ONLINE APPENDIX

A Data Appendix

A.1 CEP Management and Innovation Survey Dataset

A.1.1 The Survey Sampling Frame

We use a sub-set of the CEP Management and Organization survey in this paper (see Bloom, Sadun and Van Reenen, 2012b, for full details of larger sample) where we have ICT data (see below). Our sampling frame was based on the Bureau van Dijk (BVD) Amadeus dataset for Europe (France, Germany, Italy, Poland, Portugal, Sweden and the U.K.) and Icarus for the US. These databases all provide sufficient information on companies to conduct a stratified telephone survey (company name, address and a size indicator). These databases also typically have some accounting information, such employment, sales of capital assets. Apart from size, we did not insist on having accounting information to form the sampling population, however.

Amadeus is constructed from a range of sources, primarily the national registries of companies (such as Companies House in the UK). Icarus is constructed from the Dun & Bradstreet database, which is a private database of over 5 million US trading locations built up from credit records, business telephone directories and direct research. In every country the sampling frame was all firms with a manufacturing primary industry code with between 100 and 5,000 employees on average over the most recent three years of data (typically 2002 to 2004).[1]

Interviewers were each given a randomly selected list of firms from the sampling frame. This should therefore be representative of medium sized manufacturing firms. The size of the sampling frame appears broadly proportional to the absolute size of each country's manufacturing base, the US, has the most firms and Sweden and Portugal the least.[2] In addition to randomly surveying from the sampling frame described above we also tried to resurvey the firms we interviewed in the 2004 survey wave used in Bloom and Van Reenen (2007). This was a sample of 732 firms from France, Germany, the UK and the US, with a manufacturing primary industry code and 50 to 10,000 employees (on average between 2000 and 2003). This sample was drawn from the Amadeus dataset for Europe and the Compustat dataset for the U.S. Only companies with accounting data were selected.[3] As a robustness test we also drop the firms that were resurveyed from 2004.

[1] In the US only the most recent year of employment is provided. In Portugal the population of firms with 100 to 5000 employees was only 242, so we supplemented this with the 72 firms with 75 to 100 employees. We checked the results by conditioning on common size bands (above 150 in all countries).

[2] The size of the manufacturing sector can be obtained from http://laborsta.ilo.org/, a database maintained by ILO.

[3] So, for the UK and France this sampling frame was very similar to the 2006 sampling

A.1.2 Sample Representativeness

Comparing the aggregate number of employees for different size bands from our sampling frame with the figures for the corresponding manufacturing populations in each of the countries (obtained from national census data), we find that in all countries but two the sampling frame broadly matches up with the population of medium sized manufacturing firms. This suggests our sampling frame covers the population of all firms. In Germany and Portugal the coverage is less complete as the frame appears to cover around a third of manufacturing employees. To address this problem we always include country fixed-effects to try to control for any differences across countries. Second, we control for size and industry. This should help to condition out some of the factors that lead to under/over sampling of firms. Finally, we made sure the results were robust to dropping Germany and Portugal.

45% of the firms we contacted took part in the survey: a high success rate given the voluntary nature of participation. Of the remaining firms 17% refused to be surveyed, while the remaining 38% were in the process of being scheduled when the survey ended. The decisions to reject the interview is uncorrelated with revenues per worker, listing status of the firm or firm age. Large firms and multinationals were more likely to respond although the magnitude of this effect is small (e.g. multinationals were about 7% more likely to agree to the interview and firms about 4 percentage points more likely for a doubling in size).

A.2 Harte Hanks Data

The ICT data used is constructed using the Ci Technology Database (CiDB) produced by the international marketing and information company Harte Hanks (HH). Harte-Hanks is a NYSE listed multinational that collects IT data primarily for the purpose of selling on to large producers and suppliers of IT products (e.g. IBM, Dell etc). Their data is collected for over 160,000 plants across 20 European countries, and another 250,000 across the US. The US branch has the longest history with the company beginning its data collection activities in the mid 1980s.

Harte Hanks surveys plants (referred to as "sites" in the CiDB database) on a rolling basis with an average of 11 months between surveys. This means that at any given time, the data provides a "snapshot" of the stock of a firm's IT. The CiDB contains detailed hardware, equipment and software information at the plant level. Areas covered by the survey include PCs, many types of software, servers, storage and IT staff (including development staff such as programmers). The fact that HH sells this data on to major firms like IBM and Cisco, who use this to target their sales efforts, exerts a strong market discipline on the data quality. If there were major discrepancies in the collected data this would

frame. For Germany it is more heavily skewed towards publicly quoted firms since smaller privately held firms do not report balance sheet information. For the US it comprised only publicly quoted firms. As a result when we present results we always include controls for firm size.

rapidly be picked up by HH's clients when they placed sales calls using the survey data, and would obviously be a severe problem for HH future sales.[4] Because of this HH run extensive internal random quality checks on its own data, enabling them to ensure high levels of data accuracy.

Another valuable feature of the CiDB is its consistency of collection across countries. The data for Europe is collected via a central call centre in Dublin and this ensures that all variables are defined on an identical basis across countries. This provides some advantages over alternative strategies such as (for example) harmonizing government statistical register data collected by independent national agencies.

HH samples all firms with over 100 employees in each country. Thus, we do lose smaller firms, but since we focus on manufacturing the majority of employees are in these larger firms. It is also worth noting this survey frame is based on firm employment - rather than plant employment - so the data contains plants with less than 100 employees in firms with multiple plants. Furthermore, HH only drops plants from the survey if they die or repeatedly refuse to answer over several years, so that the sampling frame covers all firms that have had at 100 employees in any year since the survey began. In terms of survey response rate HH reports that for the large European countries (UK, France, Germany, Italy, and Spain) they had a response rate of 37.2% in 2004 for firms with 100 or more employees. Bloom, Draca and Van Reenen (2011) provide further information on the HH dataset.

A.3 Firm level accounting data

Our firm accounting data on sales, employment, capital, profits, shareholder equity, long-term debt, market values (for quoted firms) and wages (where available) came from Amadeus dataset for Europe (France, Germany, Italy, Poland, Portugal, Sweden and the U.K.) and on Icarus for the US

A.4 Leased Line Data

The data on cross national prices is given by OECD (2007). Although European prices have been falling over the past decade due to liberalizations and pressures from the regulators (e.g. European Commission DG-Competition), there remains considerable concern about differential degrees of competition and regulation generating cross-national price disparities. "Local leased line prices remain of concern where there is insufficient competition. For users in these areas this means that incumbents can continue to charge prices that are not disciplined by competition. For new entrants it means that incumbents may price local leased circuits in an anti-competitive manner" (OECD Communication Outlook, 2005).

"Leased lines are provided by traditional telecom operators. New market entrants have their own networks but need to link their customers' premises to

[4] HH also refunds data-purchases for any samples with error levels above 5%.

it. This link is called a 'leased line part circuit' and is usually provided by the incumbent. The availability at the wholesale level of these links at reasonable prices is a necessary condition for a competitive leased lines retail market and for pro-competitive downstream 'knock-on' effects" (European Commission Report, 2002)

A major turning point in the pricing of leased lines took place in 1998 when a significant number of European countries fully liberalized their telecommunication markets. The impact of increasing liberalization is evident in the OECD's Index of leased line prices. At the distances of 50 and 200 kilometers the leased lines (2Mbit/s) index fell from 77 in 1997 to 31 by 2004. This process happened at a much faster rate in some countries than others (see OECD, 2005).