to the CMMI [4] and the MPS.BR [20] models. There are companies in these families with diverse natures, such as military and public service organizations, as well as private companies that develop software in-house with diverse businesses, such as construction, engineering, electronics, automotive, finance, communication and healthcare, apart from software companies. The statistical profile of this specific data set is presented in Table 4. Later on, we filter the resulting composed data set to focus just on software companies, producing an outline of the quality and productivity levels maintained by these companies in Brazil.

As a selection criterion for inclusion of an observation in a data set, we reject observations without external validation. That is why we only take into account results of external appraisals which were performed by accredited assessors and reported to the managing institutions of the CMMI and the MPS.BR models. Concerning corporate data, large size companies listed in the Bovespa Stock Market upload their financial statements into the publically accessible CVM[1] site, whereas most medium size companies in Brazil are obliged to publish their financial statements in large circulation

[1] The Brazilian Securities and Exchanges Commission.

newspapers. Many others share these data, as well as their employment numbers, with trusted research institutes, which perform varying levels of cross checking. We use only this kind of observation to populate our corporate data set.

Having organized our research data, we adopt the following research methodology:

1. Data adjustment;
2. Data filtering;
3. Derived data computation;
4. Statistical analysis;

There is a lot of adjustment required by our raw collected data. The whole data set is organized according to corporate tax payer registry unique number (CNPJ), company name, main business code (according to the Brazilian national system of economic activity classification – CNAE) and main origin of invested capital (local or foreign). In many cases, these data are clearly incorrect and have to be adjusted manually. We also transform all categorical data on maturity levels into discrete numbers, using the indexes in Table 1, with some automated support provided by spreadsheets.

Next, we join our data set on economic observations to our maturity level data set based on the unique registry number of each company. In doing so, some inconsistencies appear, for instance between the CNAE code of a company and its main source of revenue, requiring an additional data adjustment iteration. We thus begin our filtering process, since we are only interested in studying the relationship of productivity to quality levels in software companies. The business classification code of each company is used to create a category of software product companies, which develop packaged or customized software, and another one for software service companies, which is quite broad, since it encompasses from single software development services and consulting provision to full business process outsourcing activities. It is important to mention that the use of this aggregation criterion based on the adjusted CNAE code of each company results in analyses not directly comparable to our previous work [8, 7].

We are forced to adopt a temporal filter in treating our data. Although our time series on economic data begin in 1990, we choose to conduct our analyses just in the period from 2006 to 2012, not only due to the increase in the volume of data on appraisals since the beginning of this period but also due to comparability reasons, since the Software Engineering Institute performed a major change in the CMM implementation in 2006 to require from this year onwards that appraisals be performed just according to the CMM Integration model [3]. As mentioned earlier, nowadays both the CMMI and the MPS.BR adopt specific models for software development and service provision [14], but to simplify our analyses, we only take the former into account here.

Finally, we perform derived data computation based on our filtered data. Since our historical economic data is sparse, in the sense that there are missing observations in the middle of the period, we use, as in [10], interpolation to estimate interior points in the growth curve of revenues and employment of each company, based on the points already present in each time series. The statistical profile of our software company data set is presented in Table 5. We also compute the productivity ratio of each company for all those years in which there are original or computed simultaneous observations of revenues and employment. The statistical profile of our productivity data set appears in Table 6.

It is important to mention that, in subsequent statistical analyses correlating productivity to quality in software companies, we adopt the productivity figures of the whole company or its economic group whenever there is no economic data available concerning the entity that was evaluated in appraisals, provided that they all have their main business based on software.

4. DATA ANALYSIS & FINDINGS

We have studied two distinct software process improvement models not only due to our interest in obtaining a data set with a higher number of observations but also as a research strategy to perform independent statistical tests relating labor productivity to quality levels appraised according to each of these models. If we confirm the same hypothesis testing it independently using each model data, that would provide stronger evidence of the robustness of our conclusions. We thus investigate the relationship between the CMMI and the MPS.BR models in Section 4.1 and the the relationship of quality levels to labor productivity in Section 4.2.

4.1 CMMI / MPS.BR Relationship

In Table 4, we present the population of CMMI and MPS.BR appraisals performed in Brazil, as well as the statistical profile of our maturity level data set. Our hypothesis concerning the relationship between these models is:

(HYP1) CMMI and MPS.BR appraised maturity levels are correlated;

Our null hypothesis is that there is no correlation between the populations of maturity levels appraised according to these models (population correlation coefficient equal to to zero). In order to test these hypotheses, since the respective populations are not normally distributed having positive skews, we compute the respective Spearman correlation coefficient. For the 78 concurrent appraised maturity level observations in our sample, the computed coefficient is 0.3551, which shows a positive sample correlation. We use the Fisher transformation to compute the associated z-score, which is 3.1236. Considering that each pair of maturity levels corresponds to two independent observations, this kind of score approximately follows a normal distribution. Therefore, the probability that corresponds to the computed z-score is 0.9991, yielding a p-value equal to 0.0009, which is less than 0.0250, the level of significance adopted in our research. This allows us to reject the null hypothesis and confirm **HYP1**.

It is interesting to analyze this correlation in more detail. If we take into account just software product companies, the computed Spearman correlation coefficient is 0.7518, showing that there are outliers in our original sample. Coincidentally, they all correspond to software service companies. The difference between the correlation coefficients of software product and service companies suggests that we should deal with them separately in our subsequent statistical analyses.

A graphical representation of the correlation between the CMMI and MPS.BR appraised maturity levels in our sample appears in the scatter plot of Figure 1. Therein, an appraised CMMI level determines the vertical coordinate of a point, while the corresponding appraised MPS.BR level determines its horizontal coordinate. We present the number of occurrences of coincident observations behind each point. Note that outliers in our sample correspond precisely to appraisals that generate points far from the diagonal of our plot.

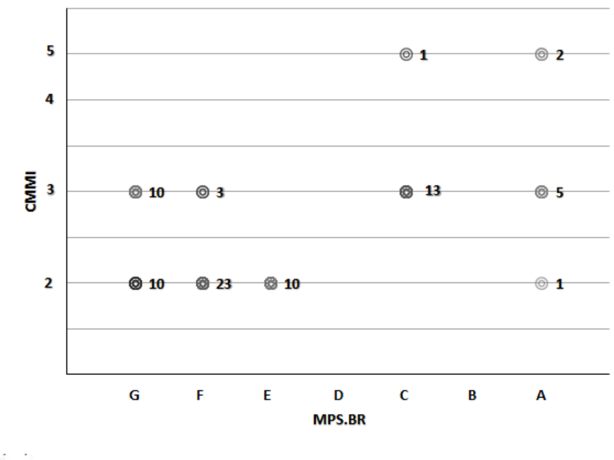

Figure 1: Scatter plot of the concurrent CMMI and MPS.BR appraised maturity levels in our sample.

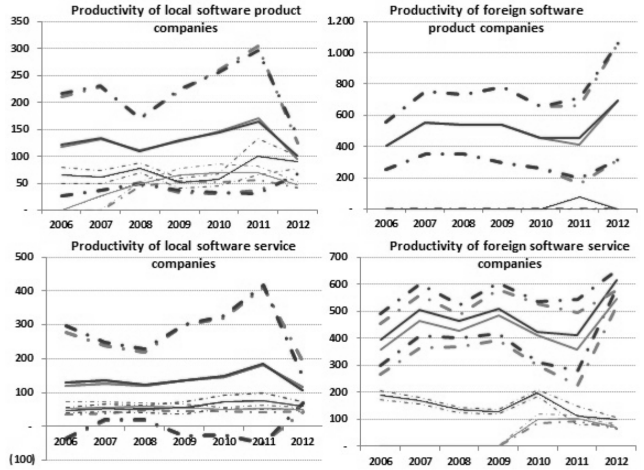

Figure 2: Control graphs of software company average productivities in our samples.

Figure 1 results from the association of maturity levels to indexes presented in Table 1, which was defined taking into account the theoretical definition of both models [16]. Using a least squares technique to minimize the indirect distances between the CMMI level and the MPS.BR level attributed at the same time to each entity, there are other possible index associations that best describe our sample. Such alternatives differ from the association presented in Table 1 due to the assignment of some CMMI levels to lower indexes than those in the table, suggesting that MPS.BR assessors have required more from software companies than prescribed by the theoretical definition of the model. Even when we adopt any other of these possible associations, **HYP1** still holds.

4.2 Quality / Productivity Relationship

Now we study the relation of appraised quality levels to labor productivity considering the population of Brazilian software companies. In Table 6, we note a growing average productivity trend among the companies in our data set. Investigating the reasons for such productivity gains, we realize that that foreign capital companies keep operations in Brazil exploring businesses with higher labor productivities than those of local companies. For software product companies, this is due to the commercial nature of their operations: many such companies just sell in the country software developed abroad, having small, if any, local development costs. For software service companies, this is due to their choice to provide in the country value added services with high margins. These facts point out that revenue and employment in foreign owned companies are different from those of local companies, leading us to further stratify our analyses according to the origin of company capital.

Consequently, we test the following hypothesis in each of the four partitions of our data set, that is, considering the business nature and the capital origin of each company, both for CMMI and MPS.BR appraised quality levels, generating eight statistical test results:

(HYP2) On average, labor productivity in companies with appraised quality levels is higher than in other companies;

Since our productivity data set is positively skewed, we apply a logarithmic transformation to generate an approximate normal distribution. Next, we use Welch's t-test for two independent samples with unequal sizes and variances to attempt to validate **HYP2**. The t-test suggests as a null hypothesis that the difference between the average values of the respective populations is equal to zero. In the context of our analyses, this means that the average labor productivity in companies with appraised quality levels is equal to that of other companies. The test is computed taking into account the sizes, average values and variances of each sample partition (those containing the productivity of companies with and without appraised quality levels), in the same way that the required varying numbers of degrees of freedom (DoF) are computed. We perform such test on an annual basis, but present in Table 2 the computed results for the whole period, since they are similar to the annual results. As it can be seen, the negative t-test result values in the table do not allow us to discard the null hypothesis.

We provide an alternative analysis of this result through the control graph in Figure 2. Therein, the average productivity figures in each sample partition are presented using continuous lines, surrounded by dashed lines corresponding to the same data plus and minus 0.5 standard deviations. We use light grey lines to make reference to companies with MPS.BR appraised levels and dark grey ones to denote companies with CMMI appraisals, whereas line thickness denotes a subject of a statistical analysis (thin lines denote the average productivity of companies in the analysed sample partition, in opposition to thick ones, which represent those of remaining companies). Since each thin line in a graph appears under the corresponding thick line of the same color, our graphs validate the corresponding statistical test results.

The adequacy of **HYP2** can be questioned, since the returns on investments in quality assurance based on the implementation of software process improvement models tend to appear and cease as time passes. Indeed, the CMMI Institute and the Softex Society establish the validity period of appraisals in three years [4, 20] and this obsolescence time was reflected in the computation of the valid maturity levels

Table 2: Results of Welch's t-test on the average productivity of companies in our samples.

NATURE OF BUSINESS	ORIGIN OF CAPITAL	MPS.BR			CMMI		
		N	t-test	DoF	N	t-test	DoF
Software products	Local	38+340	-2,9566	356	31+347	-2,3824	110
	Foreign	0+67	N.A.	N.A.	1+66	N.A.	N.A.
Software services	Local	67+741	-6,5356	570	119+689	-4,6482	691
	Foreign	6+217	-4,1315	221	35+186	-1,7806	218

reported in Table 4. We therefore formulate and test an alternative hypothesis concerning productivity growth:

(HYP3) On average, labor productivity growth in companies with appraised quality levels is higher than in other companies.

Again we use Welch's t-test on our log-transformed productivity data set, this time to attempt to validate **HYP3**. As our null hypothesis, we conjecture that the average labor productivity growth in companies with appraised quality levels is equal to that of companies without appraisals. The resulting sample sizes, t-test result values and required degrees of freedom are presented in Table 3.

The results in Table 3 are harder to interpret. Unfortunately, in this case, we face a drastic reduction in sample sizes, since each growth rate is computed considering the whole period of study. Due to this, it was impossible to perform tests concerning foreign capital software product companies, as it was also the case concerning the data presented in Table 2, since there is not even a pair of such companies simultaneously with computed productivity ratio and appraisals performed in Brazil in the studied period, for some sample partitions. Regarding the test results on locally owned software product companies, they would allow us to discard the null hypothesis for the CMMI case, however, when we go back to our original sample data (illustrated by Figure 2), we notice that the average productivity growth of companies with appraised CMMI quality levels is smaller than those outside this partition, concluding that the test result was biased by the logarithmic transformation. Regarding software service companies, although some sample partition sizes are rather small (containing 2 and 9 observations only), while others are populated with companies whose appraisals were regarded as outliers in Section 4.1, the negative test results suggest that we cannot discard the null hypothesis.

It is important to point out that the negative t-test results obtained in most of our analyses would allow one to confirm the opposite of **HYP2** and **HYP3**, that is, average productivity and productivity growth in software companies are negatively related to appraised quality levels. We believe that this conclusion requires further confirmation, perhaps by using refined non-parametric statistical methods.

5. RESEARCH RESULT ASSESSMENT

Some results presented in this paper appear to be quite strong, since they suggest that our common intuition is not universally valid concerning a positive relationship between labor productivity and quality assurance methods based on the implementation of software process improvement models. This requires a careful assessment, as presented below.

Our research was performed with real data obtained from trusted sources in all cases. As a methodological principle, we did not accept without cross-checking any data provided directly by the studied companies. The main sources of data were published financial statements, published market research studies and reports of the managing institutions of the CMMI and the MPS.BR models. Therefore, we are confident in the high data quality of our research.

It should be recognized that, although our productivity data set contains the productivity figures of companies of different sectors, sizes and businesses, data samples defined out of it may not be randomly selected, since the data set itself mostly reflects the productivity of mid-size and large companies, for which public data is available. Although this limits the possibility of generalization of our findings, we believe that the observed negative relationship between quality levels and productivity would be worsened if more small size companies were taken into account.

Unfortunately the author only had access to research data concerning the Brazilian software industry. This alerts us that the obtained results may only be regionally valid. In spite of this, to attempt to dissociate the these results from the local reality, two distinct software process improvement models were investigated. The CMMI model has worldwide adoption, while the MPS.BR model begins to be adopted in other countries of South America. The fact that each performed pair of statistical tests agrees on the obtained individual results, despite the fact that each of them was produced based on a data set partition reflecting the existence of either CMMI or MPS.BR appraisals, provides evidence that our results may not just be regionally valid.

The temporal stability of our results deserves further investigation. Herbsleb and Goldenson [11] reported in 1996 the existence of a positive relationship between quality assurance and software process productivity due to the early adoption of the CMMI model, while Kalinowski and others [13] suggested in 2011, based on extensive data collected from companies that implemented the MPS.BR model, that the studied time frame matters when software productivity is related to process maturity. Although both studies address software process productivity but not software company productivity, we believe that it is worthwhile investigating this kind of relationship within longer time frames.

6. CONCLUDING REMARKS

In this paper, we investigated the relationship between labor productivity and quality levels, whose maturity was examined in appraisals performed according to two distinct models (CMMI and MPS.BR). We performed statistical analyses showing that CMMI and MPS.BR appraised maturity levels are correlated, but we could not find any statistical evidence that implemented quality assurance methods are related to higher levels of productivity or productivity growth.

We believe that it is worthwhile to continue the reported re-

Table 3: Results of Welch's t-test on the average productivity growth of companies in our samples.

NATURE OF BUSINESS	ORIGIN OF CAPITAL	MPS.BR			CMMI		
		N	t-test	DoF	N	t-test	DoF
Software products	Local	10+65	-13,8810	59	8+68	5,5518	11
	Foreign	0+12	N.A.	N.A.	0+12	N.A.	N.A.
Software services	Local	18+142	-27,6130	99	30+138	-22,1300	149
	Foreign	2+43	-25,8770	42	9+39	-17,3530	46

search performing additional statistical analyses using larger data sets, populated with more economic and software quality data, as well as within longer time frames, so as to obtain, if possible, further confirmation of our conclusions. It also appears to be worthwhile using refined statistical methods in these analyses, such as non-parametric ANOVA, in order to avoid the biases of log-transformations and, more importantly, to determine data partitions according to the criteria that better explain the respective productivity variations.

Another subject that deserves further investigation is the relationship between software company productivity and software process productivity. As recognized by Boehm in [2], the implementation of quality assurance methods may have costs that counter the obtained productivity gains. Our research suggests that understanding and improving software productivity, be it at the corporate or at the software process level, is still an important goal.

7. ACKNOWLEDGMENTS

The author wishes to thank the Softex Society administration for providing historical data concerning MPS.BR appraisals and both to Luiz Paulo Alves Franca and the anonymous reviewers for their comments on earlier versions of this paper.

8. REFERENCES

[1] Barry W. Boehm. *Software Engineering Economics.* Prentice Hall, 1981.

[2] Barry W. Boehm. Improving software productivity. *Computer*, 20(9):43–57, 1987.

[3] Mary B. Chrissis, Mike Konrad, and Sandy Shrum. *CMMI: Guidelines for Process Integration and Product Improvement.* SEI Series in Software Engineering. Addison-Wesley, 2006.

[4] CMMI Institute. Published appraisal results [Online], 2013. Available: `https://sas.cmmiinstitute.com/pars/pars.aspx`.

[5] Carlos Henrique C. Duarte. Moving software to a global platform. *IEEE Spectrum*, 33(7):40–43, July 1996.

[6] Carlos Henrique C. Duarte. Brazil: Cooperative development of a software industry. *IEEE Software*, 19(3):84–87, 2002.

[7] Carlos Henrique C. Duarte. A decade of continued support to the information and communication technology sector in Brazil: The most relevant events and the role of BNDES. *Revista do BNDES*, 19(37):91–126, 2012. In Portuguese.

[8] Carlos Henrique C. Duarte and Carlos Eduardo C. Branco. Social and economic impacts of the brazilian policy for information technologies. *Revista do BNDES*, 15:125–146, 2001. In Portuguese.

[9] Tony Gorschek and Alan Davis. Requirements engineering: In search of the dependent variables. *Information and Software Technology*, 50(1-2):67–75, January 2008.

[10] Zvi Griliches. Productivity, R&D and basic research at the firm level in the 1970s. *The American Economic Review*, 76(1):141–154, March 1986.

[11] James D. Herbsleb and Dennis R. Goldenson. A systematic survey of CMM experience and results. In *Proc. 18th International Conference on Software engineering (ICSE'96)*, pages 323–330. IEEE Computer Society, 1996.

[12] Capers Jones and Olivier Bonsignour. *The Economics of Software Quality.* Addison-Wesley, 2012.

[13] Marcos Kalinowski et al. From software engineering research to brazilian software quality improvement. In *Proc. 15th Brazilian Software Engineering Symposium (SBES'2011)*, 2011.

[14] Mike Konrad and Sandy Shrum. *CMMI for development: version 1.3.* SEI series in software engineering. Addison-Wesley, 3rd edition, 2011.

[15] Mayuram S. Krishnan, Charles H. Kriebel, Sunder Kekre, and Tridas Mukhopadhyay. An empirical analysis of productivity and quality in software products. *Management Science*, 46(6):745–759, June 2000.

[16] Mariano A. Montoni, Ana R. Rocha, and Kival C. Weber. MPS.BR: a successful program for software process improvement in Brazil. *Software Process: Improvement and Practice*, 14(5):289–300, 2009.

[17] OECD. *Measuring Productivity: Measurement of Aggregate and Industry-Level Productivity Growth.* 2001.

[18] Mark C. Paulk, Charles V. Weber, Bill Curtis, and Mary B. Chrissis. *The Capability Maturity Model: Guidelines for Improving Software Processes.* Addison-Wesley, 1995.

[19] Harvey P. Siy et al. Making the software factory work: Lessons from a decade of experience. In *Proc. 7th International Symposium on Software Metrics (METRICS'01)*, pages 317–326. IEEE Computer Society, 2001.

[20] Softex Society. MPS.BR evaluations [Online], 2013. Available: `http://www.softex.br/mpsbr`.

[21] Mark Staples et al. An exploratory study of why organizatons do not adopt CMMI. *Journal of Systems and Software*, 80(6):883–895, 2007.

APPENDIX

The statistical profiles of the studied data sets appear in Tables 4, 5 and 6.

Table 4: **Statistical profile of our normalized maturity level data set.**

NUMBER OF APPRAISALS	2006	2007	2008	2009	2010	2011	2012
MPS.BR - Population	12	55	51	81	72	72	84
Sample	12	55	51	80	70	72	82
Software product companies	2	23	11	27	25	26	26
Software service companies	9	30	35	48	41	41	45
CMMI - Population	16	14	27	39	40	29	25
Sample	10	10	14	33	27	28	25
Software product companies	0	1	2	9	3	7	8
Software service companies	8	5	8	21	19	16	12
MATURITY LEVELS	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
MPS.BR - Sample							
Mean value	2.00	1.00	1.00	1.00	1.00	1.00	1.00
Standard deviation	1.97	1.30	1.16	1.13	1.09	1.29	1.31
Kurtosis	2.23	11.02	15.07	13.42	9.29	3.98	3.21
Skewness	1.73	3.23	3.73	3.45	2.87	2.12	1.94
CMMI - Sample							
Mean value	2.00	2.00	2.00	4.00	2.00	2.00	2.00
Standard deviation	1.74	1.87	1.72	1.56	1.63	1.52	1.59
Kurtosis	1.01	0.05	0.60	0.55	0.44	0.72	0.47
Skewness	1.58	1.24	1.34	1.06	1.14	1.16	1.13

Table 5: **Statistical profile of our software company data set.**

REVENUES (in US$ 1.000)	2006	2007	2008	2009	2010	2011	2012
Observations	284	286	294	275	278	243	145
Sum	19.190.351	24.992.070	24.409.512	27.635.061	32.062.515	31.532.153	29.212.361
Maximum	2.438.600	3.401.500	2.192.381	2.657.800	2.715.000	2.153.000	3.248.817
Average	67.572	87.385	83.026	100.491	115.333	129.762	201.465
Mean	15.088	18.462	17.480	20.847	23.603	31.000	39.123
Standard deviation	187.020	260.267	224.510	270.107	280.255	285.446	410.478
Minimum	141	371	699	625	341	1.324	2.217
Kurtosis	95	95	47	46	35	21	24
Skewness	8	8	6	6	5	4	4
NUMBER OF WORKERS	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Observations	218	232	235	243	215	194	95
Sum	326.520	402.559	417.699	428.049	471.524	442.115	429.177
Maximum	54.415	67.032	74.756	75.000	86.000	91.922	140.000
Average	1.498	1.735	1.777	1.762	2.193	2.279	4.518
Mean	219	227	248	250	376	363	600
Standard deviation	5.360	6.475	7.273	7.231	8.277	9.470	16.977
Minimum	5	15	5	10	22	18	20
Kurtosis	73	76	82	83	75	74	49
Skewness	8	8	9	9	8	8	7

Table 6: **Statistical profile of our software company productivity data set.**

PRODUCTIVITY	2006	2007	2008	2009	2010	2011	2012
Observations	234	238	246	237	227	197	97
Sum	37.834	44.500	45.205	44.487	41.348	37.891	19.319
Maximum	2.913	2.876	2.777	2.699	3.085	3.468	3.219
Average	162	187	184	188	182	192	199
Mean	61	74	80	74	77	83	76
Standard deviation	299	333	343	373	362	397	411
Minimum	9	2	8	9	7	12	5
Kurtosis	36	25	26	21	29	35	34
Skewness	5	4	5	4	5	5	5