the review of income and wealth

Review of Income and Wealth
Series 66, Number 4, December 2020
DOI: 10.1111/roiw.12455

PRODUCTIVITY DISPERSION AND MEASUREMENT ERROR

by Thomas von Brasch, Diana-Cristina Iancu and Terje Skjerpen*

Statistics Norway

Several reasons have been put forward to explain the high dispersion of productivity across establishments: quality of management, different input usage and market distortions, to name but a few. Although it is acknowledged that a sizable portion of productivity dispersion may also be due to measurement error, little research has been devoted to identifying how much they contribute. We outline a novel procedure for identifying the role of measurement error in explaining the empirical dispersion of productivity across establishments. The starting point of our framework is the errors-in-variable model consisting of a measurement equation and a structural equation for latent productivity. We estimate the variance of the measurement error and subsequently estimate the variance of the latent productivity variable, which is not contaminated by measurement error. Using Norwegian data on the manufacture of food products, we find that about one percent of the measured dispersion stems from measurement error.

JEL Codes: C23, J24, L11

Keywords: establishment performance, labor productivity, measurement errors, productivity dispersion

1. Introduction

It is widely accepted that the dispersion of productivity across establishments and industries is high. Dispersion is commonly measured by means of the standard deviation across establishments, where the productivity of each establishment is measured relative to a reference point, such as the mean productivity level at a given point in time. Using this procedure, it is typically found that the standard deviation across establishments is large and lies in the range of 30 to 100 percent; see Bartelsman and Wolf (2018).

Several reasons have been put forward to explain this high productivity dispersion: noisy selection (Jovanovic, 1982), sunk cost of entry (Hopenhayn, 1992), quality of management (Bloom and Van Reenen, 2010), different input usage, as the intensity of R&D or other intangible capital (Crepon *et al.*, 1998), product substitutability (Syverson, 2004), product market rivalry (Bloom *et al.*, 2013), market distortions (Hsieh and Klenow, 2009), skill-biased technical change and technological adoption (Dunne *et al.*, 2004) and innovation dynamics (Foster *et al.*, 2018), to name but a few. Although it is acknowledged that the high productivity dispersion may also be due to measurement error, little research has been devoted to identifying how much they contribute.

Note: We are grateful for helpful comments from two anonymous referees and the editor.
Correspondence to Terje Skjerpen, Statistisk sentralbyrå, Forskningsavdelingen, PO Box 2633 St. Hanshaugen, NO-0131 Oslo, Norway. (terje.skjerpen@ssb.no)

In this paper, we outline a novel procedure for identifying the role of measurement error in explaining empirical productivity dispersion across establishments. We define productivity as the log of the ratio between gross nominal output and the number of employee man-hours. One reason for the presence of measurement error in productivity is that our labor input variable relates to input according to the labor contract, which may deviate from the actual man-hours executed. Another source of measurement error is misclassification, which occurs when the main part of the establishment's production belongs to an industry other than the one considered, see e.g. Bartelsman *et al.* (2009, p. 28).

The starting point is the typical errors-in-variable framework consisting of a measurement equation and a structural equation for latent productivity. The key idea in our identification strategy is to estimate the variance of measurement error in a consistent way so that we can then estimate the variance of the companion latent variable, which is not contaminated by measurement error. To this end we build on the econometric theory of measurement error in dynamic models, see e.g. Komunjer and Ng (2014). Specifically, we remove time effects by means of a transformation in which we from each observation subtract time specific means of observation units that are present in all years. Unobserved establishment-specific heterogeneity is removed by differencing over time. The resulting model is a first-order autoregressive process in demeaned productivity growth rate. Measurement error variance and the productivity shock variance related to the development in latent productivity may be estimated by utilizing the covariance structure of the composite error terms. We estimate the amount by which productivity dispersion is reduced when measurement error is accounted for. Our findings indicate that about one percent of measured productivity dispersion is attributable to measurement error.

The rest of this paper is organized as follows: Section 2 outlines the procedure and the model for establishment-specific productivity. Section 3 describes the data and presents the results. Section 4 provides a conclusion.

2. Modeling Framework

Correcting for measurement error when assessing productivity dispersion across establishments presents a conceptual challenge. It can be illustrated analytically that measurement error increases the empirically observed dispersion compared with the dispersion in the latent productivity variable by considering the following econometric model

$$y_{it} = y_{it}^{*} + u_{it}, \tag{1}$$

where y_{it} (i.e. the log of the ratio between gross production and man-hours) denotes the observed productivity and y_{it}^{*} the latent productivity of establishment i in year t. The last symbol in Equation (1), u_{it}, denotes a random measurement error, i.e. it is assumed that $E(u_{it}) = 0 \forall i,t$ and that $E(u_{it}u_{j\tau}) = \delta_{ij}\delta_{t\tau}\sigma_{uu}^{2}$, where δ_{ij} and $\delta_{t\tau}$ denote Kronecker deltas such that $\delta_{ij} = 1$ if $i = j$, $\delta_{ij} = 0$ if $i \neq j$, $\delta_{t\tau} = 1$ if $t = \tau$ and $\delta_{t\tau} = 0$ if $t \neq \tau$. The symbol σ_{uu}^{2} denotes measurement error variance. Furthermore $E(y_{it}^{*}u_{j\tau}) = 0 \forall i,j,t,\tau$. Thus, the two terms on the right-hand side of Equation (1) are assumed to be uncorrelated. We assume that y_{it} and y^*_{it} follow trend stationary processes. In the empirical part of the paper we conduct a test to provide support for the trend stationary hypothesis. Let the time-invariant variances of the observed productivity and latent productivity variables be denoted σ^2_{yy} and $\sigma^2_{y^*y^*}$, respectively.[1] Under the imposed assumptions, it follows from taking the variance on both sides of Equation (1) that the presence of measurement error leads to wider productivity dispersion, i.e. $\sigma^2_{yy} > \sigma^2_{y^*y^*}$.

To identify how much of the variance of observed productivity is due to measurement error, we need a model for the latent level of establishment-specific productivity. Our point of departure is the standard model of technology diffusion used in the literature. The key idea in this model is that there should be an underlying driving force causing equalization of productivity if information can flow freely and know-how can be adopted easily. The further away an establishment is from the technology frontier, the higher the potential for technological catch-up and the higher the growth in productivity will be. Analytically, this may be represented by the following model for establishment-specific latent productivity:

$$y^*_{it} = y^*_{i,t-1} + \mu_i + \lambda(y^*_{F,t-1} - y^*_{i,t-1}) + \varepsilon_{it}, \quad for\ i = 1, \ldots, N_{t},$$

where μ_i captures the establishment's own rate of innovation through its underlying capabilities and ε_{it} is the stochastic shock to productivity growth. It is assumed that $E(\varepsilon_{it}) = 0 \forall i,t$ and that $E(\varepsilon_{it}\varepsilon_{j\tau}) = \delta_{ij}\delta_{t\tau}\sigma^2_{\varepsilon\varepsilon}$, where $\sigma^2_{\varepsilon\varepsilon}$ denotes the variance of the productivity shock. Furthermore, $E(\varepsilon_{it}u_{j\tau}) = 0 \forall i,j,t,\tau$. The term $y^*_{F,t-1} - y^*_{i,t-1}$ measures the distance between the technology level of establishment i and the frontier F, and $0 < \lambda < 1$ determines the speed of catch-up or technological adoption. The symbol N_t denotes the number of establishments present in year t. This model has been applied in numerous books (Banks, 1994; Benhabib and Spiegel, 2005; Acemoglu, 2009) and articles covering both technology adoption between countries (Griffith *et al.*, 2004; Madsen *et al.*, 2010) and technology adoption among establishments within countries (Cameron *et al.*, 2005; Griffith *et al.*, 2009). To proceed with the analysis of this model, we let the establishment's own rate of innovation μ_i be an unobserved establishment-specific fixed effect, and we let latent productivity at the frontier follow a deterministic function represented by fixed time effects, which leads to the following specification for development in latent establishment-specific productivity:

$$y^*_{it} = \beta y^*_{i,t-1} + \mu_i + \alpha_t + \varepsilon_{it}, \quad for\ i = 1, \ldots, N_t, \tag{2}$$

where $\beta = 1 - \lambda$ and μ_i and α_t are a fixed establishment and a fixed time effect, respectively. That β lies between zero and unity is implied by the model for technology diffusion, since $0 < \lambda < 1$.

[1]In the empirical application, we carry out sub-sample estimation by considering shorter time periods. Indirectly, this sheds some light on the assumption of time-invariant variances.

Inserting for y^*_{it} from Equation (1) into Equation (2) yields[2]

(3) $$y_{it} = \mu_i + \alpha_t + \beta y_{i,t-1} + \eta_{it},$$

where $\eta_{it} = \varepsilon_{it} + u_{it} - \beta u_{i,t-1}$. It follows from our assumptions that

$$\sigma^2_{\eta\eta} = Var(\eta_{it}) = \sigma^2_{\varepsilon\varepsilon} + (1+\beta^2)\sigma^2_{uu}.$$

To remove the time effects, we make use of a transformation given by the difference between the observation and the time-specific mean of the observations for establishments present in all years. Let B denote a set containing all units observed in all years and let N_B denote the number of such units. Later we will refer to B as the reference group. All other observational units are in the set $\underline{B}$.[3] Let

$$\bar{y}^B_t = \frac{1}{N_B}\sum_{k\in B} y_{kt},$$

$$\bar{\mu}^B = \frac{1}{N_B}\sum_{k\in B} \mu_k,$$

$$\bar{\varepsilon}^B_t = \frac{1}{N_B}\sum_{k\in B} \varepsilon_{kt}$$

and

$$\bar{u}^B_t = \frac{1}{N_B}\sum_{k\in B} u_{kt}.$$

It therefore follows that

(4) $$y^d_{it} = \mu^d_i + \beta y^d_{i,t-1} + \varepsilon^d_{it} + u^d_{it} - \beta u^d_{i,t-1},$$

where

$$y^d_{it} = y_{it} - \bar{y}^B_t,$$

$$\mu^d_i = \mu_i - \bar{\mu}^B,$$

$$\varepsilon^d_{it} = \varepsilon_{it} - \bar{\varepsilon}^B_t$$

[2]One could also consider the case of systematic measurement error, such that Eq. (1) is augmented with an intercept. In that case Eq. (3) will also contain an intercept. However, this parameter is not identifiable since fixed time effects for all years also are present in the equation.

[3]To avoid additional symbols, we also use B and $\underline{B}$ in sub- and superscripts to indicate that the measures relate to establishments in the sets B and $\underline{B}$, respectively.

and

$$u_{it}^d = u_{it} - \overline{u}_t^B.$$

Equation (4) is in the form used by Komunjer and Ng (2014). By differencing over time, we may filter out the time-invariant term μ_i^d. Such a transformation yields

$$\Delta y_{it}^d = \beta \Delta y_{i,t-1}^d + \Delta \varepsilon_{it}^d + \Delta u_{it}^d - \beta \Delta u_{i,t-1}^d. \tag{5}$$

The transformation underlying Equation (5) implies introducing heteroscedasticity, which can easily be corrected for and which vanishes asymptotically. For observational units outside the reference group, i.e. $i \in \underline{B}$, we multiply Equation (5) by $\left[(N_B+1)/N_B\right]^{-0.5}$ and for observational units within the reference group, i.e. $i \in B$, we multiply Equation (5) by $\left[(N_B-1)/N_B\right]^{-0.5}$. After this rescaling, we obtain the following equation

$$\Delta y_{it}^w = \beta \Delta y_{i,t-1}^w + \Delta \varepsilon_{it}^w + \Delta u_{it}^w - \beta \Delta u_{i,t-1}^w, \tag{6}$$

where
$\Delta y_{it}^w = \left[(N_B+1)/N_B\right]^{-0.5} \Delta y_{it}^d$ for $i \in \underline{B}$ and $\Delta y_{it}^w = \left[(N_B-1)/N_B\right]^{-0.5} \Delta y_{it}^d$ for $i \in B$. The other symbols in Equation (6) are defined by analogous expressions. Note that $\Delta y_{i,t-1}^w$ is correlated with the composite error term, $\Delta \varepsilon_{it}^w + \Delta u_{it}^w - \beta \Delta u_{i,t-1}^w$. The same is true for $\Delta y_{i,t-2}^w$, since this lagged difference is correlated with $\beta \Delta u_{i,t-1}^w$. Hence, we employ the variable $\Delta y_{i,t-3}^w$, which is not correlated with the composite error term, as an identifying instrument. The IV estimate obtained for β is referred to as $\tilde{\beta}_{IV}$.

Let the composite error term in Equation (6) be defined by:

$$\xi_{it} = \Delta \varepsilon_{it}^w + \Delta u_{it}^w - \beta \Delta u_{i,t-1}^w.$$

It follows from our assumptions that the following holds true:

$$Var(\xi_{it}) = \sigma_{\xi\xi}^{iitt} = 2\left(\sigma_{\varepsilon\varepsilon}^2 + (1+\beta+\beta^2)\sigma_{uu}^2\right),$$

$$Cov(\xi_{it}, \xi_{i,t-1}) = \sigma_{\xi\xi}^{iit,t-1} = -\left(\sigma_{\varepsilon\varepsilon}^2 + (1+\beta)^2\sigma_{uu}^2\right),$$

$$Cov(\xi_{it}, \xi_{i,t-2}) = \sigma_{\xi\xi}^{iit,t-2} = \beta\sigma_{uu}^2$$

and

$$Cov(\xi_{it}, \xi_{i,t-s}) = 0 \forall s \geq 3.$$

The transformation undertaken introduces some correlation between the observational units. There are three different cases. If $i,j \in \underline{B}$ we obtain

$$\sigma_{\xi\xi}^{ijtt\underline{B}} = Cov(\xi_{it}, \xi_{jt}) = \frac{2}{N_B+1}\left(\sigma_{\varepsilon\varepsilon}^2 + (1+\beta+\beta^2)\sigma_{uu}^2\right),$$

$$\sigma_{\xi\xi}^{ijt,t-1\underline{B}} = Cov(\xi_{it},\xi_{j,t-1}) = -\frac{1}{N_B+1}\left(\sigma_{\varepsilon\varepsilon}^2 + (1+\beta)^2\sigma_{uu}^2\right),$$

$$\sigma_{\xi\xi}^{ijt,t-2\underline{B}} = Cov(\xi_{it},\xi_{j,t-2}) = \frac{\beta\sigma_{uu}^2}{N_B+1}$$

and

$$\sigma_{\xi\xi}^{ijt,t-s\underline{B}} = Cov(\xi_{it},\xi_{j,t-s}) = 0 \forall s \geq 3.$$

Second, if both observational units are within the reference group, i.e. $i,j \in B$, we have

$$\sigma_{\xi\xi}^{ijttB} = Cov(\xi_{it},\xi_{jt}) = \frac{2}{N_B-1}\left(\sigma_{\varepsilon\varepsilon}^2 + (1+\beta+\beta^2)\sigma_{uu}^2\right),$$

$$\sigma_{\xi\xi}^{ijt,t-1B} = Cov(\xi_{it},\xi_{j,t-1}) = -\frac{1}{N_B-1}\left(\sigma_{\varepsilon\varepsilon}^2 + (1+\beta)^2\sigma_{uu}^2\right),$$

$$\sigma_{\xi\xi}^{ijt,t-2B} = Cov(\xi_{it},\xi_{j,t-2}) = \frac{\beta\sigma_{uu}^2}{N_B-1}$$

and

$$\sigma_{\xi\xi}^{ijt,t-sB} = Cov(\xi_{it},\xi_{j,t-s}) = 0 \forall s \geq 3.$$

Third, if observational unit $i \in B$ and observational unit $j \in \underline{B}$ we have

$$Cov(\xi_{it},\xi_{j,t-s}) = 0 \forall s.$$

The variance and autocovariances of the composite error term may be estimated from the residuals. Furthermore, in Appendix A we show how we estimate the covariances between the composite errors of different observational units. Let the estimates of $\sigma_{\xi\xi}^{iit,t-s}, \sigma_{\xi\xi}^{ijt,t-s\underline{B}}$ and $\sigma_{\xi\xi}^{ijt,t-sB}$ (s = 0, 1, 2) be $\hat{\sigma}_{\xi\xi}^{iit,t-s}, \hat{\sigma}_{\xi\xi}^{ijt,t-s\underline{B}}$ and $\hat{\sigma}_{\xi\xi}^{ijt,t-sB}$, respectively. Consider the following vector equation

$$(7) \quad \begin{bmatrix} \hat{\sigma}_{\xi\xi}^{iitt} \\ \hat{\sigma}_{\xi\xi}^{iit,t-1} \\ \hat{\sigma}_{\xi\xi}^{iit,t-2} \\ \hat{\sigma}_{\xi\xi}^{ijtt\underline{B}} \\ \hat{\sigma}_{\xi\xi}^{ijt,t-1\underline{B}} \\ \hat{\sigma}_{\xi\xi}^{ijt,t-2\underline{B}} \\ \hat{\sigma}_{\xi\xi}^{ijttB} \\ \hat{\sigma}_{\xi\xi}^{ijt,t-1B} \\ \hat{\sigma}_{\xi\xi}^{ijt,t-2B} \end{bmatrix} = \begin{bmatrix} 2 & 2(1+\tilde{\beta}_{IV}+\tilde{\beta}_{IV}^2) \\ -1 & -(1+\tilde{\beta}_{IV})^2 \\ 0 & \tilde{\beta}_{IV} \\ 2/(N_B+1) & [2/(N_B+1)](1+\tilde{\beta}_{IV}+\tilde{\beta}_{IV}^2) \\ -1/(N_B+1) & -[1/(N_B+1)](1+\tilde{\beta}_{IV})^2 \\ 0 & [1/(N_B+1)]\tilde{\beta}_{IV} \\ 2/(N_B-1) & 2/(N_B-1)](1+\tilde{\beta}_{IV}+\tilde{\beta}_{IV}^2) \\ -1/(N_B-1) & -[1/(N_B-1)](1+\tilde{\beta}_{IV})^2 \\ 0 & [1/(N_B-1)]\tilde{\beta}_{IV} \end{bmatrix} \begin{bmatrix} \sigma_{\varepsilon\varepsilon}^2 \\ \sigma_{uu}^2 \end{bmatrix} + \begin{bmatrix} r_0 \\ r_1 \\ r_2 \\ r_0^{\underline{B}} \\ r_1^{\underline{B}} \\ r_2^{\underline{B}} \\ r_0^B \\ r_1^B \\ r_2^B \end{bmatrix},$$

where the last vector on the right-hand side contains errors. We estimate the two second-order parameters, $\sigma^2_{\varepsilon\varepsilon}$ and σ^2_{uu}, by applying the OLS formula to Equation (7). To assess estimation uncertainty related to the estimates of $\sigma^2_{\varepsilon\varepsilon}$ and σ^2_{uu}, we apply bootstrapping; see Appendix B.

In empirical work, attention is often devoted to the standard deviation of productivity less the mean productivity of the establishments that are present in a specific year. Within our (superpopulation) framework, this measure corresponds approximately to $\sqrt{\sigma^2_{yy}}$ and $\sqrt{\sigma^2_{y^*y^*}}$, where the former is the standard deviation of observed log-productivity and the latter a model-based measure after correction for measurement error. Thus, our estimates of $\sqrt{\sigma^2_{yy}}$ and $\sqrt{\sigma^2_{y^*y^*}}$ should produce numbers which are comparable to those reported elsewhere in the literature.

3. Empirical Application

We apply our framework using unbalanced panel data for manufacture of food products in Norway from the years 2000–2014. Our data on gross output are from the Central Register of Establishments and Enterprises and data for labor input from the State Register of Employers and Employees.[4] Table 1 provides information about the number of observations and about properties of the unbalanced panel data set for the industry, whereas Table 2 provides summary statistics on the (untransformed) productivity variable. Nominal gross production is measured in 1,000s of NOK, whereas labor input is measured as number of man-hours. The mean value added per man hour worked over the sample period is NOK 1,410 per man hour, while the median is somewhat lower, at NOK 1,003 per man hour. In line with findings in other countries, there is also a wide labor productivity spread across establishments that manufacture food products in Norway. The standard deviation is about 85 percent of the mean labor productivity level.

We have looked at the time series properties of the observed (log-transformed) productivity variable to see whether trend stationarity is a reasonable assumption. To this end, we have employed the balanced part of the panel data set[5] for the industry and considered the test provided by Harris and Tzavalis (1999).[6] The test statistic is based on a first-order autoregressive regression augmented with establishment-specific fixed effects and establishment-specific linear trends and fixed-T asymptotics. Under the null hypothesis non-stationarity prevails. We find that the null-hypothesis is firmly rejected. The significance probability is for practical reasons equal to zero.

Besides showing estimation results based on the full data set, we also present, in Table 3, results from two sub-periods, 2005–2014 and 2000–2009. These two periods are considered for reasons of robustness. Using the full sample, the autoregressive slope parameter, β, is estimated to be 0.794. The estimate is clearly significant, with

[4]See https://www.nav.no/en/Home/Employers/NAV+State+Register+of+Employers+and+Employees

[5]See Table 1.

[6]This part of the calculations has been carried out using Stata version 15.1. TSP version 5.1 was used for all other calculations.

TABLE 1
Properties of the unbalanced panel data set

No. of obs.	No. of obs. Units	No. of obs. units present in all years	No. of obs. units without contiguous time series
25,953	3,875	600	668

TABLE 2
Summary statistics based on establishment-specific means of untransformed labor productivity[a]

No. of obs. units	Mean	Std. dev.	First quartile	Median	Third quartile	Min.	Max.
3,875	1.410	1.203	0.574	1.003	1.848	0.103	10.098

[a]The total number of observations is 25,953.

a t-value (based on a robust estimate of the standard error) of about 4.3. Table 3 also reports the estimates of the two variance parameters. When the full sample is used, the estimates of productivity shock variance, $\sigma^2_{\varepsilon\varepsilon}$, and of measurement error variance, σ^2_{uu}, are 0.136 and 0.018, respectively. The corresponding results using data for the two sub-periods are not very far from those obtained using all data. In Appendix B, we report the results of an exercise in which we used bootstrapping in the full data case to generate standard errors of the estimates of σ^2_{uu} and $\sigma^2_{\varepsilon\varepsilon}$. The t-value of the estimate of measurement error variance, σ^2_{uu}, is about 2.4. In a one-sided test, this corresponds to a p-value of about 0.009. Thus, the estimate of the measurement error variance is significant. As mentioned in Appendix B, some of the replications needed to be disregarded because of a negative estimate of the measurement error variance or because the estimate of the autoregressive parameter, β, exceeded 1. The occurrence of negative estimates of error component variances under unconstrained estimation is a well-known problem in panel data econometrics; see for instance Maddala (1971) and more recently Bun *et al.* (2017). Thus, the quality of the obtained standard errors must be evaluated in view of this feature.

It is possible to estimate the two variance parameters, $\sigma^2_{\varepsilon\varepsilon}$ and σ^2_{uu}, without involving cross-moments of residuals between different observational units. This option corresponds to omitting the last six rows of Equation (7). The estimation results using this simplified procedure are very similar to those reported in Table 3. The reason is that all moments related to different observational units are very small, as are the corresponding values in the 9 × 2 matrix multiplied by the vector consisting of the two variances in Equation (7).

From the estimates of the two variance parameters we can derive the proportion of the variation of the composite error terms stemming from productivity shock and measurement error, respectively. The results are reported in Table 4. We carry out the decomposition both for the full sample and for the two subperiods. The last column of Table 4 shows the results for the full sample covering the years 2000–2014. When this period is considered, about 18 per cent of the variation of the composite error can be attributed to measurement error, whereas the remaining 82 percent can be attributed to productivity shocks. Thus, measurement error captures a substantial part of the variation in composite error.

TABLE 3
ESTIMATES OF FIRST AND SECOND ORDER PARAMETERS[a]

	β		$\sigma^2_{\varepsilon\varepsilon}$	σ^2_{uu}
Time period	Estimate	*t*-value[b]	Estimate	Estimate
2000-2014	0.794	4.274	0.136	0.018
2005-2014	0.626	3.529	0.120	0.025
2000-2009	0.671	2.622	0.125	0.013

[a]Using data for the full-time period 2000–2014, the number of observations used to estimate β and the two variance parameters are 12,635 and 8,831, respectively. Using data for the period 2005–2014, the number of observations used to estimate β and the two variance parameters are 6,290 and 3,592, respectively. Using data for the period 2000–2009, the numbers of observations used to estimate β and the two variance parameters are 7,483 and 4,282, respectively.
[b]Based on analytical formula for robust standard errors.

To relate our results to the applied literature on productivity dispersion, we focus in Table 5 on the standard deviation of productivity. We report results showing the difference between the observed standard deviation, $\sqrt{s^2_{yy}}$, and the estimated standard deviation of latent productivity, operationalized as $\sqrt{\hat{\sigma}^2_{y^*y^*}} = \sqrt{s^2_{yy} - \hat{\sigma}^2_{uu}}$. The observed variance of log productivity based on data from all years, i.e. s^2_{yy} is taken as the estimator of σ^2_{yy}. The results reported in Table 5 provide information about the positive bias caused by neglecting measurement error when reporting figures on productivity dispersion. The effect is fairly small, amounting to about one percent for the full sample, nor is it very far from one percent when the two sub-periods are considered.

A one percent contribution from measurement error to productivity dispersion is relatively small. By way of comparison, it should be noted that the contribution from measurement error in Norway may be lower than in many other countries. The reason is the long-standing tradition in Scandinavian countries of using administrative data for research purposes and in the construction of the National Accounts. As pointed out by Barth (2012), administrative data are accurate, because they are entered for purposes such as accounting, tax reporting etc. that are subject to strict control and auditing rules. In addition, and in contrast to survey data, many of the administrative registers contain the entire population, which eliminates the problem of sampling error. In this article, we have used data on gross output from the Central Register of Establishments and Enterprises and data on labor input from the State Register of Employers and Employees, which is a matched employer-employee data set. Given that data based on administrative registers are less prone to measurement error, the contribution of measurement error to productivity dispersion may be larger in countries where productivity data are based on surveys. The empirical framework we have outlined in this article can be used to test the merit of this hypothesis, or to analyze the extent to which measurement error can explain the size of productivity dispersion in other countries, but this is an area we leave open for future research.

TABLE 4
DECOMPOSITION OF THE ESTIMATED VARIANCE OF THE COMPOSITE ERROR

Time period	Variance of composite error[a]	Contribution of variance of composite error stemming from productivity shocks (in %)[b]	Contribution of variance of composite error stemming from measurement error (in %)[c]
2000-2014	0.165	82.251	17.749
2005-2014	0.155	77.411	22.589
2000-2009	0.143	86.674	13.326

[a]Recall that $\sigma^2_{\eta\eta} = \sigma^2_{\varepsilon\varepsilon} + (1+\beta^2)\sigma^2_{uu}$.

[b]The contribution is given by $100 \times \sigma^2_{\varepsilon\varepsilon} / \sigma^2_{\eta\eta}$.

[c]The contribution is given by $100 \times (1+\beta^2)\sigma^2_{uu} / \sigma^2_{\eta\eta}$.

TABLE 5
Spread in observed and latent productivity. Percent

Spread	Period		
	2000–2014	2005–2014	2000–2009
$\sqrt{s_{yy}^2}$	84.6[a]	88.2[b]	81.3[c]
$\sqrt{\sigma_{y^*y^*}^2} = \sqrt{s_{yy}^2 - \sigma_{uu}^2}$	83.5	86.8	80.5

[a]Empirical standard deviation based on 25,953 observations.
[b]Empirical standard deviation based on 16,151 observations.
[c]Empirical standard deviation based on 18,476 observations.

4. Conclusion

In this article, we have outlined a novel procedure for identifying the role of measurement error in explaining empirical productivity dispersion across establishments. The starting point of our framework is the classical errors-in-variable model consisting of a measurement equation and a structural equation for latent productivity. The key idea in our identification strategy has been to estimate measurement error variance in order to deduce the variance of the latent productivity variable. Specifically, we have estimated a differenced demeaned dynamic panel data model where establishment-specific productivity is modelled as a first-order autoregressive process. Using the case of manufacture of food products in Norway as an illustrative example, we found that about 1 percent of the measured dispersion is due to measurement error.

A topic that deserves more attention in further work is the presence of negative estimates of error component variances, see e.g. Bun *et al.* (2017). In this article, this feature emerged when obtaining standard errors of the estimate of measurement error variance by means of non-parametric bootstrapping. Some of the replications had to be disregarded.

References

Acemoglu, D., *Introduction to Modern Economic Growth*, Princeton University Press, Oxford, 2009.

Banks, R., *Growth and Diffusion Phenomena: Mathematical Frameworks and Applications*, Springer-Verlag, Heidelberg, 1994.

Bartelsman, E. J. and Z. Wolf, "Measuring Productivity Dispersion," in Grifell-Tatjé, E., C. A. K. Lovell, and R. C. Sickles (eds), *The Oxford Handbook of Productivity Analysis*, Oxford University Press, Oxford, Chap. 9, 593–624, 2018.

Bartelsman, E., J. Haltiwanger, and S. Scarpetta, "Measuring and Analyzing Cross-Country Differences in Firm Dynamics," inT. Dunne, J. B. Jensen, and M. J. Roberts (eds), *Producer Dynamics: New Evidence from Micro Data*, University of Chicago Press, Chicago, IL, Chap. 1, 15–76, 2009.

Barth, E. Administrative Data for Research and Evaluation. Peer Review on "Evaluation of Labour Market Policies and Programmes: The Use of Data-Driven Analyses", *Mutual Learning Programme, Belgium*, 19-20 November. Eurostat, 2012.

Benhabib, J., and M. M. Spiegel, "Human Capital and Technology Diffusion," in P. Aghion and S. N. Durlauf (eds), *Handbook of Economic Growth*, Vol. 1., North-Holland, Amsterdam, Chap. 13, 935–66, 2005.

Bloom, N., and J. Van Reenen, "Why Do Management Practices Differ Across Firms and Countries?," *Journal of Economic Perspectives*, 24, 203–24, 2010.

Bloom, N., M. Schankerman, and J. Van Reenen, “Identifying Technology Spillovers and Product Market Rivalry,” *Econometrica*, 81, 1347–93, 2013.
Bun, M. J. G., M. A. Carree, and A. Juodis, “On Maximum Likelihood Estimation of Dynamic Panel Data Models,” *Oxford Bulletin of Economics and Statistics*, 79, 463–94, 2017.
Cameron, G., J. Proudman, and S. Redding, “Technological Convergence, R&D, Trade and Productivity Growth,” *European Economic Review*, 49, 775–807, 2005.
Crepon, B., E. Duguet, and J. Mairesse, “Research, Innovation and Productivity: An Econometric Analysis at the Firm Level,” *Economics of Innovation and New Technology*, 7, 115–58, 1998.
Dunne, T., L. Foster, J. Haltiwanger, and K. R. Troske, “Wage and Productivity Dispersion in United States Manufacturing: The Role of Computer Investment,” *Journal of Labor Economics*, 22, 397–429, 2004.
Foster, L., C. Grim, J. C. Haltiwanger, and Z. Wolf, “Innovation, Productivity Dispersion, and Productivity Growth,” NBER Working Paper 24420, September 2018.
Griffith, R., S. Redding, and J. Van Reenen, “Mapping the Two Faces of R&D: Productivity Growth in a Panel of OECD Industries,” *Review of Economics and Statistics*, 86, 883–95, 2004.
Griffith, R., S. Redding, and H. Simpson, “Technological Catch-Up and Geographic Proximity,” *Journal of Regional Science*, 49, 689–720, 2009.
Harris, R. D. F. and E. Tzavalis, “Inference for Unit Roots in Dynamic Panels where the Time Dimension is Fixed,” *Journal of Econometrics*, 91, 201–26, 1999.
Hopenhayn, H. A., “Entry, Exit, and Firm Dynamics in Long Run Equilibrium,” *Econometrica*, 60, 1127–50, 1992.
Hsieh, C.-T. and P. J. Klenow, “Misallocation and Manufacturing TFP in China and India,” *Quarterly Journal of Economics*, 124, 1403–48, 2009.
Jovanovic, B., “Selection and the Evolution of Industry,” *Econometrica*, 50, 649–70, 1982.
Komunjer, I. and S. Ng, “Measurement Errors in Dynamic Models,” *Econometric Theory*, 30, 150–75, 2014.
Maddala, G. S., “The Use of Variance Components Models in Pooling Cross Section and Time Series Data,” *Econometrica*, 39, 341–58, 1971.
Madsen, J. B., M. R. Islam, and J. B. Ang, “Catching Up to the Technology Frontier: The Dichotomy between Innovation and Imitation,” *Canadian Journal of Economics*, 43, 1389–411, 2010.
Syverson, C., “Product Substitutability and Productivity Dispersion,” *Review of Economic Statistics*, 86, 534–50, 2004.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher’s web site: