Industrial and Corporate Change, 2019, Vol. 28, No. 1, 123–138
doi: 10.1093/icc/dty066
Original article

OXFORD

Impacts of innovation, export, and other factors on firm employment growth in Chinese manufacturing industries

Jacques Mairesse[1] and Yilin Wu[2,*]

[1]CREST (ENSAE, France), UNU-MERIT Maastricht University, NBER (USA), The Netherlands.
e-mail: jacques.mairesse@ensae.fr and [2]Renmin University of China, Beijing, China.
e-mail: wuyilin@ruc.edu.cn

*Main author for correspondence.

Abstract

This article presents empirical evidence on the impacts of product innovation, export, and other important factors on employment growth in Chinese manufacturing industries over the period 2000–2006. The results of our analysis show that the overall demand effect of firms' output growth on their average yearly employment growth amounts to 7.0%, of which 3.5%, 1.2%, 1.6%, and 0.7% correspond to the output growth of, respectively, domestic-old, export-old, domestic-new and export-new products. The displacement effect from process and organizational innovations, as measured by firms' productivity efforts to catch up with industry regional productivity frontiers, accounts for a 5.4% average reduction of yearly employment growth. We also observe a trade-off between growth of productivity and growth of employment, which could have been on average higher by 2% for productivity (16.8% instead of 14.8%) and lower by 2% for employment (1.4% instead of 3.4%).

JEL classification: D2, J23, L1, O31, O33

1. Introduction

1.1 The issue

Innovation is widely considered to be one of the major factors of firm competitiveness and productivity, and a primary source of macro-economic growth. In China, like in nearly all developed and developing countries, policies to encourage firm-level innovation are very high on the economic agenda. The most critical concern of such policies is nonetheless whether innovation is by and large more beneficial than detrimental to employment. This issue is a particularly complex and difficult one. It is of paramount importance for China which faces the formidable challenge of a sustainable development with a very large pool of underemployed workers or workers in the informal sector.[1]

On the one hand, at the macro-economic level, product innovation and process innovation have been major sources of technological progress, and their long-run outcome has usually been accompanied by employment growth instead of the ever-decreasing levels of jobs that many predicted. On the other hand, at the micro-economic level, although evidence suggests that innovating firms are more likely to grow and survive than non-innovating firms, our

1 See Lundin *et al.* (2006, 2007); Ping *et al.* (2008) for our study period.

knowledge and understanding of the impacts of innovation on their employment remain to be much improved. Product innovation at the firm level contributes to stimulating demand, both domestic and foreign, for the firm's products, which leads, other things equal, to boost employment. Conversely, process innovation results in productivity gains for the firm, which brings about, other things equal, reductions in employment. Overall it is unclear to what extent, and through which other factors and mechanisms, firms' employment is affected.

At national level, in many countries and China particularly, exports are another crucial factor of employment, and export policies, as well as innovation policies, are most important. At the firm level, likewise and in conjunction to innovation, the employment impacts of exports remain to be better assessed and understood. Exports create demand for certain products, which requires an increase in labor demand. At the same time, exports may concur to stimulate product and process innovation and productivity, which also impact the growth of firm employment.

Many other correlates, determinants, or factors can be important to account for firm employment. Taking advantage of the firm-level data available for our study, we are able to measure not only firms' product innovation and export but also their investment in capital equipment and other assets, and their efforts to catch up to industry productivity frontiers, which proxy for process and organizational innovation. We can also consider the firm potential trade-off between increasing their employment and their workers' average wage. We know that these various factors have played a central role in China's extremely rapid economic development, and the main purpose of our econometric analysis is to better assess the absolute and relative magnitude of their effects on the growth of firm employment.

1.2 Present contribution

The issue of the impact of technological change on employment is an old one with many and rich scholarly contributions and debates on the relative importance of displacement and compensation effects in the short and long run. The empirical firm-level literature has been rapidly increasing in recent years; it has been spurred in particular by the development of innovation surveys such as the harmonized European Community Innovation Surveys (CIS).[2] Although it is now abundant, it remains unsatisfying and inconclusive in many respects. For example, while in many studies the evidence of significant, sizeable, and positive impacts for employment is well established for product innovation, it remains ambiguous and uncertain for process innovation. It is also the case that firm studies on the employment impacts of innovation have tended to be very focused and separated from studies focusing themselves on other very important employment determinants, such as exports, capital to labor substitution, wage increase, trade-off between employment growth, and productivity gains. Although analyzing a well-defined research question is generally preferable, it can be beneficial to consider together specific but closely related questions, whenever permitted by data availability.

Our study thus focuses on the comparative assessment of the effects of product innovation, exports, and other important factors on the employment growth in Chinese manufacturing industries as a whole, as well as within the various industries, different provinces, and categories of firms grouped by size and types of ownership, which is of particular interest in view of China immensity and diverseness. Among the other factors, we include investment in capital equipment and other assets, we account for the changes in wage levels, and we consider the firm distance to regional industry productivity. This variable measures firms' productivity catching up efforts, and proxy for process and organizational innovation indicators which are not available in our data.

Our study is based on the data of the yearly national industrial survey organized by the China National Bureau of Statistics (NBS), which was available to us for the 1999–2006 period. Precisely, as we shall explain, we have been able to construct an unbalanced panel data study sample by pooling five panel sub-samples, which are, respectively, balanced over the 4-year period 1999–2002, 2000–2003, 2001–2004, 2002–2005, and 2003–2006. To possibly include or use as instrument up to 2-year lagged variables in our model, we have also focused our econometric analysis on the yearly growth rate evidence for the past 2 years of each of these five sub-samples, that is on the growth rates for the years 2001–2002, 2002–2003, 2003–2004, 2004–2005, and 2005–2006.

Although rather short and not covering the more recent years, the study period 2001–2006 is highly interesting to consider. Since China transition from a centrally and fully controlled system in 1978 to a market economy and its

2 See for example, the literature reviews or overviews in Vivarelli and Pianta (2000), Chennells and Van Reenen (2002), Spiezia and Vivarelli (2002), Hall *et al.* (2008), Mairesse and Mohnen (2010), Vivarelli (2014), and Hou *et al.* (2019) in this issue.

accession to the World Trade Organization (WTO) in 2001, manufacturing industries have grown extremely fast and have been quite innovative and successful in exporting worldwide. One can also think that a number of our findings and conclusions hold well enough for the following decade during which the Chinese economy has pursued its extremely rapid development.

In Section 2 after this introduction, we explain how we generalize the original regression model of Harrison, Jaumandreu, Mairesse, and Peters (HJMP), which only considers the employment impact of innovation. We also explain in this section how we estimate our generalized model, how we address endogeneity and heterogeneity issues, and how we summarize and interpret our estimation results in terms of the contributions to firm employment growth of innovation, export, and the other factors considered. In Section 3, we provide the needed information on our data, study sample construction, and variable measurement, and we give some useful descriptive statistics on the growth of employment, output and productivity, and the other main variables of our analysis. In Section 4, we compare and discuss a series of estimation results for manufacturing as a whole, either based on instrumental variables (IVs) or ordinary least squares (OLS) estimators, and for different regression specifications going from the simple one of the HJMP model to the extended one of our generalized model. In Section 5, we present and comment the estimates of our model in terms of the employment growth decomposition in nine components associated to output growth in old and new, domestic and export products; distance to the productivity frontier; growth in average wage; and growth in capital equipment and other assets. We conclude in Section 6.

In Appendix Table A1 we document in detail the decomposition of growth employment in nine components for our study sample and separately for each of its the five four-year sub-samples (2001–2002 to 2005–2006). We provide in Table A2 a glossary of variables and in Table A3 information on the scope of our selected sample before cleaning (total number of employees and aggregate values for main variables per year). In complementary materials available from the authors, we also document the decomposition of growth employment separately for the 29 two-digit manufacturing industries, the five large Chinese regions (Bohai Rim, Yangtze River Delta, Pearl River Delta, Middle China, and West China), and for six groupings of firms defined by their size (large, medium, and small) and their types of ownership (state, private, and foreign).

2. Modeling the impact of innovation, export, and other factors on employment growth

2.1 The original model

Our model generalizes the orginal HJMP model to take advantage of the information of the NBS Annual Survey of Industrial Enterprise, especially the twofold decomposition of firm total output in new and old product output and in domestic and export output. The HJMP model (first circulated in 2005 and published in 2014) was itself purposely specified to take advantage of the CIS survey, mainly its specific measure of firm product innovation. This measure follows directly from the firms' answers to a survey question asking for the share of innovative sales in the total sales of the current year, where innovative sales correspond to the new or substantially improved products introduced in the current year but also during the two preceding years. From this measure it is also easy to retrieve two other variables, corresponding to the decomposition of total sales in the sales of "old" and "new" products. Note, as we shall see in Section 3, that in the NBS survey the measure of innovative or new products output is directly available as one of the survey variables but corresponds to innovative products introduced in the current year only (and not in the two preceding years).[3]

The HJMP model is built up as a linear regression equation between the growth rate of employment and the growth rates of old and new product output, with the constraint that the coefficient of growth rate of old product is equal to 1. Using already notations consistent with the ones we use for our generalized model, we can write the HJMP regression equation (1) as:

$$grl_{it} = \alpha + \beta_o wgrqo_{it} + \beta_n ngrqn_{it} + \epsilon_{it} \text{ with } \beta_o = 1 \tag{1}$$

3 Mairesse *et al.* (2011) explain how one can approximately relate the NBS survey measure of innovative products in the current year to CIS-type measures in terms of share of innovative sales over 3 years. They also document that the econometric estimates such as ours here for China provide qualitative evidence very close to the evidence which would have been obtained with such approximate CIS measures.

where grl_{it}, $wgrqo_{it}$, $ngrqn_{it}$ and ϵ_{it} denote for the firm (i) and year (t), respectively, the growth rates of employment, old product output, new product output, and a random disturbance with mean zero or idiosyncratic error, and where α, β_o, and β_n stand for the constant, and old and new product innovation parameters.

While grl_{it} is measured as the usual rate of growth of employment for firm (i) in the current year (t), that is the change between the employment level in the current year and in the previous year ($t-1$) relative to its level in this previous year, one must remember that the output product new in the current year (t) is by definition zero in the previous year ($t-1$), and hence $wgrqo_{it}$ and $ngrqn_{it}$ are measured slightly differently. We can write precisely:

$$grl_{it} = \frac{l_{it} - l_{i(t-1)}}{l_{i(t-1)}}, \quad wgrqo_{it} = \frac{qo_{it} - q_{i(t-1)}}{q_{i(t-1)}} \text{ and } ngrqn_{it} = \frac{qn_{it}}{q_{i(t-1)}}$$

where q_{it}, qo_{it} *and* qn_{it} denote, respectively, the levels of total output, old output, and new output. It is easy to verify that $wgrqo_{it}$ and $ngrqn_{it}$ simply add up to the rate of growth of total rate of growth, that is:

$$grq_{it} = \frac{q_{it} - q_{i(t-1)}}{q_{i(t-1)}} = wgrqo_{it} + ngrqn_{it}$$

Note that the HJMP model also includes the binary indicator of process innovation available in the CIS surveys. Such a variable is not available in our data, but we include in our econometric analysis a variable measuring the distance to the industry productivity frontier, which provides us with a broader and probably better measure for process and organizational innovation.[4, 5]

The constraint specification that β_o is unity means that β_n is the only parameter of interest and is estimated in terms of the relative impact of ngrqn with respect to wgrqo. In other words, the employment impact of new product growth output is less or higher than that of old product growth output if it is positive and less or higher than 1. Hence, the contribution of new product growth to employment growth will be smaller or higher than that of old product growth, depending on both the magnitude of β_n and the proportion of total output growth accounted for by new and old products, that is precisely on whether: $wgrqo > \beta_n ngrqn$ or $wgrqo < \beta_n ngrqn$.

2.2 The generalized model

In the version of the regression model we consider here, the growth rate of labor is the dependent variable, and the basic explanatory variables are, first, the growth rates for four kinds of products, not only old and new products but also domestic and export products, second, the distance to the technological frontier and its growth rate, and, third, the growth rates of average wage per employee and of total capital. A very large set of different indicators or dummies is also taken care of.

From now on, to simplify notations, we shall not always mention the firm and year indices i and t, and we shall denote $l.x$ or $l2.x$ the 1- or 2-year lagged values of variable x. Our generalized model can be written as the following linear regression model (2):

$$grl = \beta_{do} wgrqdo + \beta_{eo} wgrqeo + \beta_{dn} ngrqdn + \beta_{en} ngrqen + \sum_{k=1}^{K} \delta_k var(k) + \sum_{j=1}^{J} \alpha_j dum(j) + \epsilon \text{ with } \beta_{do} + \beta_{eo} = 1 \tag{2}$$

4 Note also that in the original HJMP model *wgrqo* is in fact measured as a log-growth rate . Here we keep *wgrqo* as a natural growth rate, as in Hall, Lotti and Mairesse (2008). It does not make a difference as long as log-growth rates remain a good approximation of natural growth rates, that is as long as they are small enough, which is not case for our Chinese data of output.

5 Note also that the HJMP model also includes the binary indicator of process innovation available in the CIS surveys. Such a variable is not available in our data, but we include in our econometric analysis a distance to the industry productivity frontier, which provides us with a broader and probably better measure for process and organizational innovation.

While *grl* is again the growth rate variable, the four product variables are the weighted growth rates of domestic old product and export old product, noted *wgrqdo* and *wgrqeo*, and the growth rates of domestic new product and export new product, noted *ngrqdn* and *ngrqen*. They are measured in the following way:

$$wgrqdo = \frac{(qdo - l.qd)}{l.q} = \frac{(qdo - l.qd)}{l.qd} * \frac{l.qd}{l.q}$$

$$wgrqeo = \frac{(qeo - l.qe)}{l.q} = \frac{(qeo - l.qe)}{l.qe} * \frac{l.qe}{l.q}$$

$$ngrqdn = \frac{qdn}{l.q} \text{ and } ngrqen = \frac{qen}{l.q}$$

We can verify that these four variables add up to the rate of growth *gq* of total output, that is:

$$grq = \frac{(q - l.q)}{l.q} = wgrqdo + wgrqeo + \ ngrqdn + ngrqen$$

The set of other variables $\sum_{k=1}^{K} \delta_k var(k)$ includes the distance to the 2-year lagged industry productivity frontier $l2.dis$, the growth rate of this frontier grf, the growth rate of the average firm wage per employee grw, and the growth rate of beginning of year total capital and 1-year lagged grc and $l.grc$. The productivity frontier is defined as the 95th percentile $p95$ of the log-labor productivity distribution at the two-digit industry level within the five large Chinese regions, that is $p95\lceil pr_{it} \rceil$, where pr_{it} denotes $(q_{it} - l_{it})$ the log-labor productivity in level of firm i in year t. The distance of the industry–region productivity frontier to the firm productivity is measured by the following gap: $dis_{it} = p95\lceil pr_{it} \rceil - pr_{it}$.[6]

The set of indicators or dummies $\sum_{j=1}^{J} \alpha_j dum(j)$ corresponds to five types of dummies: *dyr*, *reg*, *ind*, *own*, and *sca*. They stand precisely for the 5 four years balanced periods: 1999–2002, 2000–2003, 2001–2004, 2002–2005, and 2003–2006, 31 provinces grouped in five regions (Bohai Rim, Yangtze River Delta, Pearl River Delta, Middle China, and West China), 29 two-digit industries grouped in four technology classes (high-tech, medium-high-tech, medium-low-tech, and low-tech firm), three types of firm ownership (state, private, and foreign), and three groups of firm size (large, medium, and small).

2.3 Estimation and decomposition of employment growth

The issues of endogeneity of the innovation variables and that of correlated unobserved heterogeneity is carefully considered in HJMP and is addressed by instrumenting the new product sales growth variable by a variety of instruments (and by constraining the elasticity of old product output growth to be equal to 1). The HJMP econometric analysis is based on the 2000 harmonized European CIS for France, Germany, Spain, and the UK, which provide rich but mainly cross-sectional information for the year.[7] Unless one is able to match a CIS survey to previous CIS surveys or other sources of firm data, it is not possible to use lagged variables to instrument core variables of the model, which are a priori strongly endogenous, or to directly include in the model lagged variables, which are also relevant and thus would not need to be instrumented. This is in fact what we do here choosing as instruments of *wgrqdo*, *wgrqeo*, *ngrqdn*, *and ngrqen* lagged variables, which a priori are valid and strong enough. We use precisely as instruments seven variables lagged by 2 and 3 years, which are the following: indicators of the occurrence that firms invest in R&D, the level of R&D if they do, indicators of the occurrence that firms export, the level of export if they do, indicators of the occurrence that firms innovate, the level of new product output if they do, and the level of labor productivity.

6 We have also measured *dis* and *grf* by defining the productivity frontier as the p95 percentile of productivity distribution at the three-digit industry level within the five large Chinese regions, and we have found negligible differences in our econometric results when we used it instead of the two-digit industry level.

7 Germany performs the CIS survey every year, but the study used only data pertaining to the 2000 survey for the purpose of comparison with the three other countries.

On the basis of both the OLS and IV estimates for our most complete and preferred model, we interpret the economic significance of our results in terms of the decomposition of the overall employment growth in nine components: respectively, associated with the output growth of domestic old and new products $I(wgrqdo > 0)\hat{\beta_{do}}wgrqdo$ and $I(ngrqdn > 0)\hat{\beta_{dn}}ngrqdn$; with the output growth of export old and new products $I(wgrqeo > 0)\hat{\beta_{eo}}wgrqeo$ and $I(ngrqen > 0)\hat{\beta_{en}}ngrqen$; with the average productivity growth; with the lag distance to the frontier $\hat{\delta_1} l2.dis$ and with the growth of the frontier $\hat{\delta_2} gfr$; with growth of the average wage per employee $\hat{\delta_3} grw$; and with the growth of total capital $\hat{\delta_4} grc + \hat{\delta_5} l1.grc$.

3. Data, sample, variables, and descriptive statistics

3.1 Data source, cleaning, and adjustment

The basic source of our data is the Annual Survey of Industrial Enterprise organized by the Chinese NBS, which is actually a census of all state-owned firms, and private and foreign firms with sales higher than 5 million Yuan or RMB (RenMinBi). We have had access to nearly the full survey, with the exception of firms' names and addresses, for the years 1999–2006.

A general issue of micro-econometrics analysis is the need to "clean" thoroughly the raw data to delete from the study sample extreme outliers and most likely erroneous observations, as well as missing ones, for the variables central to the analysis. This is indeed an important task, which requires to be explained at least broadly. We thus have cleaned the raw data first in the levels of the main variables and second in their rate of yearly growth. In levels we have deleted firm*year observations with less than 10 employees, with gross output or sales revenue smaller than 5 million Yuan, and with average wage per employee negative or less than 1000 Yuan.[8] In terms of growth rates, we have deleted firm observations with extremely low or extremely high values of growth rates for employment, productivity, total gross output, total assets, and wage per employee.[9]

Besides cleaning, we had to deal with two specific missing data difficulties. Both could have limited strongly our study sample but proved to be minor. The solution to the first difficulty, which could have also prevented us to formalize our model the way we did, is at first glance somewhat complex. To implement our model, we need to know the four-way decomposition of firms' total output in "domestic-old," "domestic-new," "export-old," and "export-new" outputs, while we observe separately in the data the two-way decomposition of total output in domestic and export output and in old and new product output. We could have ignored the problem, since it concerns in fact only 6.5% of the observations, which we could have simply discarded from the study sample. However, we can do a little better, first, if we consider the distribution and occurrence of the different possible cases for the non-problematic 93.5% observations, which is very informative by itself, and, second, by assigning plausible values to the missing components for the 6.5% remaining cases, which is actually straightforward.

The 93.5% observations cover eight (i.e., 2^3) different cases: four of them with one component corresponding to domestic-old (56%), export-old (8.5%), domestic-new (0.4%), and export-new (0.1%); the other four with two components corresponding to domestic-old and export-old (22%), domestic-old and domestic-new (5.7%), export-old and export-new (0.5%), and domestic-new and export-new (0.3%). The 6.5% observations with missing components correspond to two situations, which can be simply described using our model notations. The first situation is the one where qdo and qdn are positive (with q = qdo + qdn); qe is positive and less than q; hence, qeo and qen are missing, and where we have obviously two symmetric possibilities: qeo = qe with qen =(q − qe) or qen = qe with qeo =(q − qe). The second situation is the one where similarly qeo and qen are positive (with q = qeo + qen); qn is

8 We also did some minor and rare adjustment or correction to the variables we use in the study when it appeared there were clear errors of declaration or compilation. In particular, we replaced negative export values or negative new product values by 0. We also replaced export output or new product output by the maximum of total output minus domestic output or minus old product output when it appeared they were too high (that their sums with domestic output or with old product output were larger than total output). We also adjusted export and new product output to 0 or to total output when their shares in total output were smaller than 1% or larger than 99%.

9 Precisely we only kept the consecutive observations for which all five growth rates are higher than 0.5 and lower than 7.4 for employment and capital growth rates, and higher than 0.5 and lower than 2.7 for productivity, gross output and wage per employee, which in terms of log-growth rates is equivalent to be in the range of −0.7 to 2 , or in the range of −0.7 and 1.

Table 1. Numbers of firms and growth rates observations

Balanced sub-samples	Numbers of firms	Number of growth rates observations								
		2001–2006	1999	2000	2001	2002	2003	2004	2005	2006
1999–2002	32,318	32,318	32,318	32,318	32,318	32,318	25,525	18,195	15,502	13,111
2000–2003	31,786	31,786	25,525	31,786	31,786	31,786	31,786	22,488	19,030	16,045
2001–2004	34,050	34,050	18,195	22,488	34,050	34,050	34,050	34,050	28,360	23,689
2002–2005	38,568	38,568	15,502	19,030	28,360	38,568	38,568	38,568	38,568	31,944
2003–2006	46,122	46,122	13,111	16,045	23,689	31,944	46,122	46,122	46,122	46,122
Total	74,527	182,844	32,318	38,579	50,141	60,349	67,734	58,436	52,746	46,122

positive and less than q; hence, qdo and qdn are missing, and we have the two symmetric possibilities: qdo = qd with qdn =(q − qe) or qdn = qe with qdo =(q − qe). In practice in these two situations we have simply chosen at random between the two alternative possibilities.

The econometric results reported here in the article correspond to this random imputation. We have checked that these results are hardly affected if we simply delete the 6.5% observations with missing four-way components when constructing our study sample.[10]

The second important missing data difficulty arose from the fact that the information on the new product output variable and the R&D variable, which we use as instrument to compute our IV estimates, were in 2004 both part of another NBS national survey focusing on research and innovation. This information was thus not collected for 2004 in the Industrial Enterprise NBS survey to which we had access. We just simply made up for this absence of R&D and new product output by taking the simple average of the corresponding values in 2003 and 2005 (whether they are positive or zeros).

3.2 Sample construction

Our study sample is a very unbalanced panel data sample due to the fact that many firms are leaving and entering our raw data set, and also as the result of missing firm*year observations following our extensive cleaning procedure. After several experiments, we have adopted what we found to be an appropriate compromise to preserve a partially balanced panel data structure without losing a large number of firms and observations.

As can be seen from Table 1, we have assembled an overall unbalanced panel data sample as the union of five panel sub-samples, respectively, balanced over the 4-year period 1999–2002, 2000–2003, 2001–2004, 2002–2005, and 2003–2006. Furthermore, to possibly include or use as instrument up to 2-year lagged variables in our model, we focus our econometric analysis on the yearly growth rate evidence for the last 2 years of each period, that is on the growth rates 2001–2002, 2002–2003, 2003–2004, 2004–2005, and 2005–2006.

Table 1 shows that the size of our overall study sample is of 182,844 observations for 74,527 firms, while the sizes of the five four-year sub-samples vary from roughly 32,000 to 46,000. If we had constrained it to be fully balanced, the size of our study sample would have been much smaller with respectively 78,666 observations for 13,111 firms. If we had constructed a study sample as the union of seven sub-samples balanced over 2 consecutive years, its size would have been much larger with, respectively, 333,898 observations for the same sample of 74,527 firms. Table A2 in Appendix A shows the detailed structure of our unbalanced panel data study sample.

3.3 Variables and descriptive statistics

Much can already be learned from simple descriptive statistics alone. They are also essential to understand the results of econometric analysis and interpret them rightly. Table 2 documents the four-way decomposition of total output in domestic and export output and old and new product output, by showing the number and proportion of observations with positive exports and positive new product output as well as their shares in total output. It also shows the

10 Our results are also basically unchanged if we rely on an average type of imputation assuming that the proportions of the missing components in total output are equal to the average proportions of corresponding components computed at an industry*region*year level.

Table 2. Four-way decomposition: proportions and shares of export and new output components

	Number of observations	Proportion of observations (%)	New output (%)	Export output (%)	Employment growth (%)	Productivity growth (%)
Domestic and old	102,313	56.0	0.0	0.0	2.2	15.4
Export and old	55,733	30.5	0.0	63.0	5.4	12.6
Domestic and new	11,064	6.0	34.5	0.0	2.9	16.8
Export and new	13,734	7.5	34.7	36.0	4.5	16.8
Total	182,844	100.0	4.7	21.9	3.4	14.8

Table 3. Simple descriptive statistics of variables in growth rate (%)[a]

	Number of employees	Employment growth grl	Productivity growth grp	Total output growth grq	Wage growth grw	Capital growth grc	Frontier distance in logs dis	Frontier distance delogged dist	Frontier growth grf
Mean	457.3	3.4	14.8	16.3	18.7	7.8	1.3	3.7	11.2
SD	1594.0	23.0	32.4	33.8	74.8	21.2	0.8	2.2	14.1
p5	40.0	−26.5	−30.0	−28.5	−34.6	−24.1	0.0	1.0	−6.1
p25	94.0	−6.1	−6.1	−3.9	−6.0	−3.6	0.8	2.2	5.3
p50	182.0	0.0	10.1	11.4	7.4	5.2	1.4	4.1	10.6
p75	396.0	8.4	30.0	30.8	26.1	18.5	1.9	6.7	17.1
p95	1514.0	42.9	76.1	76.1	91.9	45.8	2.6	13.5	30.2
Sample size	182,844	182,844	182,844	182,844	182,844	182,844	182,844	182,844	182,844

[a]Except the numbers of employees and the variable of distance to frontiers are in level, with dist and distdl respectively in logs and not in logs.

corresponding average growth rates of employment and labor productivity. We see that the case of domestic and old product output is by far the most frequent (56%), followed by that of export of old products (30.5% with a share of 63%). The occurrence and relative importance of the two cases of export or new product output are close, respectively, 7.5% or 6.0% with a share of 34.7% or 34.5%. Overall we see that the average share of total exports in total output amounts to a substantial 38.0%, while that of total of new product output only attains 13.5%. We can also stress that the overall average productivity growth rate is an impressive 14.8% per year compared to a much smaller but nonetheless sizeable average employment growth rate of 3.4% per year. The gap between productivity and employment growth rates is also very wide for each four components of total output but with significant differences: for export positive on employment and uncertain on productivity, and for new product output positive on productivity and uncertain on employment.

The simple statistics for most of the variables in our econometric analysis: mean, standard deviations, median, first and third quartiles, and 5th and 95th percentiles are given in Table 3. They are expressed in growth rates (except the number of employees and the distance to frontier), and they all vary considerably, in spite of the fact they are weakly correlated with size, at the difference of variables in levels. We see in particular that the range between the first and third quartiles p25 and p75 and that between the 5th and 95th percentiles p5 and p95 are extremely wide. For example, they vary, respectively, from (−6.1% to 8.4%) and (−26.5% to 42.9%) for the growth rates of employment and even much more from (−6.1% to 30.0%) and (−30.0% to 76%) for the growth rate of labor productivity. Interestingly, we observe that the mean growth rate of the wage per employee is higher than that of productivity (18.7% against 14.8%), while for median growth rate the contrary is true (7.4% against 10.1%), as could be expected. Note also that the distance to the frontier productivity variable is extremely dispersed with a median productivity 4.1.

4. Impact of innovation, export, and other factors on employment: overall results

We present, respectively, in Tables 4 and 5 the OLS and IV estimates of our generalized model (Columns LS-6 and IV-6) and the original HJMP model (Columns LS-2 and IV-2), as we specified them precisely in Section 2. Although we

Table 4. Main OLS estimates for total manufacturing

Specification	LS_1	LS_2	LS_3	LS_4	LS_5	LS_6
wgrqdo	1	1	0.411***	0.410***	0.410***	0.411***
	(0)	(0)	(0.001)	(0.001)	(0.001)	(0.001)
wgrqeo	1	1	0.589***	0.590***	0.590***	0.589***
	(0)	(0)	(0.001)	(0.001)	(0.001)	(0.001)
ngrqdn	1	0.883***	0.411***	0.391***	0.378***	0.377***
	(0)	(0.004)	(0.001)	(0.003)	(0.003)	(0.003)
ngrqen	1	0.883***	0.589***	0.533***	0.526***	0.523***
	(0)	(0.004)	(0.001)	(0.006)	(0.006)	(0.006)
l2.dis					−0.042***	−0.041***
					(0.001)	(0.001)
grf					−0.034***	−0.034***
					(0.004)	(0.004)
grw						−0.035***
						(0.001)
grc						0.044***
						(0.002)
l.grc						0.035
						(0.002)
constant	−0.129***	−0.121***	−0.042***	−0.041***	0.023***	0.021***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)
RMSE	0.324	0.324	0.221	0.221	0.219	0.217
Sample size	182,844	182,844	182,844	182,844	182,844	182,844

prefer a priori more complete and instructive specification, we also show in these tables the OLS and IV estimates for other intermediate specifications of interest, which allows us to judge the corresponding changes in the estimated parameters and their respective quality of fit in terms of root mean square errors (RMSE). We thus consider a sequence of six regression (R1–R6), which we can characterize first by the constraints they impose on the parameters $\beta_{do}, \beta_{eo}, \beta_{dn},$ and β_{en} of the four-way components of total output growth rate *wgrqdo*, *wgrqeo*, *wgrqdn*, and *wgrqen*, and next by the inclusion or not in the regression of the four important factors of log-distance to the productivity frontier and its growth rate *dis* and *grf*, and of average wage and total assets growth rates *grw* and *gfa*. These six regressions, which are sequentially embedded (except the second and third), are precisely defined as follows:

$$(\mathrm{R1})\ \ \beta_{do} = \beta_{eo} = \beta_{dn} = \beta_{en} = 1$$

$$(\mathrm{R2})\ \ \beta_{do} = \beta_{eo} = 1\ \ \beta_{dn} = \beta_{en} \neq 0$$

$$(\mathrm{R3})\ \ \beta_{do} = \beta_{dn} \neq 0\ \ \beta_{eo} = \beta_{en} \neq 0\ \text{ and }\ \beta_{do} + \beta_{eo} = 1$$

$$(\mathrm{R4})\ \ \beta_{do} \neq \beta_{eo} \neq \beta_{dn} \neq \beta_{en} \neq 0\ \text{ and }\ \beta_{do} + \beta_{eo} = 1$$

$$(\mathrm{R5})\ \ \beta_{do} \neq \beta_{eo} \neq \beta_{dn} \neq \beta_{en} \neq 0\ \text{ and }\ \beta_{do} + \beta_{eo} = 1\ \text{ with }\ dis,\ gfr$$

$$(\mathrm{R6})\ \ \beta_{do} \neq \beta_{eo} \neq \beta_{dn} \neq \beta_{en} \neq 0\ \text{ and }\ \beta_{do} + \beta_{eo} = 1\ \text{ with }\ dis,\ gfr,\ grw,\ gfa.$$

Regression (R1) only includes the five groups of indicators or dummies, respectively, *dyr* for the 5 yearly growth rates 2001–2002, 2002–2003, 2003–2004, 2004–2005, and 2005–2006 of our study sample, *reg* for the 31 Chinese provinces and 5 large regions, *ind* for the 29 two-digit and four technology classes industries, *own* for the three types

Table 5. Main IV estimates for total manufacturing

Specification	IV_1	IV_2	IV_3	IV_4	IV_5	IV_6
wgrqdo	1	1	0.695***	0.770***	0.436***	0.431***
	(0)	(0)	(0.013)	(0.021)	(0.019)	(0.018)
wgrqeo	1	1	0.305***	0.230***	0.564***	0.569***
	(0)	(0)	(0.013)	(0.021)	(0.019)	(0.018)
ngrqdn	1	1.124***	0.695***	0.861***	0.376***	0.368***
	(0)	(0.017)	(0.013)	(0.022)	(0.019)	(0.019)
ngrqen	1	1.124***	0.305***	0.584***	0.192***	0.205***
	(0)	(0.017)	(0.013)	(0.044)	(0.039)	(0.038)
l2.dis					−0.043	−0.042***
					(0.001)	(0.001)
grf					−0.035***	−0.035***
					(0.004)	(0.004)
grw						−0.035***
						(0.002)
grc						0.042***
						(0.003)
l.grc						0.033***
						(0.003)
constant	−0.067***	−0.064***	−0.024***	−0.020*	−0.006*	−0.004
	(0.013)	(0.013)	(0.010)	(0.010)	(0.008)	(0.008)
RMSE	0.324	0.326	0.256	0.277	0.221	0.219
Sample size	182,844	182,844	182,844	182,844	182,844	182,844

of firm ownership, and *sca* for the three groups of firm size. These indicators account for an important part of the fixed effect heterogeneity of the employment growth rate *grl* variability and four-way components of total output growth rate *grq* and other factors. Since they are also included in the five other regressions, regression (R1) is simply reported as a benchmark.[11]

Regression (R2) corresponds to our point of reference, the original HJMP model, since it only includes as an explicit factor of employment the growth rate of the new product output component (which implies $wgrqo = wgrqdo + wgrqeo$ and $ngrqn = ngrqdn + ngrqen$). Regression (R3) is an alternative to regression (R2), in which we consider that the export output component is the only explicit factor of employment growth rate (which implies $wgrqd = wgrqdo + ngrqdn$, and $ngrqe = wgrqeo + ngrqen$). We observe in Tables 4 and 5 that both OLS and IV estimated parameters of the four output components are statistically very significant but differ widely between the two regressions. This is not a priori too surprising since firm innovation and export behaviors are very distinct and specific, even if they can be related to a certain extent. Furthermore, we find that both the OLS and IV estimated RMSE are much smaller for the regression (R3) than for the regression (R2), which shows that overall the importance of export in accounting for the growth of employment is much higher than that of a new product. This is also not surprising, since as shown in Table 2 both the occurrence and shares in total output are much higher for exports than for new products.

Regression (R4), (R5), and (R6) combine in a sense regressions (R2) and (R3) by including as separate regressors the four-way new product and export output components. Regression (R5) takes also into account, respectively, the two year lagged log-distance to the region–industry productivity frontier *dis* and the growth rate of this frontier *grf*. Regression (R6) also considers the growth rates of the average wage per employee *grw*, and of the beginning of year and one year lagged total assets *grc* and *l.grc*. We find that the OLS estimates of the impact elasticities of the four new and export output components are quite stable in terms of order of magnitude across the three regressions with $\beta_{do} = 0.4,\ \beta_{eo} = 0.6,\ \beta_{dn} = 0.4,\ \beta_{en} = 0.5$. We also observe that they are similar to the ones of regression (R3),

11 As usual, if only for obvious reason of space, the estimated coefficients of the five groups of fixed effects are not recorded. However, we can note in Tables 4 and 5 that the OLS and IV estimated RMSE of regression (1) are equal as they should be with an RMSE = 0.324.

confirming the larger impact of export than of product innovation on employment growth. At first glance, the IV estimates can appear different from the corresponding OLS estimates and less stable across regressions. This is the case in particular of the IV and OLS estimates of regression (R2), the HJMP original specification, with $\beta_{dn} = \beta_{en} = 1.124$ versus $\beta_{dn} = \beta_{en} = 0.883$. A more careful look shows that the IV estimates of regression (R4) are similar enough to the ones of regression (R3) with $\beta_{do} = 0.8,\ \beta_{eo} = 0.2,\ \beta_{dn} = 0.9,\ \beta_{en} = 0.6$ versus $\beta_{do} = 0.7,\ \beta_{eo} = 0.3,\ \beta_{dn} = 0.7,\ \beta_{en} = 0.3$, both being quite different from the corresponding OLS estimates. Much more strikingly, we observe that the IV estimates of regressions (R5) and (R6), which differ much from the ones of regressions (R3) and (R4), are not only extremely close to each other but also very close to their OLS counterparts, with the only exception of a lower estimate for the impact of export of new products: $\beta_{en} = 0.2$ versus $\beta_{en} = 0.5$.

The presence of the log-distance to the productivity frontier *dis* and its growth rate *gfr*, which can be interpreted as a measure of process and organizational innovation in a broad sense, appears both as the main source of the sizeable differences of the IV estimates of regressions (R5) and (R6) with those of regressions (R3) and (R4), and of the fact of their closeness with the corresponding OLS estimates of regressions (R5) and (R6).

In view of the endogeneity of the four new and export components of total output such closeness of IV and OLS estimates for regressions (R5) and (R6) may seem surprising. It shows that the inclusion of *dis* and *gfr* is enough to remove a great part of the regression error ϵ accounting for the differences between our IV and OLS differences. It appears in other words that the origin of these differences is mainly the exclusion of process and organizational innovation as proxied by *dis* and *gfr*. Since OLS estimates are much more precisely estimated and usually more robust than IV estimates, they might thus be preferred to them.

It remains possible, however, that the IV and OLS estimates of regressions (R5) and (R6) would both similarly suffer from specification errors more important than the endogeneity biases and thus would be similarly biased. We have tried to minimize heterogeneity and errors in variables biases by adopting a growth rate specification taking care of firm-level fixed effects, by thoroughly cleaning the data and by including a very large set of dummy indicators. In supplementary materials where we present our results for the different two-digit 29 manufacturing industry, we also take care of the important heterogeneity in the various elasticities across industries. We have also chosen as instruments lagged variables that a priori are likely to be reasonably valid. Although the Sargan test of overidentification is strongly rejected for our overall sample, we also document in supplementary materials that it is often accepted at the industry level.

To summarize so far as we can accept that our instruments are good enough to correct for endogeneity and that in our choice of specification we have been able to avoid other potential sources of large biases, we can say that OLS as well as IV provide consistent estimates of our a priori preferred regression (R6).

5. Employment growth decomposition

As we have explained (last paragraph of subsection 2.2), in order to assess the economic importance of the various factors accounting for employment growth, we have computed their respective contributions, and present them in the bar charts of Figures 1 and 2 corresponding to the LS and IV regression estimates LS6 and IV6 of Tables 4 and 5. We can see right away that the bar-charts in the two Figures are basically the same with only few slight differences. This reflects of course the closeness of the estimates of regression (R6) we already stressed, in spite of the fact that differences in estimates are amplified when multiplied by the large growth rates of the different factors.[12]

Such amplifications of the differences in the estimated elasticities of the instrumented variables explain in particular why the employment growth rate contribution of *wgrqdo*, the domestic-old product output (3.5%) is larger by 2.3% than the contribution of growth rate of *wgrqeo*, the export-old product output (1.2%), although the estimated coefficient of the former (=0.41) is smaller by 30% than the estimated elasticity of the later (=0.59). The reason lies of course in the difference (5.5% per year) of the yearly average rates of *wgrqdo* (8.68%) and *wgrqeo* (2.1%).

The pattern of the distinct contributions of the different factors is very clear, irrespective of whether one prefers or not IV to LS estimates. For the sake of simplicity, however, we shall comment here the contributions shown of Figure 1 based on the LS estimates. For an average growth rate of employment of 3.4% per year over the six-year period 2000–2006 in Chinese manufacturing industries, we find, on the basis of the OLS estimates, that the growth

12 Note in this respect that the contributions of the variables not instrumented *l2.dis*, *grf*, *grw*, *grc* and *l.grc* are basically equal, since their coefficients in the LS and IV regressions do not practically differ. Note also that for clarity of the bar charts the contributions of the beginning of year total capital *grc* and its lag *l.grc* are added.

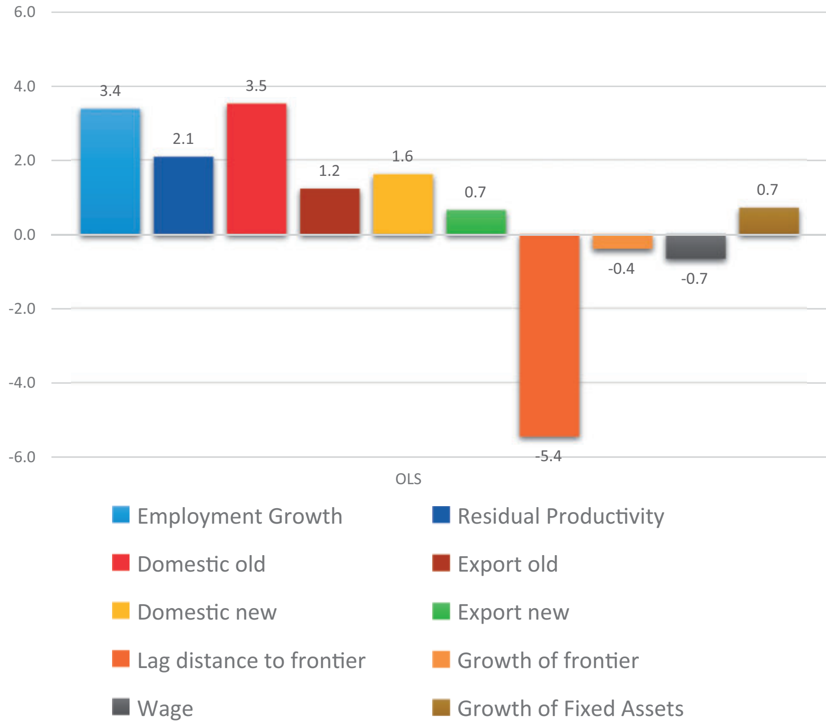

Figure 1. Employment growth decomposition (based on OLS estimates).

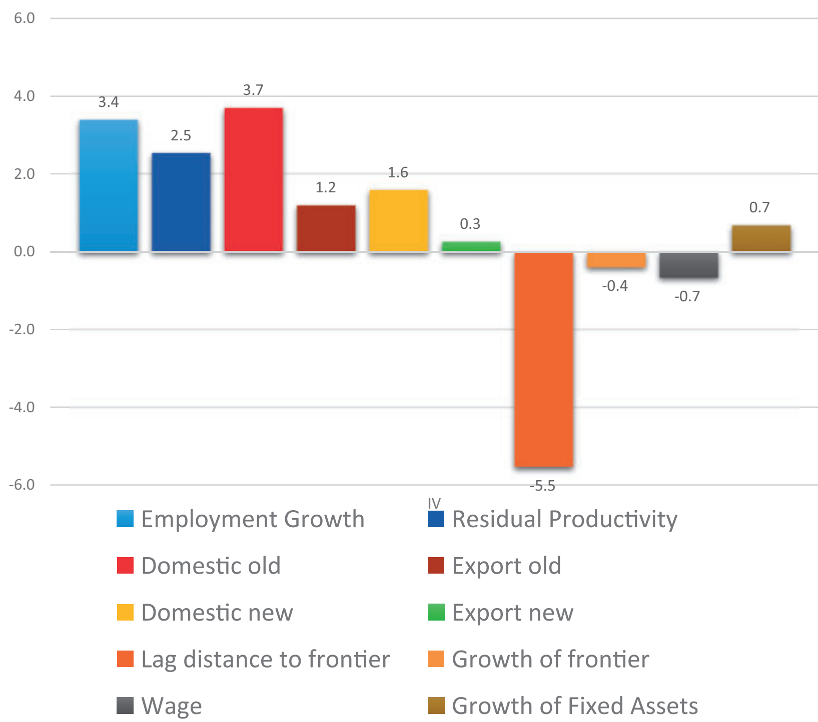

Figure 2. Employment growth decomposition (based on IV estimates).

of output for domestic and old product output *wgqdo* plays a major role, accounting for an average employment growth of 3.5%; followed by the growth for export and old product output wgqeo, for domestic and new product output *ngrqdn*, for domestic and new product output ***ngrqen***, of respectively 1.2%, 1.6% and 0.7%. Overall the change in demand for total output, every other factor being constant, would contribute to an employment growth rate of 7.0%, a doubling of the actual rate. This contribution on the basis of the IV estimates would be barely smaller of 6.8% (plus 0.2% for *wgrqdo*, less 0.4% for *ngrqen*).

The distance to the productivity frontier *l2.dis* and the growth of the productivity frontier *grf*, which largely reflect the firms' catching up efforts, are a massive countervailing factor accounting for a negative contribution of −5.8%. The tradeoff between increasing employment or average wages per worker *grw* also account for a much smaller but still significant negative contribution of −0.7%, which is exactly counterbalanced by the positive contribution of investment in total capital *grc* and *l.grc*. Interestingly, as reflected in the residual average productivity *grpres*, we see that the average employment growth *grl* could have been lower by about 2.0%, that is only 1.4% instead of 3.4%, and thus the average productivity growth grp would have been higher by 2.0%, that is 16.8% instead of 14.8%.

6. Conclusions and some further considerations

"*Who knows what is unattainable and yet will be doing it.*"[13]

Explaining and predicting well the growth of employment, likewise economic growth, is a formidable, probably an unattainable challenge. In the present study, we have merely tried to account for employment growth in Chinese manufacturing in relation to several major factors, mainly domestic and export production, product innovation, process and organizational innovation. In a nutshell, we have found that overall, in per year average, domestic producion has a very high contribution of at least 5% much higher than the positive one of at most 2% of export; and contribution of product innovation also of around 2%, which is far from compensating the strong negative one of nearly 6% for process and organizational innovation, as reflected in firms' efforts to catch up with the technological productivity frontier and management best practices.

Furthermore, we have found a tradeoff between increasing the growth of the average wage and that of employment around 0.7%, only one fifth of the 3.4% employment yearly growth rate. This can be regarded as a sign that firms do care about fighting unemployment. We observe a surprisingly low contribution of the same order of magnitude, 0.7%, for the growth of capital equipment and other assets, which would need particular investigation. Last but not least, as shown by the residual productivity trend of about 2%, which is actually of the same magnitude as the contribution of product innovation, there is also a substantial tradeoff between labor productivity and employment growth. The yearly average employment growth could have been of 1.4% only, instead of 3.4%, with a productivity growth of 16.8%, an even higher achievement than its actual 14.8%. This is another, even stronger, indication that firms do care about fighting unemployment, and not only about their productivity performance.

At this point, we also want to conclude by mentioning some limitations and making some tentative conjectures. The limitations to our study largely arise from data. The supplementary materials that can be obtained from the authors can already mitigate a few of them.[14] We just already alluded to the surprisingly low employment growth contribution of total capital. Part of it may be due to attenuation biases due to important errors in variables, pointing to the need of better measures, based on long enough investment series as well as detailed and accurate enough balance sheet information. But obviously the basic data limitation for our present study — and actually for many other economic researchers, foreign as well as Chinese, so far as we know — is its restricted availability to the period 1999 to 2006. Let us cite in particular Dosi and Yu (2019) in this issue, who use the same data for this period to address from a different perspective the same issues than us, with basically and comfortingly the same findings. Although we

13 Quote from Tchouang Tseu concerning Confucius, cited in German in a newspaper article by Herman Hesse (Neue Zurcher Zeitung, 7 april 1945), and translated in French as part of a book of a series of his articles ('Mes lectures préférées' in "Une Bibliothèque Idéale", Payot et Rivages, 2010). In Chinese: "Zhi qi bu ke wei er wei zhi".

14 As already said in the outline at the end of introductory Section 1, we also provide descriptive statistics and document Least Squares based contributions to employment growth at the very detailed level of twenty-nine manufacturing industries, five large Chinese regions, and six groupings of firms defined by size and types of ownership.

have not made any formal attempt to test for causality, we are fairly confident that our analysis, as it stands has causal implications. Nonetheless, its final empirical soundness and policy relevance basically lie on its confirmation and validation using data that would cover a much longer period until the recent years.

In the absence of such updated data, we can at least make a few tentative conjectures. The most important conjecture is that of an increasing policy relevance of the growth of employment agenda in the recent decade or so, with different implications on export, product, process and organizational innovation activities, as well as the wage-employment and productivity-employment tradeoffs. We can expect increasingly intence economic challenges with innovative activities for both the domestic and export markets having a growing importance, the productivity-employment tradeoff keeping its predominance and the wage-employment tradeoff becoming essential.

Acknowledgment

This article deepens and expands greatly a previous paper coauthored with professor Yanyun Zhao and professor Feng Zhen (Mairesse *et al.*, 2011), which emulates for China the framework of Harrison *et al.* (2014) for France, Italy, Spain, and the UK. The authors have benefitted from comments of participants to the various seminars and conferences, where drafts of this study were presented in the past five years. The authors are particularly thankful to Bronwyn Hall, Can Huang, Jordi Jaumandreu, Pierre Mohnen, Bettina Peters, Yanyun Zhao, and Feng Zhen. Yilin Wu also acknowledges that she has benefitted from the support of the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (Project No. 13XNK022).

References

Chennells, L. and J. Van Reenen (2002), 'Technical change and the structure of employment and wages: a survey on the microeconometric evidence', in N., Greenan, Y. L'Horty and J. Mairesse (eds), *Productivity, Inequality and the Digital Economy*. MIT Press: Cambridge, MA.

Doms, M., T. Dunne and M. J. Roberts (1995), 'The role of technology use in the survival and growth of manufacturing plants,' *International Journal of Industrial Organization*, **13**(4), 523–542.

Dosi, G. and X. Yu (2019). 'Technological catching-up, sales dynamics and employment growth: evidence from China's manufacturing', *Industrial and Corporate Change*, 79–107.

Entorf, H. and W. Pohlmeier (1990), 'Employment, innovation and export activity', in J. Florens, M. Ivaldi, J. Laffont and F. Laisney (eds), *Microeconometrics: Surveys and Applications*. Basil Blackwell: Oxford.

Hall, B. H., F. Lotti and J. Mairesse (2008), 'Employment, innovation, and productivity: evidence from Italian micro-data,' *Industrial and Corporate Change*, **17**(4), 813–839.

Harrison, R., J. Jaumandreu, J. Mairesse and B. Peters (2014), 'Does innovation stimulate employment? A firm-level analysis using comparable micro-data from four European countries,' *International Journal of Industrial Organization*, **35**, 29–43. http://www.eco.uc3m.es (2005) and NBER WP 14216, 2008.

Hou, J., C. Huang, G. Licht, J. Mairesse, P. Mohnen, B. Mulkay, B. Peters, Y. Wu, Y. Zhao and F. Zhen (2019), 'Does innovation stimulate employment? Evidence from China, France, Germany, and The Netherlands,' *Industrial and Corporate Change*, 109–121.

Lundin, N., S. Fredrik and Q. Jinchang (2006), 'The role of small firms in China's technology development', *Working Paper 227*, Stockholm School of Economics: Stockholm.

Lundin, N., S. Fredrik, P. He, and Q. Jinchang (2007), 'Technology development and job creation in China', *IFN Working Paper No. 697*, Research Institute of Industrial Economics: Stockholm.

Mairesse, J., W. Yilin, Z. Yanyun and Z. Feng (2011), *Employment Growth and Innovation in China: A Firm Level Comparison across Regions, Industries, Ownership Types and Size Classes*. ENSAE-CREST Mimeo paper (available J. Mairesse and Y. Wu).

Mairesse, J. and P. Mohnen (2010), 'Using innovation surveys for econometric Analysis', in B. H. Hall and N. Rosenberg (eds) *Handbook of the Economics of Innovation*, vol. 2. North-Holland: Amsterdam.

Peters, B. (2008), 'Innovation and firm performance: an empirical investigation for German firms', *Zew Economic Studies*, Vol. **38**. Physica: Heildeberg.

Ping, H., Q. Jinchang, L. Nannan and S. Fredrick (2008), 'Technology development and job creation in China's manufacturing sector', Mimeo Presented at the 2d MEIDE Conference, Renmin University: Beijing.

Spiezia, V. and M. Vivarelli (2002), 'Technical change and employment: a critical survey', in N. Greenan, Y. L'Horty and J. Mairesse (eds), *Productivity, Inequality and the Digital Economy*. MIT Press: Cambridge, MA.

Vivarelli, M. (2014), 'Innovation, employment and skills in advanced and developing countries: a survey of economic literature,' *Journal of Economic Issues*, **48**(1), 123–154.

Vivarelli, M. and M. Pianta (2000), *The Employment Impact of Innovation: Evidence and Policy*. London: Routledge.

Appendix A

Glossary of variables, main aggregate statistics, and employment growth decomposition by year

Table A1. Glossary of variables

Var	Explanation	Mean
grl	*Growth rate of employment*	*3.4*
grq	*Growth rate of output*	*16.3*
grp	*Growth of labor productivity defined as the growth rate of the ratio of output to employment*	*14.8*
grpres	*Growth rate of the productivity residual computed as the residual of the regressions of Tables 4 and 5*	*2.1*
wgrqdo	*Growth rate of domestic old product*	*8.6*
wgrqeo	*Growth rate of export old product*	*2.1*
ngrqdn	*Growth rate of domestic new product*	*4.3*
ngrqen	*Growth rate of export new product*	*1.3*
grw	*Growth rate of wage per employee*	*18.7*
grc	*Growth rate of beginning of year total capital*	*7.8*
dis	*The distance to the industry productivity frontier in the period t − 2 based on the two-digit industry level*	*1.3*
grf	*The growth rate of the frontier based on the two-digit industry level*	*11.2*
dyr	*Dummy variables for years of the research period. Here we construct the combination of five 4-year balanced sub-samples: respectively, on 1999–2002, 2000–2003, 2001–2004, 2002–2005, and 2003–2006*	–
reg	*Dummy variables for 31 provinces of mainland China*	–
ind	*Dummy variables for 29 industries of manufacturing industry*	–
own	*Dummy variables for three different ownership groups include state-owned group (state-owned firms, collective firms, cooperative stock enterprise, joint venture, limited company, and company with limited liability), private group (private firms), and foreign group (Hong Kong-, Macao-, and Taiwan-invested enterprises and foreign-invested enterprise).*	–
sca	*Dummy variables for three different scales of firms include big firms with numbers of employees more than 2000, median firms with numbers of empoyees between 300 and 2000, and small firms with numbers of employees less than 300*	–

Table A2. Numbers of firms and employees and aggregate values of main variables for the total sample by year after cleaning, and overall before cleaning

Year	Number of observations	Number of employees Million	Value added Trillion yuan	Sales Trillion yuan	Gross output Trillion yuan	New output Trillion yuan	Exports Trillion yuan	Wage Trillion yuan	Fixed assets Trillion yuan	Total assets Trillion yuan
1999	32,318	16.4	0.8	2.6	2.7	0.3	0.5	0.2	1.6	3.7
2000	38,579	18.3	1.0	3.6	3.7	0.5	0.7	0.2	2.0	4.6
2001	50,141	21.2	1.3	4.5	4.7	0.6	0.9	0.3	2.4	5.6
2002	60,349	23.8	1.6	5.7	5.9	0.8	1.2	0.3	2.7	6.6
2003	67,734	25.3	2.0	7.3	7.5	1.0	1.5	0.4	3.0	7.7
2004	58,436	21.8	2.1	7.8	7.9	1.0	1.7	0.4	2.8	7.3
2005	52,746	20.8	2.3	8.6	8.7	1.3	1.9	0.4	2.9	7.6
2006	46,122	19.1	2.4	9.2	9.2	1.4	1.9	0.4	3.0	7.8
Total (CS)	406,425	166.8	13.5	49.2	50.1	6.9	10.3	2.7	20.4	50.9
Total (RS)	1,365,159	384, 1	28, 6	106, 0	108, 8	11, 8	23, 0	5, 6	39, 4	103, 1

The row TOTAL(CS) before the last in Table A2 gives the total figures for the study sample (sum of the yearly figures for the _ years 1999 to 2006), while the last raw Total(RS) gives the corresponding figures for the Raw Sample before cleaning.

Table A3. Employment growth decomposition by year in percent

Years	*grl*	*wgqrdo*	*wgqreo*	*ngrqdn*	*ngrqen*	*l2.dis*	*grf*	*grw*	*grc*	*l1.grc*	Nb. obs. / grpres
Descriptive statistics[a]											
01–02	2.4	5.7	1.6	4.0	0.8	1.34	5.5	15.4	5.1	7.5	32,318
02–03	3.9	9.3	2.4	4.2	0.9	1.33	6.5	14.4	7.3	6.5	31,786
03–04	4.2	9.4	3.8	4.3	1.2	1.33	10.2	22.3	10.7	9.9	34,050
04–05	3.7	9.1	1.6	4.5	1.5	1.32	13.0	17.8	9.5	12.7	38,568
05–06	3.0	9.1	1.5	4.5	1.7	1.33	11.4	21.8	9.8	11.5	46,122
Study sample	3.4	8.6	2.1	4.3	1.3	1.33	11.2	18.7	8.6	9.9	182,844
Decomposition based on the whole sample estimates[b]											
01–02	2.4	2.3	0.9	1, 5	0.4	−5.5	−0.1	−0.5	0.2	0.3	2.9
02–03	3.9	3.8	1.4	1, 6	0.5	−5.5	−0.2	−0.5	0.3	0.2	2.1
03–04	4.2	3.9	2.2	1, 6	0.6	−5.4	−0.2	−0.8	0.5	0.3	1.5
04–05	3.7	3.7	0.9	1, 7	0.8	−5.4	−0.3	−0.6	0.4	0.4	2.0
05–06	3.0	3.7	0.9	1.7	0.9	−5.5	−0.3	−0.8	0.4	0.4	1.5
Study sample	3.4	3.5	1.2	1.6	0.7	−5.4	−0.4	−0.7	0.4	0.3	2.1
Decomposition based on estimates on separate five year sub-samples[b]											
01–02	2.4	2.3	1.0	1.4	0.4	−5.4	−0.1	−0.5	0.3	0.3	2.7
02–03	3.9	3.7	1.4	1.5	0.5	−5.4	−0.3	−0.6	0.5	0.3	2.2
03–04	4.2	4.0	2.2	1.6	0.6	−6.7	−0.3	−0.6	0.6	0.4	2.4
04–05	3.7	3.9	0.9	1.7	0.8	−5.1	−0.4	−0.7	0.2	0.4	2.0
05–06	3.0	3.6	0.9	1.8	0.9	−5.1	−0.2	−0.9	0.3	0.4	1.4
Study sample	3.4	3.5	1.2	1.6	0.7	−5.4	−0.4	−0.7	0.4	0.3	2.1

[a]The last column of the upper panel of Table A3 shows the number of observations Nb. obs.

[b]The last column of the two lower panels of Table A3 shows the growth rate of the productivity residual grpres.