21COE, University of Tokyo *MMRC Discussion Paper* No. 39

***MMRC***
*DISCUSSION PAPER SERIES*

MMRC-F-39

# Productivity Convergence at the Firm Level

Before March 31
Faculty of Economics, University of Tokyo, and
Economic and Social Research Institute, Cabinet Office,
Government of Japan
After April 8
Policy Board, Bank of Japan
Kiyohiko G. Nishimura

Faculty of Business and Commerce, Keio University
Takanobu Nakajima

International Graduate School of Social Sciences, Yokohama
National University
Kozo Kiyota

March 2005

東京大学21世紀COE（センター・オブ・エクセレンス）
ものづくり経営研究センター

# Productivity Convergence at the Firm Level*

Kiyohiko G. Nishimura†
Takanobu Nakajima‡
Kozo Kiyota§

March 12, 2005

**Abstract**

Productivity convergence among countries has been investigated extensively with mixed results. This paper extends the analysis to the firm level to shed light on the debate of convergence or non-convergence. We find productivity convergence among firms widely in Japan, in both manufacturing industries and non-manufacturing ones. We obtain these results taking explicit account of exiting firms as a source of selection biases. The convergence rate is much faster among firms than countries. We also find that there are substantial differences among industries in the convergence speed. IT industries that heavily rely on technological progress show faster rates of convergence.

**Key words**: firm-level productivity, convergence, technology diffusion, selection bias

**JEL classification code**: D21 (Firm Behavior); D24 (Production; Capital and Total Factor Productivity; Capacity); O47 (Measurement of Economic Growth; Aggregate Productivity); L11 (Production, Pricing and Market Structure; Size Distribution of Firms)

---

*Views expressed here are strictly personal and in no way represent those of Economic and Social Research Institute, University of Tokyo, Keio University, or Yokohama National University. This paper is a part of our ongoing joint research with Flora Bellone, Patrick Musso, and Michel Quéré of Institut de Droit et d'Economie de la Firme et de l'Industrie.

†Faculty of Economics, University of Tokyo and Economic and Social Research Institute, Cabinet Office, Government of Japan

‡Faculty of Business and Commerce, Keio University

§International Graduate School of Social Sciences, Yokohama National University

¶Takanobu Nakajima, Faculty of Business and Commerce, Keio University, 2-15-45 Mita Minatoku, Tokyo, 108-8345, Japan; nakajima@fbc.keio.ac.jp, +81-3-5427-1231(Tel), +81-3-5427-1578(Fax).

# 1 Introduction

"Comparisons of productivity performance across countries are central to many of the questions concerning long-run economic growth" (Bernard and Jones, 1996, p. 1216). Thus, there has been a large body of literature that investigated cross-country productivity convergence both at the country level (Dollar and Wolff, 1988; Dorwick and Nguyen, 1989; Wolff, 1991) and at the industry level (Baumol, 1986; Bernard and Jones, 1996; Pascual and Westermann, 2002). Bernard and Jones found little evidence of convergence in manufacturing.[1] Others have obtained results that support productivity convergence in countries with lower productivity at the initial period but growing rapidly in the subsequent periods.

However, it should be noted that the growth of a country results from the growth of industries, which comes from the growth of firms. Ultimately, the productivity growth of a country is attributed to the productivity growth of firms. Thus, to examine characteristics of across-country productivity convergence, it is of the utmost importance to examine characteristics of the productivity convergence among them. If productivity convergence is important across countries, it should be all the more important within a country.

In spite of this importance, there has been little empirical work at the firm level on productivity convergence whose scope is wide enough to cover a major part of an entire economy, including not only manufacturing but also non-manufacturing sectors. This wide coverage is necessary to assess *macroeconomic* effects of technology diffusion. Although there are some recent studies of productivity convergence at the firm or establishment levels, they are also narrow in their scope in that the studies only focused on manufacturing sector.[2]

This paper fills this gap between theory and empirical analysis. We examine the growth of productivity at the firm level using a large-scale data base of firms, covering not only manufacturing but also non-manufacturing industries. The data used in this paper is a micro database of firms called *Kigyou Katsudou Kihon Chousa Houkokusho* (*The Results of the Basic Survey of Japanese Business Structure and Activities*) prepared by the Research and Statistics Depart-

[1] However, their finding is controversial. Sørensen (2001) criticized in his comment that the established evidence of non-concurrence in manufacturing by Bernard and Jones heavily depended on the choice of base year. Pascual and Westermann (2002) stressed the importance of similar technologies in productivity convergence, disaggregating the industry classification in more detail.

[2] For instance, see Fukao and Kwon (2004) for Japan and Griffith, Redding and Simpson (2002) for the United Kingdom.

ment, Ministry of Economy, Trade and Industry (METI) for the period 1994-2000.

This paper is, to the best of our knowledge, the first study to analyze technology diffusion with careful consideration on the effects of entry/exit. Note that countries hardly enter/exit while firms often do. Recent studies by Aw, Chen and Roberts (2001) and Kimura and Kiyota (2004) confirmed that not only firms' own productivity growth but also entries/exits played large roles in aggregate (national-level) productivity growth. This paper shows that the analysis without taking account of the effects of exits causes a 1.5 percent point downward bias in the speed-of-convergence.

The results of this paper indicate that there exists strong evidence of productivity convergence among firms in most industries. Further, the speed of convergence of firms is faster than has been observed in previous national- or industry-level studies which indicate an average 10.3 percent annual rate.

This paper is organized into five sections as follows. The next section explains the model, data and econometric issues. Section 3 reports the estimation results of the baseline model. Section 4 discusses the results obtained in this paper, with special emphasis on industrial difference in the speed of convergence. The final section concludes the paper.

# 2 Productivity Convergence among Firms

## 2.1 Model

The simple model of productivity convergence proposed by Bernard and Jones (1996)is our initial model. This model has been extensively utilized in the studies of cross-country productivity convergence studies. It has also been employed in recent establishment- or firm-level productivity studies (Fukao and Kwon, 2004; Griffith, Redding, and Simpson, 2002). The strength of this approach is that the specification does not depend on the form of production function.

Let us denote TFP (total factor productivity) for a firm $i$ in year $t$ by $\theta_{it}$. The TFP growth is assumed to be described as:

$$\ln\theta_{it} = \gamma_i + \lambda\left\{\ln\theta_{1t-1} - \ln\theta_{it-1}\right\} + \ln\theta_{it-1} + \ln\epsilon_{it}, \tag{1}$$

where $\ln\theta_{1t-1} - \ln\theta_{it-1}$ represents a catch-up variable, which represents the distance in productivity between firm 1 with the highest productivity and firm $i$. The speed of catching-up therefore is captured by $\lambda$ while the asymptotic rate of productivity growth of firm $i$ is denoted by $\gamma_i$. Finally, $\ln \epsilon_{it}$ represents the disturbance term.[3]

This framework yields the following catch-up path of productivity:

$$\ln \hat{\theta}_{it} = (\gamma_i - \gamma_1) + (1-\lambda) \ln \hat{\theta}_{it-1} + \ln \hat{\epsilon}_{it},$$

where $\hat{\theta}_{it} = \theta_{it}/\theta_{1t}$ and $\hat{\epsilon}_{it} = \epsilon_{it}/\epsilon_{1t}$, respectively. In the long-run, the annual average TFP growth rate of firm $i$ relative to firm 1 between year 0 and year $T$ is written as

$$\frac{1}{T}(\ln \hat{\theta}_{iT} - \ln \hat{\theta}_{i0}) = -\frac{1-(1-\lambda)^T}{T} \ln \hat{\theta}_{i0} + \frac{1}{T}\sum_{\tau=1}^{T}(1-\lambda)^{T-\tau}(\gamma_i - \gamma_1 + \ln \hat{\epsilon}_{i\tau}).$$

This is the basis of familiar convergence models of long-run average growth on initial level and we specify a baseline model as follows.

$$\Delta \ln \hat{\theta}_{iT} = \frac{1}{T}(\ln \hat{\theta}_{iT} - \ln \hat{\theta}_{i0}) = \beta_0 + \beta_1 \ln \hat{\theta}_{i0} + \mu_{iT}, \tag{2}$$

where catch-up is denoted by a negative coefficient of $\beta_1 = -\{1-(1-\lambda)^T\}/T$. We assume $\mu_{iT} \sim N(0, \sigma)$.

## 2.2 Econometric Issues

The standard procedure of a cross-country productivity convergence is to estimate equation (2) and then to get the estimate of the speed-of-convergence $\lambda$ from the estimate of $\beta_1$. However, this procedure cannot be applied to cross-firm convergence directly, since we have a problem of data truncation in firm data that is not present in country data. In the case of country data, whether or not a particular country is in the data set we consider is not correlated to the country's productivity level. In contrast, whether or not a particular firm is in the data set is quite likely to be correlated to the firm's productivity level. If a particular firm's productivity goes under some threshold level, this usually implies a serious profitability problem for the firm. Then, the firm is likely to be closed down and/or bankrupt so that it drops out of the data set we consider. This

[3] There are other factors influencing productivity movement, such R&D activities and patent purchases. In a companion paper (Nishimura, Nakajima and Kiyota, 2005), we explicitly consider effects of these factors. However, for the sake of simplicity, we assume that effects of these factors are represented by the disturbance term $\ln \epsilon_{it}$.

correlation between data truncation and productivity may produce a well-known bias of sample selection if we estimate equation (2) by an ordinary least squares (OLS).[4]

Let us consider firms' entry and exit patterns between year 0 and year $T$, which are classified into three types as summarized in Table 1. Type-1 firms are survivors that continue to stay in the market between year 0 and year $T$. Type-2 firms are exiters that stay in year 0 but exit from the market before year $T$. Type-3 firms are entrants that start their business after year 0. Whether or not productivity is observed depends on the patterns of entry and exit. The entry and exit variations lead to the missing data for the dependent or independent variables. This is often called a sample selection problem.

Table 1: Entry, Exit, and Observability of Variables

| | Year 0 | Year $T$ | Independent variables | Dependent variables |
|---|---|---|---|---|
| | $\theta_{i0}$ | $\theta_{iT}$ | $x_{i0}$ | $y_{iT}$ |
| Type-1 Survivors | observable | observable | observable | observable |
| Type-2 Exiters | observable | missing | observable | missing |
| Type-3 Entrants | missing | observable | missing | missing |

The sample selection problem does not always cause a biased estimator. Let $y_{iT} = \Delta \ln \hat{\theta}_{iT}$, $\beta' = (\beta_0, \beta_1)$ and $x'_{i0} = (1, \ln \hat{\theta}_{i0})$. Hereafter, "$'$" denotes a transpose. The regression equation is rewritten as:

$$y_{iT} = x'_{i0}\beta + \mu_{iT}.$$

We first consider the firms that are observed in year 0 (Type-1 and Type-2 firms in Table 1). For these two types of firms, the independent variables $x_{i0}$ are always observable. The problem is that we cannot observe $y_{iT}$ for Type-2 firms because Type-2 firms exit from the market between year 0 and year $T$.[5]

Suppose that the firm $i$'s survival in year $T$ ($s_{iT} = 1$) depends on the productivity and other

[4]For more detail, see Heckman (1976).

[5]If firm $i$ exits from the market between year 0 and year $T$, $\hat{\theta}_{iT}$ it is not observable, and the growth rate of TFP, $\Delta \ln \hat{\theta}_{iT}$ cannot be defined. This means that $y_{iT}$ is a missing variable.

firm characteristics in year 0, $Z'_{i0} = (1, \ln\hat{\theta}_{i0}, z^1_{i0}, z^2_{i0}, \cdots)$:

$$s_{iT} = \begin{cases} 1 & \text{if } Z'_{i0}\gamma + \nu_{iT} \geq 0 \\ 0 & \text{otherwise} \end{cases}, \quad E(\nu_{iT}|Z_{i0}) = 0 \text{ and } \nu_{iT} \sim N(0,1). \tag{3}$$

Assume that $(\mu_{iT}, \nu_{iT})$ is independent of $Z_{i0}$ and bivariate normal with zero means. This in turn implies that

$$E(y_{iT}|Z_{i0}, \nu_{iT}) = x'_{i0}\beta + \rho\nu_{iT},$$

where $\rho$ is some parameter that satisfies $E(\mu_{iT}|\nu_{iT}) = \rho\nu_{iT}$.[6]

If $\rho = 0$, $E(y_{iT}|Z_{i0}, \nu_{iT}) = E(y_{iT}|Z_{i0}) = E(y_{iT}|x_{i0}) = x'_{i0}\beta$. Since $s_{iT}$ is a function of $Z_{i0}$ and $\nu_{iT}$, $E(y_{iT}|Z_{i0}, s_{iT}) = E(y_{iT}|Z_{i0}, \nu_{iT})$. Therefore, $E(y_{iT}|Z_{i0}, s_{iT}) = x'_{i0}\beta$. That is, OLS using only Type-1 firms yields a consistent estimator when $\rho = 0$.

In contrast, if $\rho \neq 0$, we have, using iterated expectations,

$$E(y_{iT}|Z_{i0}, s_{iT}) = x'_{i0}\beta + \rho E(\nu_{iT}|Z_{i0}, s_{iT}) = x'_{i0}\beta + \rho h(Z_{i0}, s_{iT}), \tag{4}$$

where $h(Z_{i0}, s_{iT}) = E(\nu_{iT}|Z_{i0}, s_{iT})$. OLS using the observed data thus produces inconsistent estimates of $\beta$.

It is, however, possible to estimate $\beta$ consistently by a maximum likelihood (ML) method. The log-likelihood function of the sample selection model for firm $i$, $\ln\mathcal{L}_i$, is

$$\ln\mathcal{L}_i = \begin{cases} \ln\Phi\{\frac{Z'_{i0}\gamma + (y_{iT} - x'_{i0}\beta)\rho/\sigma}{\sqrt{1-\rho^2}}\} - \frac{1}{2}\left(\frac{y_{iT} - x'_{i0}\beta}{\sigma}\right)^2 - \ln(\sqrt{2\pi}\sigma) & \text{if } s_{iT} = 1; \\ \ln\Phi(-Z'_{i0}\gamma) & \text{if } s_{iT} = 0, \end{cases} \tag{5}$$

where $\Phi(\cdot)$ is the standard cumulative normal distribution function. The key estimate $\beta$ is obtained from the ML estimate of the parameters of equation (5) using the data set including both surviving (Type 1) firms and exiting (Type 2) firms.

So far, we have not considered the effect of entry. However, entrants (Type-3 firms in Table 1) are not likely to have biased effects on the speed-of-convergence equation. As explained before, exiting firms may cause a bias if we estimate equation (2) by OLS using only the observed data (that is, surviving firms), because exiting firms' dropping out of the observed data set is likely to be correlated to their productivity level. In the case of entries, we do not have such

[6]The assumption that $(\mu_{iT}, \nu_{iT})$ is bivariate normal means $E(\mu_{iT}|\nu_{iT}) = \rho\nu_{iT}$ for some number $\rho_T$. For more detail, see Wooldridge (2002, p. 562).

correlation. In fact, Nishimura, Nakajima and Kiyota (forthcoming, Figure 2 and 3) find no systematic relationship between new entries and productivity levels after entry: new entrants do not always have significantly higher (or lower) productivity than existing firms.[7] If there is no systematic relationship between entry and productivity after entry, or entry is randomly determined independently of subsequent development, then the above-mentioned ML method yields a consistent estimator. More detailed explanation will be provided in the Appendix below.

## 2.3 Data

The micro database of *Kigyou Katsudou Kihon Chousa Houkokusho* (*The Results of the Basic Survey of Japanese Business Structure and Activities*) prepared by METI (1996-2002)was used for this study. This survey was first conducted in 1991, again in 1994, and annually thereafter. The main purpose of the survey is to capture statistically the overall picture of Japanese corporate firms in light of their activity diversification, globalization, and strategies on R&D and IT.

One of the strengths of this survey is its sample coverage and information reliability. The survey covers both manufacturing and non-manufacturing firms with more than 50 employees and with capital of more than 30 million yen.[8] Industry data is available at a 3-digit level.

Another strength of this survey is that it is based on *firm* rather than establishment. Japanese *Census of Manufactures* omits information about establishments *not engaged in manufacturing activities* such as headquarters, sales blanches and R&D institutions. It is clear, that the simple aggregation of manufacturing plants does not always depict the true picture of a "firm" unless it is a single-plant firm. In that sense, our data captures the activity of firms more accurately than studies which use data based only on the manufacturing plant census. A limitation of the survey is the lack of sufficiently-detailed information on financial and institutional features such as *keiretsu*NOTE: This term needs to be defined if the audience may potentially be wider than just Japan.). Similarly, information on finance, location, and intermediate inputs is not detailed enough to investigate these variables fully.

[7]However, this result may be specific to Japanese firms in our data set, and may not be generalized. Aw et al. (2001, Table 5) found that new entrants were, on average, less productive than existing firms in Taiwanese manufacturing.

[8]Some industries such as finance, insurance, and software services are not included.

From this survey, we develop a longitudinal panel data set for the years from 1994 to 2000, following procedures used by Nishimura et al. (forthcoming).[9] We dropped the firms from our sample set for which the firm-age (the year of the survey minus the year of establishment), total wages, tangible assets, value-added (sales minus purchases), and/or the number of workers are not positive and in cases with incomplete replies. However, such anomalous data are rather rare, and thus the number of firms in our final data set exceeds 22,000 each year.

## 2.4 Productivity Measurement

To make comparisons of productivity across firms and time-series, we employ the multilateral index method in computing TFP developed by Caves, Christensen and Diewert (1982) and extended by Good, Nadiri, Roeller and Sickles (1983).[10] The advantage of a multilateral index is that we do not assume *any* specific production function, which is in line with the baseline model described in Section 2.1.

This multilateral index uses a hypothetical firm that has the arithmetic mean values of log output, log input, and input cost shares over firms in each year. Each firm's output and inputs are measured relative to this hypothetical firm. The hypothetical firms are chain-linked over time. Hence, the index measures the TFP of each firm in year $t$ relative to that of the hypothetical firm in year 0 (initial year).

Specifically, the TFP index for firm $i$ in year $t$ is defined as:

$$\begin{aligned}\ln\theta_{it} \approx\ & \left(\ln Q_{it} - \overline{\ln Q_t}\right) + \sum_{\tau=1}^{t}\left(\overline{\ln Q}_\tau - \overline{\ln Q}_{\tau-1}\right) \\ & - \sum_{j=1}^{J}\frac{1}{2}\left(\alpha_{ijt} + \bar{\alpha}_{jt}\right)\left(\ln X_{ijt} - \overline{\ln X}_{jt}\right) \\ & + \sum_{\tau=1}^{t}\sum_{j=1}^{J}\frac{1}{2}\left(\bar{\alpha}_{j\tau} + \bar{\alpha}_{j\tau-1}\right)\left(\overline{\ln X}_{j\tau} - \overline{\ln X}_{j\tau-1}\right) \qquad (6)\end{aligned}$$

[9] Detailed description of our data set is found in Nishimura et al. (forthcoming) as well as that of methods used to construct TFP measures.

[10] There is an alternative method that is based on the econometric estimation of production functions, which is proposed by Olley and Pakes (1996) and extended by Levinsohn and Petrin (2003). However, this framework has to specify a production function, which is not consistent with our baseline model. Moreover, because of the limited availability of intermediate inputs, their method was not feasible. Therefore, we employ a multilateral index method in the present study.

where $\ln Q_{it}$, $\ln X_{ijt}$, and $\alpha_{ijt}$ are the log output, log input of factor $j$, and the cost share of factor $j$ for firm $i$, respectively. $\overline{\ln Q}_t$, $\overline{\ln X}_{jt}$, and $\bar{\alpha}_{jt}$ are the same variables for the hypothetical firm in year $t$ and are equal to the arithmetic mean of the corresponding variable over all firms in year $t$.

The first term of the first line indicates the deviation of the firm $i$'s output from the output of the hypothetical firm in year $t$. The second term means the cumulative change in the output of the hypothetical firm from year 0 to year $t$. The same operations are applied to each input $j$ in the second and the third lines, weighted by the average of the cost shares.

Output is defined as value-added while inputs are capital and labor. As for other additional data and their manipulation, we adopt the methodology described in Nishimura et al. (forthcoming).

# 3 Estimation Results

## 3.1 Sample Selection Biases Caused By Exits

Since we explicitly consider possible biases caused by exits, our baseline model consists of a speed-of-convergence equation and a selection one. From equation (2), the speed-of-convergence equation is specified as follows:

$$\Delta \ln \hat{\theta}_{iT} = \beta_0 + \beta_1 \ln \hat{\theta}_{i0} + \mu_{iT}, \tag{7}$$

where $\Delta \ln \hat{\theta}_{iT}$ indicates the TFP growth of firm $i$ from year 0 to year $T$, $\hat{\theta}_{i0}$ represents the initial level TFP, and $\mu_{iT}$ represents error term.[11] In the estimation, we use $\theta_{it}$ instead of $\hat{\theta}_{it}$ for simplicity. This does not cause any problems, since the log form means that the estimated $\beta_1$ using $\hat{\theta}_{it}$ is not different from $\beta_1$ using $\theta_{it}$ (it simply shifts $\beta_0$).

The selection equation captures the effects of exiting decisions of exiting firms. Dunne, Roberts and Samuelson (1989) find that plant size, age, and ownership type (single-plant firm or multi-plant firm) are statistically significant determinants of plant growth and failure. Following the findings of Dunne, Roberts, and Samuelson, we assume that exit depends on three factors: firm age ($AGE$), employment scale ($L$), and multi-plant dummy ($D^{multi}$ that takes value one if a firm has multi-plant and zero for otherwise). In addition, we assume that natural selection

[11] Although we focus on TFP, we obtained almost the same results for labor productivity.

mechanism works: firms with lower productivity exit from the market. Thus, the selection equation for the exiting firms is represented as follows:

$$s_{iT} = \begin{cases} 1 & \text{if } \gamma_0 + \gamma_1 \ln \hat{\theta}_{i0} + \gamma_2 \ln AGE_{i0} + \gamma_3 \ln L_{i0} + \gamma_4 D_{i0}^{multi} + \nu_{iT} \geq 0; \\ 0 & \text{otherwise.} \end{cases}$$

Table 2 reports the estimation results of the baseline model with and without selection equations. The results without the selection equation (Model 0) is generated by OLS while the results with the selection equation (Model 1) is generated by ML estimation using equation (5) as explained in the previous section.

To obtain the speed-of-convergence, we first estimate $\beta_1$, and then compute $\lambda$ using the relationship between $\beta_1$ and $\lambda$: $\beta_1 = -\{1 - (1-\lambda)^T\}/T$. Note that the baseline model without the selection equation uses only Type-1 (surviving) firms so that we lose 8649 observations of Type-2 (exiting) firms.

Table 2: Baseline Model and Sample Selection Bias

| | Model 0 | Model 1 |
|---|---|---|
| Selection equation | No | Yes |
| $\beta_1$ | -0.071*** | -0.080*** |
| | (0.002) | (0.002) |
| $\rho$ | — | -0.75*** |
| | | (0.013) |
| $\lambda$ | 8.8% | 10.3% |
| $N$ | 10237 | 18886 |
| LR test for $H_0 : \rho = 0$ | — | 313.39*** |

Note: Standard errors are in parenthesis. *** indicates significance level at 1%.

Two findings stand out from this table. Firstly, in both Models 0 and 1, we observe the strong evidence of productivity convergence. However, the speed-of-convergence is different between the two. The convergence speed is faster in Model 1 than in Model 0. Furthermore, the coefficient of $\rho$ is significantly negative. A log-likelihood ratio (LR) test for $H_0 : \rho = 0$ is rejected at the 1 percent level. This implies that there exists a statistically significant sample selection bias, causing about a 1.5 percentage point downward bias to $\lambda$. Thus, the results clearly show the necessity to incorporate the selection equation in the analysis of firm-level speed of productivity convergence.

Secondly, the speed of productivity convergence is significantly faster than the speed reported in the previous country-level studies. While, for instance, Dorwick and Nguyen (1989) reported that the speed-of-convergence among countries was 2.5 percent annually, the result of Model 1 shows that the speed of convergence is 10.3 percent. At first glance, this seems a very high rate, but it is not so high if one looks at its order of magnitude. Suppose that the productivity level of firm $i$ is 10 while that of the most productive firm is 100. If the speed-of-convergence is 10.3 percent (i.e., $\lambda = 0.103$), it still takes about 23 years for firm $i$ to catch up to the most productive firm.[12] Note that whether or not a firm can survive for more than 23 years is an important issue. Nishimura et al. (forthcoming, Table 3) confirmed that in Japan about half of new firms exited from the market within five years.[13]

Moreover, we should take account of the difference between firm-level analyses and country-level ones. Diffusion of technological knowledge can be much faster among firms within the same country than among different countries because of so-called ”border effects.” The recent studies on international economics emphasize the importance of ”border effects”: the trade flows within the same country is much larger than the transaction between different countries. For instance, McCallum (1995) found that trade flows among Canada's different provinces were 22 times as large as trade between the Canadian provinces and U.S. states, despite that there is virtually no trade barrier between these two countries. Engel and Rogers (1996) found that price adjustment suffered from ”border effects”, as well. They argue that national borders as well as distance was an important impediment of price diffusion between different cities located in different countries. Similar arguments can be applied to the technology diffusion. Since we focus on the diffusion within an industry in Japan, we expect that the speed of convergence is much faster than that of cross-country studies.[14]

[12] It takes more than 27 years to catch up to the most productive firm if the speed-of-convergence is estimated as 8.8 percent (Model 0).

[13] This is not a Japan specific fact. In France, about 70 percent of new firms exited from the market within 10 years. See, Bellone, Musso and Quéré (2003) for more detail.

[14] Another possibility is an effect of aggregation bias. The productivity growth of cross-country studies captures not only productivity growth itself but also macroeconomic effects such as business cycles and sectoral shifts, which might cause underestimation of the productivity growth.

## 3.2 Robustness Checks

### The Choice of the Base Year

Our results may be sensitive to the choice of the base year. For instance, in his comments on Bernard and Jones (1996), Sørensen (2001) finds that whether or not we observe convergence depends crucially on the choice of the base year. To check the sensitivity of our results to the choice of the base year, we changed the base year from 1994 to 1995, examined the productivity growth between 1994 and 2000, and re-estimated the baseline model.

Table 3: Robustness Check: Base Year and Threshold

| | Base year = 1995 | | $L \geq 55$ |
|---|---|---|---|
| | Model 0 | Model 1 | Model 1 |
| Selection equation | No | Yes | Yes |
| $\beta_1$ | -0.073*** | -0.081*** | -0.079*** |
| | (0.002) | (0.002) | (0.002) |
| $\rho$ | — | -0.713*** | -0.76*** |
| | | (0.014) | (0.013) |
| $\lambda$ | 8.7% | 9.9% | 10.2% |
| $N$ | 11449 | 19200 | 18235 |
| LR test for $H_0 : \rho = 0$ | — | 230.87*** | 321.80*** |

Note: Standard errors are in parenthesis. *** indicates significance level at 1%.

Table 3 presents the estimation resultsof changing the base year from 1994 to 1995. The results indicate that the speed-of-convergence is not very sensitive to the choice of the base year. The estimated speed-of-convergence (9.9 percent) is almost the same as the speed-of-convergence obatained in the baseline model (10.3 percent). Moreover, the coefficient of $\rho$ is significantly negative and a LR test for $H_0 : \rho = 0$ is rejected at a 1 percent level. Thus, our result is robust for the choice to base year.

Table 3 presents the estimation resultsof changing the base year from 1994 to 1995. The results indicate that the speed-of-convergence might be sensitive to the choice of the base year, as is often observed in the literature. However, the result of significant selection bias still holds true. Even after we change the base year, the coefficient of $\rho$ is significantly negative and a LR test for $H_0 : \rho = 0$ is rejected at a 1 percent level. Moreover, the estimated speed-of-convergence, $\lambda$, is faster in Model 1 than in Model 0. Thus, although the speed-of-convergence might depend on the base year, the existence of the downward bias to $\lambda$, due to selection bias, does not depend on the choice of base year.

**Threshold**

As indicated before, the survey covers both manufacturing and non-manufacturing firms with more than 50 employees.[15] One may be concerned with this truncation based on the threshold of 50 workers. In our data, firms with less than 50 workers are not covered in the survey, and thus a firm whose employment is reduced below this level is regarded as an exiting firm.

To check the sensitivity of our results with respect to this particular threshold of 50 workers, we re-estimated the baseline model for firms with more than 55 workers. The last column in Table 3 reports our estimation results. Despite the fact that 651 firms are eliminated from our sample, the speed-of-convergence is almost the same as the baseline model. Thus, our results are not sensitive to the threshold of truncation.

**Endogeneity**

Finally, one may raise the issue of endogeneity: $\ln \hat{\theta}_{i0}$ and $\mu_{iT}$ in the baseline model (7) might be correlated. To examine possible effects of this endogeneity, we applied instrumental variable (IV) methods to the speed-of-convergence equation (7).

In obtaining IV estimators, we employed Heckman's two-step estimation procedure (Heckit) rather than ML. In Section 3, we have used the ML method since this one-step procedure is generally more efficient than the two-step method of so-called Heckit estimation.[16] However, here we use the Heckit framework since it provides a straightforward extension to the case of endogeneity, which is unfortunately not the case in the ML method.

We first estimated the Mills ratio using probit model and the Mills ratio is used as an additional variable to estimate the speed-of-convergence equation. Instruments utilized are industry

[15] There is also a truncation based on the amount of paid capital. However, since paid capital is usually not a good indicator of firm size in practice, this truncation is considered not as serious as the truncation based on the number of employees.

[16] See Johnston and DiNardo (1997, p. 450) and Davidson and MacKinnon (1993, p. 545).

average capital stocks excluding firm $i$, modified industry average employment excluding firm $i$, returns-on-assets, and those deemed foreign-ownership dummy firms. Industry average capital excluding firm $i$ ($\tilde{K}_{it}$) and industry average labor excluding firm $i$ ($\tilde{L}_{it}$) are defined as follows:

$$\tilde{K}_{it} = \frac{1}{N-1}\left(\sum_{j=1, j\neq i}^{N-1} K_{jt} - K_{it}\right) \quad \text{and} \quad \tilde{L}_{it} = \frac{1}{N-1}\left(\sum_{j=1, j\neq i}^{N-1} L_{jt} - L_{it}\right),$$

where $N$ is the number of firms in an industry. In computing these industry averages of capital and labor, a firm's own capital and labor are subtracted to remove endogenous factors.

Table 4 presents the estimation results of Heckit and IV estimators. It is remarkable to note that the results from IV, taking account of Endogeneity, are almost the same as the results from Heckit ignoring endogeneity. This in turn implies that possible endogeneity between $\ln \hat{\theta}_{i0}$ and $\mu_{iT}$ does not lead to a significant bias, if any, in the speed-of-convergence estimation. Thus, we have ignored endogeneity as a problem in this paper.

Finally, the coefficient of $\beta_1$ from Heckit is smaller than the coefficient from ML in Table 2 reflecting the difference in estimation methods. In the remaining part of this paper we used ML estimation using (5) as our baseline model since the one-step ML approach is more efficient than the two-step Heckit approach.

Table 4: Robustness Check: Endogeneity

| | Model 1 | Model 1 |
|---|---|---|
| Selection equation | Yes | Yes |
| Estimation | Heckit | IV |
| $\beta_1$ | -0.075 | -0.076 |
| | (0.002) | (0.006) |
| Mills ratio | -0.039 | -0.040 |
| | (0.005) | (0.007) |
| $\lambda$ | 9.5% | 9.6% |
| $N$ | 18886 | 18886 |

Notes:

1 Standard errors are in parentheses.
2 Heckit: Heckman's two-step estimation method is used for the estimation.
3 IV: Instrumental variable method is used for the estimation.
4 For the IV results, standard errors are not adjusted.

## 3.3 Difference in Productivity Convergence across Industries

To examine the difference in the speed-of-convergence across industries, we include industry dummies to the baseline model for both constants and initial TFP levels. Thus, constants and initial TFP levels are now represented as:

$$\beta_0 + \zeta^1 D^1 + \cdots + \zeta^j D^j$$

and

$$\beta_1 \ln \hat{\theta}_{i0} + \omega^1 D^1 \ln \hat{\theta}_{i0} + \cdots + \omega^j D^j \ln \hat{\theta}_{i0},$$

where $D^j$ is a dummy variable that takes the value one if firm $i$ belongs to industry $j$. The speed-of-convergence of industry $j$, $\lambda^j$, is represented as:

$$-\frac{\{1-(1-\lambda^j)^T\}}{T} = \begin{cases} \beta_1 + \omega^j & \text{if } \omega^j \text{ is statistically significant at 10\% level;} \\ \beta_1 & \text{otherwise.} \end{cases}$$

Table 5 reports the number of industries by the magnitude of the speed-of-convergence, $\lambda$. Out of 70 industries, nearly two-thirds of the industries report less than 15 percent and 7 industries report more than 20 percent. Note that firms in manufacturing present a faster rate of convergence than firms in non-manufacturing. Similarly, firms in IT industries tend to converge faster than firms in non-IT industries.

Table 5: Distribution of the Speed-of-convergence (the number of industries)

| | Speed-of-convergence ($\lambda^j$) based on Model 1 | | | | | |
|---|---|---|---|---|---|---|
| | $< 3\%$ | $< 5\%$ | $< 10\%$ | $< 15\%$ | $< 20\%$ | $20\% \leq$ |
| All industry | 2 | 32 | 1 | 10 | 18 | 7 |
| Manufacturing | 1 | 11 | 0 | 9 | 17 | 3 |
| Non-manufacturing | 1 | 21 | 1 | 1 | 1 | 4 |
| IT industry | 0 | 1 | 0 | 2 | 3 | 1 |
| Non-IT industry | 2 | 31 | 1 | 8 | 15 | 6 |

# 4 Discussion

## 4.1 Importance of Technology Diffusion

We have confirmed strong evidence for productivity convergence among Japanese firms. But why is productivity convergence important? One reason may be that technology diffusion behind productivity convergence expands opportunities for secondary firms to catch up to the leading firm. Suppose that there was no technology diffusion. Without the diffusion, the secondary firms could not catch up to the leading firm without conducting their own R&D investment or purchasing technology through patents, which is very costly for new entrants and existing small- and medium-sized firms.

The same argument can be applied to the difference between developed and developing countries. Without technology diffusion, developing countries can catch up to developed countries only when they can afford to innovate or purchase technologies. This is also pointed out by Helpman (1993). His theoretical analysis shows that to tighten intellectual property rights benefit developed countries and hurt developing countries under some rates of imitation.

We should note, however, that instant technology diffusion causes yet a different problem. If technology diffuse too easily, no firms would have an incentive to conduct R&D investment. However, our results clearly indicate technology diffusion is not instantaneous but rather takes a long time. Thus, a technological advantage of a leading firm can still exist for a long time, which gives firms enough incentives to create innovative technologies.[17]

## 4.2 Difference in Selection Biases among Industries

Section 3.1 has shown that estimation without the selection equation causes a 1.5 percent point downward bias in the speed of convergence. We still have similar sample selection biases when we allow for different speeds of convergence in different industries.

Table 6 reports OLS regression results ignoring selection biases, that is, OLS without selection equations. The speed-of-convergence is much slower than the results presented in Table 5. Out of 70 industries, more than three-fourth of the industries report less than 15 percent and only two industries report more than 20 percent. The results clearly indicate that the estima-

[17] In a companion paper (Nishimura et al., 2005), we examine the difference of the effects between diffusion and innovation in more detail.

tion of productivity convergence without the selection equation underestimates the speed-of-convergence.

Table 6: Distribution of the Speed-of-convergence without Selection Equation

| | Speed-of-convergence ($\lambda^j$) based on Model 1 | | | | | |
|---|---|---|---|---|---|---|
| | $< 3\%$ | $< 5\%$ | $< 10\%$ | $< 15\%$ | $< 20\%$ | $20\% \leq$ |
| With selection equation | 2 | 32 | 1 | 10 | 18 | 7 |
| Without selection equation | 1 | 30 | 10 | 20 | 8 | 1 |

## 4.3 Determinants of Industry Differences: Manufacturing, IT, and Overall Growth

Tables 5 and 6 indicate that there exist large sectoral differences in the speed-of-convergence. There are at least three reasons. Firstly, the speed-of-convergence is different between manufacturing and non-manufacturing. Previous firm- or plant-level studies mainly focused on manufacturing firms (plants) and they did not give us much information about non-manufacturing sectors. Table 5 clearly shows a large difference between manufacturing and non-manufacturing firms.

Secondly, the difference between IT and non-IT industries matters. As we confirmed in Table 5, IT industry firms tend to catch up rapidly compared with non-IT industry firms. The existence of sectoral difference may be partly attributed to IT effects.

Finally, the speed-of-convergence is different between growing and declining industries. Table 7 presents the distribution of the speed-of-convergence, by industry productivity growth. Industry productivity growth is defined as the average of the productivity growth of firms in an industry. The table classifies industries into three groups: (1) of a negative productivity growth rate, (2) of a less than overall average growth rate (2.4 percent), and (3) of a more than or equal to overall average growth.

Table 7 indicates that the very rapid convergence is found in two polar cases: declining industries and rapidly-growing industries. In the declining industries having very rapid convergence, the observed high rate of convergence might be due to a "productivity collapse" of

Table 7: Distribution of the Speed-of-convergence, by the Industry Productivity Growth

| | Speed-of-convergence ($\lambda^j$) based on Model 1 | | | | | |
|---|---|---|---|---|---|---|
| Industry productivity growth | < 3% | < 5% | < 10% | < 15% | < 20% | 20% ≤ |
| < 0 % (declining industries) | 1 | 5 | 0 | 2 | 1 | 3 |
| < 2.4% (steadily growing industries) | 1 | 18 | 1 | 5 | 11 | 0 |
| ≥ 2.4 % (rapidly growing industries) | 0 | 9 | 0 | 3 | 6 | 4 |

leading firms rather than true catch-up of secondary firms. This is an interesting possibility that will be a topic of future research.

In the modest-rate range, the relatively fast productivity catch-up is mainly observed in the growing industries; while the relatively slow catch-up occurred in the steadily growing industries. To sum up, these results suggest that at least a part of the sectoral difference can be explained by the difference between growing and declining industries.

# 5 Concluding Remarks

This paper has examined the growth of productivity at the firm level, especially focusing on the effects of convergence. Our findings are summarized as follows. Firstly, the productivity convergence among firms exists not only in manufacturing but also in non-manufacturing industries. In addition, the speed-of-convergence is faster than the speed observed in the previous national- or industry-level studies, which indicate an annual average rate of 10.3 percent.

Secondly, analysis without considering the effects of exits causes a statistically significant sample selection bias in the speed-of-convergence estimation. We have used sample selection models to correct the bias and have found a 1.5 percent point downward bias in the speed-of-convergence in the analysis without the selection equation.

Finally, we have found a wide difference in the speed of productivity convergence among industries. IT industries tend to have faster rates of productivity convergence than non-IT industries. Similarly, manufacturing industries tend to have faster rates of convergence than non-manufacturing industries.

Two policy implications may be suggested from our analysis. Firstly, tightening intellectual property rights should be discussed more carefully. While the intellectual property rights could promote innovation (through R&D), the lack of technology diffusion prevents secondary firms from catching up to a leading firm. Since secondary firms are likely to be new entrants and/or small- and medium-sized firms, strong protection on intellectual property could be harmful for their growth.

Secondly, policy makers should recognize that not only innovation but also diffusion is an important source of productivity growth. The promotion of innovation is an important policy but it is important to also address a policy toward technology diffusion so that firms don not have to always invent everything by themselves. The combination of diffusion as well as innovation enables us to use limited resources more effectively.

# References

**Aw, B.Y., X. Chen, and M.J. Roberts**, "Firm-level Evidence on Productivity Differentials and Turnover in Taiwanese Manufacturing," *Journal of Development Economics*, 2001, *66* (1), 51–86.

**Baumol, W.J.**, "Productivity Growth, Convergence, and Welfare: What the Long-Run Data Show," *The American Economic Review*, 1986, *76* (5), 1072–1085.

**Bellone, F., P. Musso, and M. Quéré**, "Exiting Firms' Patterns: Evidence from a French Panel Data Set," *Paper presented at the International Industrial Organization Conference at Boston*, April 4-5 2003.

**Bernard, A.B. and C.I. Jones**, "Comparing Apples to Oranges: Productivity Convergence and Measurement Across Industries and Countries," *American Economic Review*, 1996, *86* (5), 1216–1239.

**Caves, D., L. Christensen, and W. Diewert**, "Output, Input and Productivity Using Superlative Index Numbers," *Economic Journal*, 1982, *92* (5), 73–96.

**Davidson, R. and J.G. MacKinnon**, *Estimation and Inference in Econometrics*, Oxford: Oxford University Press, 1993.

**Dollar, D. and E.N. Wolff**, "Convergence of Industry Labor Productivity among Advanced Economies, 1963-1982," *Review of Economics and Statistics*, 1988, *70* (4), 549–558.

**Dorwick, S. and Duc-Tho Nguyen**, "OECD Comparative Economic Growth 1950-85: Catch-Up and Convergence," *American Economic Review*, 1989, *79* (5), 1010–1031.

**Dunne, T., M.J. Roberts, and L. Samuelson**, "The Growth and Failure of U.S. Manufacturing Plants," *Quarterly Journal of Economics*, 1989, *104* (4), 671–698.

**Engel, C. and J.H. Rogers**, "How Wide Is the Border?," *American Economic Review*, 1996, *86* (5), 1112–1124.

**Fukao, K. and H.U. Kwon**, "Why Did Japan's TFP Growth Slow Down in the Lost Decade? An Empirical Analysis Based on Firm-Level Data of Manufacturing Firms," *Hi-Stat Discussion Paper Series*, December 2004, (50). Hitotsubashi University.

**Good, D., I. Nadiri, L.H. Roeller, and R.C. Sickles**, "Efficiency and Productivity Growth Comparisons of European and U.S/ Air Carriers: A First Look at the Data," *Journal of Productivity Analysis*, 1983, *4*, 115–125.

**Griffith, R., S. Redding, and H. Simpson**, "Productivity Convergence and Foreign Ownership at the Establishment Level," *The Institute for Fiscal Studies Working Paper*, 2002, (02/22).

**Heckman, J.**, "Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement*, 1976, *5* (4), 475–492.

**Helpman, E.**, "Innovation, Imitation, and Intellectual Propety Rights," *Econometrica*, 1993, *61*, 1247–1280.

**Johnston, J. and J. DiNardo**, *Econometric Methods*, 4 ed., New York: McGraw-Hill, 1997.

**Kimura, F. and K. Kiyota**, "Exports, FDI, and Productivity of Firm: Cause and Effect," *Faculty of Business Administration Working Paper*, 2004, (216). Yokohama National University.

**Levinsohn, J. and A. Petrin**, "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 2003, *70* (2), 317–341.

**McCallum, J.**, "National Borders Matter: Canada-U.S. Regional Trade Patterns," *American Economic Review*, 1995, *85* (3), 615–623.

**Nishimura, K.G., T. Nakajima, and K. Kiyota**, "DIffusion versus Innovation: Determinants of Productivity Growth amongJapanese Firms," *ESRI Working Paper*, 2005.

**__ , __ , and __** , "Does the Natural Selection Mechanism Still Work in Severe Recessions? Examination of the Japanese Economy in the 1990s," *Journal of Economic Behavior and Organization*, forthcoming.

**Olley, G.S. and A. Pakes**, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 1996, *64* (6), 1263–1297.

**Pascual, A.G. and F. Westermann**, "Productivity Convergence in European Manufacturing," *Review of International Economics*, 2002, *10* (2), 313–323.

**Sørensen, A.**, "Comparing Apples to Oranges: Productivity Convergence and Measurement Across Industries and Countries: Comment," *American Economic Review*, 2001, *91* (4), 1160–1167.

**Wolff, E.N.**, "Capital Formation and Productivity Convergence Over the Long Term," *American Economic Review*, 1991, *81* (3), 565–579.

**Wooldridge, J.M.**, *Econometric Analysis of Cross Section and Panel Data*, Cambridge: MIT Press, 2002.

# Appendix

We have assumed in the text that there is no systematic relationship between entry decision and productivity after entry, or in other words, that entry is randomly determined independently of variables that determine productivity after entry. Under this assumption, the ML method yields a consistent estimator.

To see this, suppose that, in addition to Type-1 and Type-2 firms, there are Type-3 firms (entrants after period 0). Assume that the firm $i$'s existence in year 0 depends on the productivity and other firm characteristics in year 0, $Z'_{i0} = (1, \ln\hat{\theta}_{i0}, z^1_{i0}, z^2_{i0}, \cdots)$:

$$w_{i0} = \begin{cases} 1 & \text{if } Z'_{i0}\zeta + \nu_{i0} \geq 0 \\ 0 & \text{otherwise} \end{cases}, \quad E(\nu_{i0}|Z_{i0}) = 0 \text{ and } \nu_{i0} \sim N(0,1).$$

Type-1 and Type-2 firms have $w_{i0} = 1$, while Type-3 firms have $w_{i0} = 0$. Assume that (i) $(\mu_{iT}, \nu_{iT})$ is independent of $Z_{i0}$ and bivariate normal with zero means as before and assume further that (ii) there is no systematic relationship between entry decision and productivity after entry, which is formally represented as an assumption that $\nu_{i0}$ is independent of $\mu_{iT}$, and $\nu_{iT}$ as well as $Z_{i0}$. This implies

$$E(s_{iT}|Z_{i0}, \nu_{i0}) = E(Z'_{i0}\gamma + \nu_{iT}|Z_{i0}, \nu_{i0}) = E(Z'_{i0}\gamma + \nu_{iT}|Z_{i0}) = Z'_{i0}\gamma$$

so that the ML method applied to estimate $s_{iT}$ using only Type-1 and Type-2 firms yields a consistent estimator.

Similarly, since $\nu_{i0}$ has no correlation with $\nu_{iT}$ and $\mu_{iT}$, we have

$$E(y_{iT}|Z_{i0}, \nu_{iT}, \nu_{i0}) = E(x'_{i0}\beta + \mu_{iT}|Z_{i0}, \nu_{iT}, \nu_{i0}) = x'_{i0}\beta + \rho\nu_{iT},$$

where $\rho_T$ is such that $E(\mu_{iT}|\nu_{iT}) = \rho\nu_{iT}$ as before. This in turn implies that $E(y_{iT}|Z_{i0}, \nu_{iT}, \nu_{i0}) = E(y_{iT}|Z_{i0}, \nu_{iT}) = E(y_{iT}|Z_{i0}, s_{iT})$, which is equivalent to equation (4). Thus, the ML method applied to estimate $y_{iT}$ using only Type-1 and Type-2 firms yields a consistent estimator.

This property is particularly important in our estimation, since crucial information about Type-3 firms in the estimating equation (2) is missing as exemplified in Table 1.