American Economic Review 2012, 102(6): 2437–2471
http://dx.doi.org/10.1257/aer.102.6.2437

Markups and Firm-Level Export Status†

By Jan De Loecker and Frederic Warzynski*

In this paper, we develop a method to estimate markups using plant-level production data. Our approach relies on cost-minimizing producers and the existence of at least one variable input of production. The suggested empirical framework relies on the estimation of a production function and provides estimates of plant-level markups without specifying how firms compete in the product market. We rely on our method to explore the relationship between markups and export behavior. We find that markups are estimated significantly higher when controlling for unobserved productivity; that exporters charge, on average, higher markups and that markups increase upon export entry. (*JEL* D22, D24, F14, L11, L60)

Estimating markups has a long tradition in industrial organization and international trade. Economists and policymakers are interested in measuring the effect of various competition and trade policies on market power, typically measured by markups. The empirical methods that were developed in empirical industrial organization often rely on the availability of very detailed market-level data with information on prices, quantities sold, characteristics of products, and more recently are supplemented with consumer-level attributes.[1] Often, both researchers and government agencies cannot rely on such detailed data, but still need an assessment of whether changes in the operating environment of firms had an impact on markups and therefore on consumer surplus. In this paper, we provide a simple empirical framework in the spirit of Hall (1986) to estimate markups. Our approach nests various price-setting models used in applied industrial organization and international trade and relies on optimal input demand conditions obtained from standard cost minimization and the ability to identify the output elasticity of a variable input free of adjustment costs. The methodology relies crucially on the insight that the output elasticity of a variable factor of production is only equal to its expenditure share

* De Loecker: Princeton University, Fisher Hall, Princeton, NJ 08540, NBER and CEPR (e-mail: jdeloeck@princeton.edu); Warzynski: Aarhus University, Hermodsvej 22, DK8230 Aarhus, Denmark (e-mail: fwa@asb.dk). This paper was previously circulated under the name "A Control Function Approach to Estimate Markups" and has benefitted from seminar and conference participants at KU Leuven, NYU, Aarhus University, IIOC 2008, Minneapolis Applied Micro Conference, EIFI 2009,CEPR ERWIT 2009, Wharton Penn, NBER PR, Chicago, Vanderbilt, SED 2010, EARIE 2010, NYU IO Day 2010, Cowles (Yale), and Stanford University. We thank our discussants at the various conferences and workshops, and especially Amil Petrin for his thoughtful discussion on an early version of the paper. Furthermore, we want to thank Dan Ackerberg, Andrew Bernard, Steve Berry, Allan Collard-Wexler, Jeremy Fox, Tim Kehoe, Joep Konings, Sam Kortum, Jim Levinsohn, Marc Melitz, Ariel Pakes, Steve Redding, Mark Roberts, Esteban Rossi-Hansberg, Jim Tybout, Patrick Van Cayseele, Hylke Vandenbussche, and Frank Verboven for discussions on an earlier draft. We also thank three anonymous referees for comments and suggestions that help improve the paper.
† To view additional materials, visit the article page at http://dx.doi.org/10.1257/aer.102.6.2437.
[1] See Goldberg (1995) and Berry, Levinsohn, and Pakes (2004) for examples.

in total revenue when price equals marginal cost of production. Under any form of imperfect competition, however, the relevant markup drives a wedge between the input's revenue share and its output elasticity.

Markup estimates are obtained using production data where we observe output, total expenditures on variable inputs, and revenue at the plant level, a condition that is satisfied in most plant-level datasets. In principle the approach relies on estimating output elasticities and we therefore require a measure of output that does not pick up price differences across firms. Ideally, we directly observe physical output and in fact those types of datasets are becoming increasingly available to empirical researchers, making our approach very much suitable to these data. For instance, US census data collects physical output for a set of industries as documented in Foster, Haltiwanger, and Syverson (2008) and Goldberg et al. (2010) and Kugler and Verhoogen (2008), who observe output in Indian and Colombian manufacturing firms, respectively. Alternatively, we need to convert revenues to physical output using price indices. When only (deflated) revenue is observed in the data, however, our approach is still informative about the correlation between markups and firm-level characteristics, such as export status in our application. We discuss the additional assumptions we need to use revenue data, but want to stress that the main approach to get at markups is not affected. We show that when relying on revenue data, only the level of the markup is potentially affected but not the estimate of the correlation between markups and firm-level characteristics or how markups change over time, which is after all the main focus of our application.

By modeling firm-specific productivity we can relax a few important assumptions maintained in previous empirical work. First, we do not need to impose constant returns to scale, and second, our method does not require observing or measuring the user cost of capital. We show that this approach leads to a flexible methodology and reliable estimates. We then use our empirical model to verify whether exporters, on average, charge higher markups than their domestic counterparts in the same industry, and how markups change upon export entry. Our framework is well suited to relate markups to any observed firm-level activity, such as research and development, foreign direct investment, import status, etc., that are potentially correlated with firm-level productivity.

The remainder of this paper is organized as follows. We briefly put our paper in the context of the literature in Section I. In particular, we contrast our methodology to current approaches and preview our empirical application and main results. Section II introduces our empirical framework and our estimation routine. Section III provides a short discussion on the relationship between markups and firm-level export status. In Section IV, we turn to the data, and in Section V we discuss our main results. Section VI provides a few robustness checks and we discuss remaining caveats. The final section concludes.

I. Recovering Markups from Production Data

Robert Hall published a series of papers suggesting a simple way to estimate (industry) markups based on an underlying model of firm behavior (Hall 1986, 1988, 1990). These papers generated an entire literature that was essentially built upon the key insight that industry-specific markups can be uncovered from production data with information on firm- or industry-level usage of inputs and total value of shipments (e.g., Domowitz, Hubbard, and Petersen 1988; Waldmann 1991; Morrison 1992; Norrbin 1993; Roeger 1995; Basu and Fernald 1997; or Klette 1999)[2]. This approach is based on a production function framework and delivers an average markup using the notion that under imperfect competition, input growth is associated with disproportional output growth, as measured by the relevant markup. An estimated markup higher than one would therefore immediately reject the perfect competitive model.[3]

A. *Challenges and Outstanding Problems*

Some important econometric issues are still not addressed in the series of modified approaches, however. The main concern is that unobserved factors can impact output growth as well, and an obvious candidate in the framework of a production function is productivity (growth).[4] Not controlling for unobserved productivity shocks biases the estimate of the markup as productivity is potentially correlated with the input choice. While previous papers relied on the use of instrumental variables or, more recently, generalized method of moments (GMM), we relate our approach to the literature on estimating production functions. Olley and Pakes (1996) and Levinsohn and Petrin (2003) introduced a full behavioral model to solve for unobserved productivity as a function of observed firm-level decisions (investment and input demand) to deal with the endogeneity of inputs when estimating a production function.[5] We refer to this approach as the proxy approach.

The increased availability of firm- or plant-level datasets further boosted empirical studies using some version of the Hall approach on micro data. Dealing adequately with unobserved productivity shocks becomes an ever bigger concern when applying the Hall method to plant-level data given the strong degree of heterogeneity, as the set of instruments suggested in the literature were mostly aggregate demand factors such as military spending and oil prices. Moreover, the Hall methodology and further refinements have become a popular tool to analyze how changes in the operating environment—such as privatization, trade liberalization, and labor market reforms—have impacted market power, measured by the change in markups. Here again, the correlation between the change in competition and productivity potentially biases the estimates of the change in the markup. Let us take the case of trade liberalization. If opening up to trade impacts firm-level productivity, as has been documented

[2] The literature also spread to international trade. See Levinsohn (1993), Harrison (1994), and Konings and Vandebussche (2005).

[3] In the original model, Hall actually tests a joint hypothesis of perfect competition and constant returns to scale. In an extended version, however, a returns to scale parameter is separately identified (Hall 1990). Importantly, our approach does not require any assumptions on the returns to scale in production as opposed to the Roeger (1995) approach.

[4] In addition, there has been quite a long debate in the literature on what the estimated markup exactly captures and how the model can be extended to allow for intermediate inputs and economies of scale among others (see Domowitz, Hubbard, and Petersen 1988 and Morrison 1992).

[5] Various refinements have since been proposed in the literature. Ackerberg et al. (2007) show, however, that the basic framework remains valid. The methodology is now widespread in industrial organization, international trade, development economics (see, e.g., Van Biesebroeck 2005 and De Loecker 2007, who apply modified versions in the context of sorting out the productivity gains upon export entry).

extensively in the literature, it is clear that the change in the markup due to a change in a trade policy is not identified without controlling for the productivity shock.[6]

We introduce the notion of a control function to control for unobserved productivity in the estimation of the output elasticity of a variable input, which, combined with standard first-order conditions on cost minimization, generate estimates of firm-level markups. Our approach provides estimates of markups while controlling for unobserved productivity and relying on clearly spelled out behavioral assumptions. In addition, we identify markups while allowing for flexible production technologies and can accommodate dynamic and/or fixed inputs of production such as capital.

We show that our approach and the Hall (1986) approach are linked in a straightforward way by considering a special case of our model where the markup is constant across producers.[7] We also compare our estimates to those obtained using an alternative suggested by Klette (1999), who relies on dynamic panel estimation techniques. Our approach relaxes a few important assumptions on how productivity shocks enter the model. In particular, we allow for unobserved serially correlated productivity, which is potentially affected by firm-level decisions. In addition, we recover firm- and time-specific markups, as opposed to an average markup for a set of producers, allowing for an analysis of how markups are related to economic variables such as productivity, firm size, and a firm's export status. Finally, we estimate our model in levels as opposed to the current literature where first differences of the production function are considered. We hereby increase the sample size and the efficiency of the estimates considerably, while reducing the role of measurement error.[8]

B. *Markups and Export Status*

In addition to providing a simple empirical framework to estimate markups using standard production data, we provide new results on the relationship between firms' export status and markups using a rich micro dataset where we observe substantial entry into export markets over our sample period. The latest generation of models of international trade with heterogeneous producers (e.g., Melitz 2003) were developed to explain the strong correlation between export status and various firm-level characteristics, such as productivity and size. In particular, the correlation between productivity and export status has been proven to be robust over numerous datasets. The theoretical models, such as Bernard et al. (2003) and Melitz and Ottaviano (2008), emphasize the self-selection of firms into export markets based on an underlying productivity distribution, creating a strong correlation between productivity and export status. These models also have predictions regarding markups and firm-level export status, however, and our empirical framework can be used to verify these predicted correlations between a firm's markup and its export status.

[6] The same is true in the case where we want to estimate the productivity response to a change in the operating environment such as a trade liberalization. See De Loecker (2011) for more on this.

[7] We are not the first to rely on the insight of Hall (1986) and adopt it to plant-level production data. Both Levinsohn (1993) and Harrison (1994) rely on a version of the Hall approach to analyze markups using micro-level production data.

[8] The sample size under first differencing is further reduced when instrumenting with lagged input growth, which requires at least three consecutive years of data for a given producer. The latter has, in addition, the potential of increasing a selection bias by conditioning on firm survival over a three-year period.

Furthermore, we explore the dynamics of export entry and exit to analyze how it impacts markups. The latter will also allow us to shed more light on the often mentioned learning by exporting hypothesis, which refers to significant productivity improvements for exporters upon export entry. This has recently been confirmed for mostly developing countries.[9] Almost all empirical studies that relate firm-level export status to (estimated) productivity, however, rely on revenue to proxy for physical output and therefore do not rule out that part of the export premium captures product quality improvements and market power effects. Related to this, recent studies by Kugler and Verhoogen (2008) and Hallak and Sivadasan (2009) report higher product quality for exporters, whereas Manova and Zhang (2012) report higher export prices for richer and more distant markets using Chinese transaction-level data. They suggest that their results are consistent with a model where firms adjust quality and markups across destinations in response to market toughness. Therefore, differences in pricing behavior between exporters and nonexporters could, at least partially, be responsible for the measured productivity trajectories upon export entry. Our framework is especially well suited to address this question since our method generates firm-level estimates of markups and productivity, while controlling for potentially endogenous productivity improvements as a result of past export participation.

We study the relationship between markups and export status for a rich panel of Slovenian firms over the period 1994–2000. Slovenia is a particularly useful setting for this. First, the economy was a centrally planned region of former Yugoslavia until the country became independent in 1991. A dramatic wave of reforms followed that reshaped market structure in most industries. This implied a significant reorientation of trade flows toward relatively higher-income regions like the European Union (EU) and led to a quadrupling of the number of exporters over a seven-year period (1994–2000). Second, it has become a small open economy that joined the EU in 2004, and its gross domestic product (GDP) per capita is rapidly converging toward the EU average. This opening to trade has triggered a process of exit of the less productive firms, while deregulation and new opportunities facilitated the entry of new firms as well as entry into export markets, which contributed substantially to aggregate productivity growth.[10]

We find that markups differ dramatically between exporters and nonexporters and are both statistically and economically significantly higher for exporting firms. The latter is consistent with the findings of productivity premia for exporters, but at the same time requires a better understanding of what these (revenue-based) productivity differences exactly measure. We provide one important reason for finding higher measured revenue productivity: higher markups. Finally, we find that markups significantly increase for firms entering export markets.

[9] See, e.g., Van Biesebroeck (2005) and De Loecker (2007). The literature also emphasizes the importance of self-selection into export markets (e.g., Clerides, Lach, and Tybout 1998).

[10] See De Loecker and Konings (2006) for more on the importance of entry in aggregate productivity growth in Slovenian manufacturing.

II. A Framework to Estimate Markups

We introduce an empirical model to obtain firm-level markups relying on standard cost minimization conditions for variable inputs free of adjustment costs. These conditions relate the output elasticity of an input to the share of that input's expenditure in total sales and the firm's markup.[11] After we derive this relationship for a general production function, we discuss the estimation of the output elasticities, which together with data on input expenditures and total sales generate estimated markups.

To obtain output elasticities, we need estimates of the production function, for which we rely on proxy methods developed by Olley and Pakes (1996; hereafter, OP), Levinsohn and Petrin (2003; hereafter, LP) and Ackerberg, Caves, and Frazer (2006; hereafter, ACF). We present our empirical framework in this particular order to highlight the flexibility of our approach with respect to the underlying production technology, consumer demand, and market structure. We view the restrictions that we impose, and which we discuss in detail in below, to be mild especially given the state of the literature.

A. *Deriving an Expression for Markups*

A firm i at time t produces output using the following production technology:

$$Q_{it} = Q_{it}(X_{it}^1, \ldots, X_{it}^V, K_{it}, \omega_{it}), \tag{1}$$

where it relies on V variable inputs such as labor, intermediate inputs, and electricity. In addition, a firm relies on a capital stock, K_{it}, which is treated as a dynamic input in production. The only restriction we impose on $Q_{it}(\cdot)$ to derive an expression of the markup is that $Q_{it}(\cdot)$ is continuous and twice differentiable with respect to its arguments.[12]

We now assume that producers active in the market are cost minimizing and we can therefore consider the associated Lagrangian function

$$L(X_{it}^1, \ldots, X_{it}^V, K_{it}, \lambda_{it}) = \sum_{v=1}^{V} P_{it}^{X^v} X_{it}^v + r_{it} K_{it} + \lambda_{it}(Q_{it} - Q_{it}(\cdot)), \tag{2}$$

where $P_{it}^{X^v}$ and r_{it} denote a firm's input price for a variable input v and capital, respectively. The first-order condition for any variable input free of any adjustment costs is

$$\frac{\partial L_{it}}{\partial X_{it}^v} = P_{it}^{X^v} - \lambda_{it} \frac{\partial Q_{it}(\cdot)}{\partial X_{it}^v} = 0, \tag{3}$$

[11] Our approach is similar to Basu and Fernald (2002) and Petrin and Sivadasan (2010).

[12] Note that this expression encompasses both value added and gross output production function. In the former, only labor and capital enter the specification while we assume that intermediate inputs are used in a fixed proportion, purging output from intermediate input use.

where the marginal cost of production at a given level of output is λ_{it} as $\frac{\partial L_{it}}{\partial Q_{it}} = \lambda_{it}$. Rearranging terms and multiplying both sides by $\frac{X_{it}}{Q_{it}}$, generates the following expression:

$$(4) \qquad \frac{\partial Q_{it}(\cdot)}{\partial X_{it}^{v}} \frac{X_{it}^{v}}{Q_{it}} = \frac{1}{\lambda_{it}} \frac{P_{it}^{X^v} X_{it}^{v}}{Q_{it}}.$$

Cost minimization implies that optimal input demand is satisfied when a firm equalizes the output elasticity of any variable input X_{it}^{v} to $\frac{1}{\lambda_{it}} \frac{P_{it}^{X} X_{it}}{Q_{it}}$. It is important to stress that the above conditions on the use of dynamic inputs of production such as capital, and potentially other inputs facing adjustment costs. It is the use of this *conditional cost function* that will allow us to uncover a firm's markup, as cost minimization implies that we can simply condition on the dynamic inputs of production and therefore not have to consider the full dynamic problem of the firm and avoid having to make additional assumptions.[13]

A final step to obtain an expression for the markup μ_{it} is to simply define it as $\mu_{it} \equiv \frac{P_{it}}{\lambda_{it}}$. This expression is robust to various (static) price setting models, and does not depend on any particular form of price competition among firms. The markup will, however, depend on the specific nature of competition among firms. One restriction we do impose on price setting is that prices are set period by period, and hereby rule out dynamics in pricing such as menu pricing or simply costly adjustment of changing prices.[14] It is important to realize that we identify the markup from the difference in price and marginal cost. Markups are determined in equilibrium, however, depending on the specific model of competition and strategic interaction between firms. We briefly discuss some leading cases of price competition in applied industrial organization and international trade in the online Appendix and cast them in our empirical framework.

For our purpose, it is sufficient to define the markup μ_{it} as the price-marginal cost fraction. Using this definition, we can rewrite equation (4) as

$$(5) \qquad \theta_{it}^{X} = \mu_{it} \frac{P_{it}^{X} X_{it}}{P_{it} Q_{it}},$$

where the output elasticity on an input X is denoted by θ_{it}^{X}. This expression will form the basis for our approach: we obtain the output elasticity from the estimation of a

[13] Note that, in the special case where marginal costs are constant across all levels of output, the output elasticity is only then equal to the input's cost share. The constant marginal cost assumption was implicitly introduced in the original Hall article. Under that assumption, the markup could in theory be measured by directly comparing revenue and cost shares.

[14] Our data is at the annual level and at this level of frequency prices are adjusted frequently, and we therefore abstract away from this issue. We refer to Bils and Klenow (2004) who find that half of goods' prices last 5.5 months or less, which implies that prices are adjusted much more at the annual level and reducing the price stickiness at the annual frequency. Although we do not want to stress this too much in our paper, since it is not the focus of the paper, our methodology can in principle deliver an estimate of the markup consistent with dynamic pricing (under adjustment costs due to say menu costs for instance). A different first order condition (FOC) on pricing will be obtained that will imply that the wedge between an input's marginal product and the real input price will not measure the markup as the relevant markup is no longer simply price over marginal cost. Under a specific structure, we can back out both parameters of the model. This lies beyond the scope of this paper.

production function and need only to measure the share of an input's expenditure in total sales. Or put differently, we obtain an expression of the markup as follows:

$$\mu_{it} = \theta_{it}^{X}\left(\alpha_{it}^{X}\right)^{-1}, \tag{6}$$

where α_{it}^{X} is the share of expenditures on input X_{it} in total sales $(P_{it}Q_{it})$. In order to obtain a measure of firm-level markups using production data, we require only an estimate of the output elasticity of one (or more) variable input(s) of production and data on the expenditure share. The latter is directly observed in most micro data. A different way to interpret the last expression is to note that the markup is identified as the ratio of an input's output elasticity and its revenue share, where we can recover an estimate of the output elasticity by estimating the production function.

Although this derivation is standard and has been used throughout the literature, our contribution is to provide consistent estimates of the output elasticities while allowing some inputs to face adjustment costs and recover firm-specific estimates of the markup that we can relate to various economic variables. We also show how our approach relaxes the current literature, which relies on a single-equation approach to estimate industry-level markups, in a few important ways.

It is important to stress that our approach can accommodate inputs with adjustment costs. The most obvious candidate is the firm's capital stock. The wedge between the firm's output elasticity of capital and its revenue share contains the expected stream of costs and revenues and adjustment costs, in addition to the current markup, and we will revisit this implication by comparing markups obtained from both variable inputs and the capital stock.

B. *Estimating Output Elasticities and Markups*

In order to obtain estimates of the output elasticities θ_{it}^{X}, we restrict our attention to production functions with a scalar Hicks-neutral productivity term and with common technology parameters across the set of producers. The latter does not imply that output elasticities of inputs across firms are constant, except for the special case of Cobb-Douglas.

The two restrictions imply the following expression for the production function

$$Q_{it} = F(X_{it}^{1},\ldots,X_{it}^{V},K_{it};\beta)\exp(\omega_{it}), \tag{7}$$

where we highlight that a set of common technology parameters β govern the transformation of inputs to units of output, combined with the firm's productivity ω_{it}.

We view this restriction to be very mild and the expression above contains most, if not all, specifications used in empirical work such as the Cobb-Douglas and the Translog production function.[15] The main advantage of restricting our attention to production technologies of this form is that we can rely on proxy methods suggested by OP, LP, and ACF to obtain consistent estimates of the technology parameters β.

[15] We can relax the technology parameters to be time variant, and have β_t. In our empirical work, we check the importance of this assumption for our results.

From now on, we consider the log version of equation (6) given that the output elasticity of a variable input v, $\theta_{it}^{X^v}$, is given by $\frac{\partial \ln F(\cdot)}{\partial \ln X_{it}^v}$ and is by definition independent of a firm's productivity level.[16] We discuss the details of how we estimate the production function parameters, which we need to compute $\theta_{it}^{X^v}$, in the next section.

Estimation Procedure.—Moving toward the empirical specification of our model, we implicitly allow for measurement error in output observed in the data and for unanticipated shocks to production, which we combine into ϵ_{it}. More precisely, we observe logged output y_{it} and assume that it is given by $y_{it} = \ln Q_{it} + \epsilon_{it}$, where ϵ_{it} are unanticipated shocks to production and i.i.d. shocks including measurement error.[17] It is important to stress that we explicitly rely on the fact that firms do not observe ϵ_{it} when making optimal input decisions. We come back to this distinction when computing markups using our estimates.

The production function we take to the data, and estimate for each industry separately, is therefore given by

$$y_{it} = f(x_{it}, k_{it}; \beta) + \omega_{it} + \epsilon_{it}, \tag{8}$$

where we subsume the constant term in productivity and collect all variable inputs in x_{it}, and β contains all relevant coefficients.[18] We consider flexible approximations to $f(\cdot)$ and therefore explicitly write the production function we estimate on the data in general terms. For instance, our main empirical specification relies on a translog production function that implies that $f(\cdot)$ is approximated by a second-order polynomial where all (logged) inputs, (logged) inputs squared, and interaction terms between all (logged) inputs are included.[19] We recover the Cobb-Douglas (CD) production function when we drop higher-order and interaction terms. The departure from the standard CD production function is important for our purpose. If we were to restrict the output elasticities to be independent of input use intensity when analyzing how markup differs across firms, we would be attributing variation in technology to variation in markups, and potentially bias our results.

Our approach nests various specifications of the production function, such as the value added and gross output production functions. The latter has the potential advantage of providing us with multiple first-order conditions to recover the markup and test for overidentifying restrictions. In order to guarantee identification

[16] We can in principle extend our model to incorporate input-biased technological change where another productivity shock enters the model, which directly affects one particular input of production.

[17] Most firm-level production data will record output as total value of shipments or value added. Therefore revenues have to be converted to physical output measures using price indices. Unobserved price variation that is uncorrelated with input choices will therefore be picked up by ϵ_{it} and our procedure explicitly corrects for this when computing markups. We revisit this measurement problem in more detail in Section VI and discuss the additional assumptions required to rely on deflated revenue data.

[18] See below for a specific case when we introduce a value added production function.

[19] In fact we can approximate $f(\cdot)$ by a higher-order polynomial and make the coefficients time-dependent without affecting our method of moments approach. In practice, however, the search over a very large set of parameters in the GMM setting becomes much more computionally intensive. For instance, when we consider a translog gross output production function in three inputs (labor, materials, and capital) we are already left with ten production functions coefficients over which we need to search jointly. Moving to a higher-order polynomial approximations raises the number of parameters substantially.

of all variable factors of production, however, we need to make explicit that all input prices of variable inputs of production vary across firms and are serially correlated. The latter allows us to rely on lagged input choices to identify the production coefficients. The value added production function relies on an extra assumption that a fixed proportion of materials is used for producing a unit of output. We discuss more details of our estimation procedure for a gross output production function in the online Appendix. We will also revisit this distinction below when discussing adjustment costs in labor demand.

In order to obtain consistent estimates of the production function, we need to control for unobserved productivity shocks, which are potentially correlated with input choices. We deal with this standard simultaneity problem by relying on the insight of OP/LP and use the ACF approach while relying on materials to proxy for productivity. The latter has the advantage of not having to revisit the underlying dynamic model when considering modifications to the original OP setup when dealing with additional state variables.[20] We do, however, describe the estimation routine while relying on a dynamic control, investment, and discuss the additional assumptions we require. In our empirical work we run both procedures on the data.

We follow Levinsohn and Petrin (2003) and rely on material demand,

$$m_{it} = m_t(k_{it}, \omega_{it}, \mathbf{z}_{it}), \tag{9}$$

to proxy for productivity by inverting $m_t(\cdot)$, where we collect additional variables potentially affecting optimal input demand choice in the vector $\mathbf{z}_{it}$. The inclusion of these additional control variables illustrates the only restriction we impose on the underlying model of competition; i.e., we need to include the relevant variables potentially affecting differences in input demand choices of firms. Once those variables are appropriately accounted for in the estimation routine to obtain output elastiticities, we do not have to take a stand on the exact model of competition and can analyze how markups are different across firms and time, and how they relate to firm-level characteristics. The exact variables to be included in $\mathbf{z}_{it}$ depend on the application but will definitely capture variables leading to differences in optimal input demand across firms such as input prices. Anticipating the application of this paper, a firm's export status, for instance, will be included in the control function.[21]

We therefore rely on $\omega_{it} = h_t(m_{it}, k_{it}, \mathbf{z}_{it})$ to proxy for productivity in the production function estimation. The use of a material demand equation to proxy for productivity is important for us. The monotonicity of intermediate inputs in productivity holds under a large class of models of imperfect competition. As long as $\frac{\partial m}{\partial \omega} > 0$ conditional on the firm's capital stock and variables captured by $\mathbf{z}_{it}$, we can use $h_t(m_{it}, k_{it}, \mathbf{z}_{it})$ to proxy for ω_{it} and rely on the latter to index a firm's productivity. This monotonicity is preserved for a wide range of models of imperfect competition. In this setting, we also find it useful to refer to Melitz and Levinsohn (2006) who also rely on intermediate inputs to proxy for unobserved productivity while allowing for

[20] When relying on investment as a proxy, all relevant state variables, both observed and unobserved, have to be incorporated into the control function. We discuss this approach in the online Appendix.

[21] Note that both additional state variables and other demand conditions are required to be included when considering intermediate inputs as a proxy. See De Loecker (2011) for another application where both additional state variables and serially uncorrelated demand factors are included.

imperfect competition. They show that this monotonicity condition holds as long as more productive firms do not set inordinately higher markups than less productive firms.[22] Just as in their setting, we therefore rule out these cases and impose this restriction in our empirical application.[23]

We do depart from Levinsohn and Petrin (2003) and give up on identifying any parameter in the first stage since conditional on a nonparametric function in capital, materials, and other variables affecting input demand, identification of the labor coefficient is not plausible.[24] Note that the latter observation is true even for a CD production function. Given that we are concerned with more flexible production functions and allow for interaction terms between the various inputs, identification of the labor coefficients in the first stage would rely heavily on functional form assumptions.

Our procedure consists of two steps and follows ACF closely. Let us consider a value added translog production function for simplicity, which is given by

$$(10) \qquad y_{it} = \beta_l l_{it} + \beta_k k_{it} + \beta_{ll} l_{it}^2 + \beta_{kk} k_{it}^2 + \beta_{lk} l_{it} k_{it} + \omega_{it} + \epsilon_{it}.$$

In a first stage, we run

$$(11) \qquad y_{it} = \phi_t(l_{it}, k_{it}, m_{it}, \mathbf{z}_{it}) + \epsilon_{it},$$

where we obtain estimates of expected output ($\hat{\phi}_{it}$) and an estimate for ϵ_{it}. Expected output is given by

$$(12) \qquad \phi_{it} = \beta_l l_{it} + \beta_k k_{it} + \beta_{ll} l_{it}^2 + \beta_{kk} k_{it}^2 + \beta_{lk} l_{it} k_{it} + h_t(m_{it}, k_{it}, \mathbf{z}_{it}).$$

Note that under a gross output production function the first stage is identical. Only expression (10) will contain extra terms related to the material input m_{it}, including interaction terms with labor and capital.[25]

The second stage provides estimates for all production function coefficients by relying on the law of motion for productivity.

$$(13) \qquad \omega_{it} = g_t(\omega_{it-1}) + \xi_{it}.$$

We can easily allow for the potential of additional (lagged and observable) decision variables to affect current productivity outcomes (in expectation), in addition to the

[22] Melitz and Levinsohn (2006, p.14) further state that "[i]n this situation, an inordinate markup difference would imply that a productivity increase would lead a firm to increase its markup by such an amount that it would lead to a decrease in the firm's input usage."

[23] For instance, De Loecker (2011) and Aw, Roberts, and Xu (2011) show that under a constant elasticity of substitution (CES) monopolistic competition setup, m_{it} is increasing in productivity. Under models of strategic interaction we require firms with higher productivity not to have disproportionally higher markups, putting restrictions on the markup-productivity elasticity. For the case of Cournot, for example, lower marginal cost (higher productivity) implies a higher use of intermediate inputs, and hence output produced, at any level of residual demand.

[24] See Ackerberg, Caves, and Frazer (2006) and Wooldridge (2009) for a discussion.

[25] To be precise, under a value added specification y_{it} is measured by subtracting material inputs from gross output. Under a gross output specification we include $\beta_m, \beta_{mm}, \beta_{lm}, \beta_{mk}$ and β_{lmk} and their corresponding variables in the specification.

standard inclusion of past productivity. By allowing firm-level decisions such as innovation, export participation, and investment more generally to directly affect a firm's future, we directly accommodate the concerns raised by De Loecker (2010) who discusses the potential problems of restricting the productivity process to be completely exogenous.[26]

After the first stage, we can compute productivity for any value of β, where $\beta = (\beta_l, \beta_k, \beta_{ll}, \beta_{kk}, \beta_{lk})$, using $\omega_{it}(\beta) = \hat{\phi}_{it} - \beta_l l_{it} - \beta_k k_{it} - \beta_{ll} l_{it}^2 - \beta_{kk} k_{it}^2 - \beta_{lk} l_{it} k_{it}$. By nonparametrically regressing $\omega_{it}(\beta)$ on its lag (and potentially a set of variables affecting productivity), $\omega_{it-1}(\beta)$, we recover the innovation to productivity given β, $\xi_{it}(\beta)$.[27]

We can now form moments to obtain our estimates of the production function, where we rely on

$$E\left(\xi_{it}(\beta)\begin{pmatrix} l_{it-1} \\ k_{it} \\ l_{it-1}^2 \\ k_{it}^2 \\ l_{it-1} k_{it} \end{pmatrix}\right) = 0 \tag{14}$$

to estimate the production function parameters and we use standard GMM techniques to obtain the estimates of the production function and rely on block bootstrapping for the standard errors.[28]

The moments above are similar to the ones suggested by ACF and exploit the fact that capital is assumed to be decided a period ahead and therefore should not be correlated with the innovation in productivity. We rely on lagged labor to identify the coefficients on labor since current labor is expected to react to shocks to productivity, and hence $E(l_{it} \xi_{it})$ is expected to be nonzero. In order for lagged labor to be a valid instrument for current labor, however, we require input prices to be correlated over time. We found very strong evidence in favor of this by running various specifications that essentially relate current wages to past wages.[29]

For a gross output production function, we identify the (five) coefficients related to materials in a similar way, where lagged material choices are used as instruments where material input prices are assumed to be serially correlated over time (which is largely supported by the data).

[26] In a similar way we can control for the nonrandom exit of firms by including the propensity to exit P_{it} as in Olley and Pakes (1996); i.e., $g_t(\omega_{it-1}, P_{it})$.

[27] If we want to allow the export status e_{it} to impact expected future productivity, we simply regress it on $(\omega_{it-1}(\beta), e_{it-1})$, and obtain $\xi_{it}(\beta)$. We refer the reader to Section V and the online Appendix for the application to exporting.

[28] Wooldridge (2009) provides a similar procedure where all coefficients are estimated in a one-step system GMM approach that delivers standard GMM standard errors and higher efficiency by relying on cross-equation restrictions. We follow the two-step procedure, however, since we only have to search over five parameters in the second stage, after recovering estimates for ϕ_{it} and ϵ_{it} in the first stage. The Wooldridge (2009) approach is computationally much more demanding since it requires to search jointly over all five parameters and all coefficients of the polynomial functions we use to approximate $h_t(\cdot)$ and $g_t(\cdot)$.

[29] We come back to this point in the online Appendix when we discuss the approach using investment, which requires including wages, and other input prices, in the investment policy function since they are serially correlated.

The estimated output elasticities are computed using the estimated coefficients of the production function. Under a translog value added production function, the output elasticity for labor (L), for instance, is given by

$$\hat{\theta}_{it}^{L} = \hat{\beta}_l + 2\hat{\beta}_{ll} l_{it} + \hat{\beta}_{lk} k_{it}, \tag{15}$$

and under a translog gross output production function we get a similar expression. Most, if not all, current work relies on a CD production, which implies that the output elasticity of labor is simply given by $\hat{\beta}_l$. We now turn to how we compute markups using our estimates and data on firm-level input expenditures and revenues. This will highlight the importance of allowing for heterogeneity in output elasticity across firms and time.

Obtaining Markups from Estimates and Data.—We now have everything in hand to compute markups. Using expression (5) and our estimate of the output elasticity, we can compute markups directly. As mentioned above, however, we do not observe the correct expenditure share for input X_{it} directly since we only observe $\tilde{Q}_{it}$, which is given by $Q_{it} \exp(\epsilon_{it})$. The first stage of our procedure does provide us with an estimate for ϵ_{it} and we use it to compute the expenditure share as follows:

$$\hat{\alpha}_{it}^{X} = \frac{P_{it}^{X} X_{it}}{P_{it} \dfrac{\tilde{Q}_{it}}{\exp(\hat{\epsilon}_{it})}}. \tag{16}$$

This correction is important as it will eliminate any variation in expenditure shares that comes from variation in output not correlated with $\phi_t(l_{it}, k_{it}, m_{it}, \mathbf{z}_{it})$, or put differently from output variation not related to variables impacting input demand including input prices, productivity, technology parameters, and market characteristics, such as the elasticity of demand and income levels.

We obtain an estimate for the markup by simply applying the FOC on input demand for a variable input in production as given by equation (5). Markups for each firm i at each point in time t are obtained while allowing for considerable flexibility in the production function, consumer demand, and competition.

Some Remarks.—Before we turn to our application we want to make four remarks. First, we briefly discuss the gross output production function setting, which potentially allows for multiple variable inputs to compute markups. Second, we summarize how our procedure changes when we were to rely on investment to proxy for productivity. Third, we show how the standard and mostly used specification, the CD production function, is a special case of our estimation routine. Finally, we briefly discuss a special case of our empirical model where markups are constant across producers in an industry, and recover the specifications suggested by Hall (1986) and the subsequent work of Klette (1999).

Gross Output and Adjustment Costs.—We presented our estimation routine under the assumption that labor is a static input into production, which is consistent with the notion that we can learn about markups from the optimal labor demand decisions. If labor is a dynamic input, however, due, for example, to adjustment costs such as hiring and firing costs, our procedure can still produce consistent estimates of the production function. In that case we can rely on current labor to identify the coefficients on labor, just like capital. It does have implications for computing markups: the wedge between a firm's output elasticity of labor and the share of the wage bill in sales will capture an additional component reflecting the adjustment costs.[30] In this case, we can rely on a gross output production function and compute the markups using the output elasticity of materials and its expenditure share. Material inputs are potentially much less prone to adjustment costs, up to inventory management, and in our empirical work we will check the robustness of our results to this. We refer the reader to the online Appendix for a detailed discussion of the estimation of the production function parameters under a gross output production function.

Using Investment to Proxy for Productivity.—In order to rely on the OP version of the ACF estimator and use investment to proxy for productivity, we need to incorporate any additional state variable in the investment policy function and check invertibility. Obvious candidates for additional state variables are serially correlated input prices and a firm's export status. Adding the extra state variables, up to showing monotonicity, has no implications on our ability to identify the coefficients of interests.[31]

Cobb-Douglas Production Function.—The Cobb-Douglas production function is obtained simply by shutting the parameters β_{ll}, β_{kk}, and β_{lk} to zero in equation (7). The rest of the procedure is unchanged. The output elasticity of labor, for instance, simply reduces to β_l and implies a constant elasticity across producers and time. Therefore, all variation in the expenditure share will carry over to the variation in markups across firms. The latter implies that under this restrictive model choice, we can immediately rank firms' markups by ranking their (corrected) expenditure shares. In our empirical work we compare markups under different production technologies.

Special Case: Constant Markup.—We can use our framework to recover the original Hall approach, and to some extent the approach of Klette (1999), by assuming that markups are constant across firms and time, $\mu_{it} = \mu$. Both Hall and Klette make further assumptions on the productivity shocks and let productivity be a fixed effect that is eliminated by first differencing the production function. We want to focus on the constant markup assumption for now.[32] Let us consider a CD production

[30] See Petrin and Sivadasan (2010) for such an application.

[31] The online Appendix provides the details of the estimation routine. We refer to Van Biesebroeck (2005) and De Loecker (2007) for a detailed discussion, and we rely on their results to use investment when considering export as a state variable.

[32] Note that Klette (1999) allows for additional productivity shocks by further instrumenting using a GMM approach. We discuss this more in Section V when we estimate Klette's model on our data. Furthermore, Klette (1999) can in principle allow for markup heterogeneity but cannot directly obtain firm-level markup estimates. Instead the focus was on recovering an average markup while trying to control for the firm-level deviations away from the average markup, as this might bias the coefficient of interest. This approach has the advantage of keeping the underlying production technology less restricted.

function for simplicity. The main estimating equation in the Hall framework is obtained by taking first differences of the production function and directly imposing the first-order conditions from cost minimization on *all* inputs of the production function. The estimating equation then reduces to

$$\Delta y_{it} = \mu \Delta x_{it} + \Delta \tilde{\epsilon}_{it}, \tag{17}$$

where $\Delta y_{it} = y_{it} - y_{it-1}$, $\Delta x_{it} = (\alpha^L \Delta l_{it} + \alpha^K \Delta k_{it})$, and $\Delta \tilde{\epsilon}_{it} = \Delta \epsilon_{it} + \Delta \omega_{it}$.[33] It is worth emphasizing that the constant markup condition can either be imposed by considering a specific model of competition and demand system, or by restricting the goal of the estimation routine to estimate the average markup. Both constraints lead to the same estimating equation, but identification of the parameter μ is obviously different. Equation (15) further highlights that capital is assumed to be a variable input since the static first-order condition is used to substitute the capital coefficient. In addition, we need to measure the user cost of capital (r_{it}), which, as discussed before, requires an additional set of assumptions and introduces additional measurement issues. Variants of this equation have been used extensively in the literature and this paper makes the assumptions required to obtain consistent markup estimates explicit.

We can directly verify the importance of relaxing the assumptions on the productivity shock by relying on our approach. In fact, we obtain an estimate of the markup directly alongside the dynamic input's coefficients as follows. In a first-stage run,

$$y_{it} = \phi_t(l_{it}, m_{it}, k_{it}, \mathbf{z}_{it}) + \epsilon_{it}, \tag{18}$$

where $\phi_t(\cdot) = \mu l_{it}^* + \beta_k k_{it} + h_t(m_{it}, k_{it}, \mathbf{z}_{it})$ and $l_{it}^* = \alpha_{it}^L l_{it}$. We then obtain an estimate of the markup parameter from the (modified) moment $E(\xi_{it} l_{it-1}^*) = 0$.

Finally, it's useful to consider the first difference version of our approach. This specification will allow us to directly verify the importance of relying on our control function on the estimated markup. Compared to equation (17), we obtain

$$\Delta y_{it} = \mu \Delta l_{it}^* + \Delta \tilde{h}_t(m_{it}, k_{it}, \mathbf{z}_{it}) + \Delta \epsilon_{it}. \tag{19}$$

The proxy for productivity has the advantage of not having to treat capital as a static input since we collect all terms on capital and materials in $\tilde{h}_t(\cdot)$, where $\Delta l_{it}^* = \alpha_{it}^L \Delta l_{it}$ and $\Delta \tilde{h}_t(m_{it}, k_{it}, \mathbf{z}_{it}) = \beta_k \Delta k_{it} + h_t(m_{it}, k_{it}, \mathbf{z}_{it}) - h_{t-1}(m_{it-1}, k_{it-1}, \mathbf{z}_{it-1})$. Note that we could, in principle, identify the markup parameter in the first stage by making additional assumptions. We rely on similar moment conditions as discussed extensively to identify μ, however, although efficiency is further sacrificed by requiring lagged differenced inputs as instruments.

In our empirical work we will compare our estimates to those obtained with the Hall/Klette approach where we rely on an adjusted version of the GMM approach described in Klette (1999). Note that the first differencing approach of Hall/Klette has the potential disadvantage of increasing the role of measurement error compared

[33] In general, the revenue shares are firm and time specific. In the case of CD with a constant markup, however, they need to be constant across firms since $\beta_l = \alpha^L \mu$.

to our approach in levels and can lead to a downward bias of the constant markup parameter. We will compare estimates of the level and first difference model in our empirical application.

III. Exporters, Productivity, and Markups

We rely on our empirical framework to analyze how markups differ between exporters and nonexporters. In addition, we are interested in how export entry impacts markups. To answer this, we correlate markups with a firm's export status and check whether markups change with export entry, while controlling for input usage. We further explain our empirical model in detail once we have introduced the data and discuss the information we can rely on. We stress that we want to verify whether exporters charge different markups without taking a stand on any specific model of international trade. When interpreting the estimated markup parameters, however, we can turn to various models to interpret and explain our findings.

A number of models of international trade with heterogeneous producers and firm specific markups have predictions on the relationship between a firm's export status and its productivity level. Most of the empirical work in this literature has focused on the latter, while not much attention has gone to analyzing the relationship between markups and firm-level export behavior. These models generate the result that more productive firms set higher markups, and, given that those firms can afford to pay an export entry cost, therefore predict that exporters will have higher markups. Bernard et al. (2003) rely on a Bertrand pricing game while allowing for firm-level productivity difference and find that on average, exporters have higher markups. Recently, Melitz and Ottaviano (2008) model firms in an international trade setting that compete in prices where products are horizontally differentiated. This model generates a firm-specific markup that is a function of the difference between the firm's marginal cost and the cut-off marginal cost where the firm is indifferent between staying in the industry or exiting. Therefore, when a firm is relatively more productive, it can charge a higher markup and enjoy higher profits. Markups therefore drive a wedge between actual and measured productivity, and disproportionately so for exporting firms.

A wide range of models will predict the aforementioned relationship, which essentially comes from a single source of heterogeneity on the supply side (productivity). Another strand of the trade literature explores the role of quality differences between exporters and nonexporters. If exporters produce higher-quality goods, while relying on higher-quality inputs, all things equal, they can charge higher markups (see Kugler and Verhoogen 2008 and Hallak and Sivadasan 2009 for an empirical analysis). In the industrial organization literature, Foster, Haltiwanger, and Syverson (2008, 2010) also consider two-dimensional firm heterogeneity: productivity and idiosyncratic demand shocks. They show that both dimensions are important to explain firm exit, so that selection can be explained by both productivity and profitability.[34]

[34] We study the reverse relationship; i.e., how entry and exit into exporting are related to a change in the markup. Investigating the link between markup and selection into export markets would require additional assumptions, as we discuss in Section V.

Table 1—Firm Turnover and Exporting in Slovenian Manufacturing

Year	No. firms	Exit rate	Entry rate	No. exporters	Labor productivity
1995	3,820	3.32	13.14	1,738	14.71
1996	4,152	2.60	5.44	1,901	16.45
1997	4,339	3.43	4.47	1,906	18.22
1998	4,447	3.94	4.14	2,003	18.81
1999	4,695	3.26	3.30	2,192	21.02
2000	4,906	2.69	3.38	2,335	21.26

Note: Labor productivity is in thousands of tolars (deflated by industry-specific producer price index).

Both mechanisms are thus expected to generate higher markups for exporters in the cross section. In the time series dimension, however, it is not clear how markups change as firms enter export markets compared to already-exporting firms and domestic producers. We therefore see this paper as providing both a check of current models of international trade generating a relationship between export status and markups, as well as new evidence on markup dynamics and export status. Since most theories are static in nature, they cannot speak to this time dimension. More recently, Cosar, Guner, and Tybout (2009) develop a dynamic general equilibrium trade model to explain certain features of the labor market, and their model implies that exporters charge higher markups because factor market frictions prevent them from freely adjusting their capacity as exporting opportunities come and go over time.

Taking stock of the above, we therefore expect higher markups for exporters. It is clear, however, that markup differences are related to both supply and demand factors impacting both costs and prices. Our procedure delivers both markup and productivity estimates and allows us to further decompose the markup difference between domestic producers and exporters and verify whether, after controlling for differences in marginal costs (i.e., productivity), exporters still have higher markups. In this way, once we have established our main results, we can eliminate the productivity component from the markup difference and provide some suggestive evidence on the role of other factors impacting price. We therefore relate our results to a recent literature that has put forward the importance of these factors, such as differences in elasticities of demand across markets and product quality, for instance.

IV. Background and Data

We rely on a unique dataset covering all firms active in Slovenian manufacturing during the period 1994–2000. The data are provided by the Slovenian Central Statistical Office and contains the full company accounts for an unbalanced panel of 7,915 firms.[35] We also observe market entry and exit, as well as detailed information on firm-level export status and export sales. At every point in time, we know whether the firm is a domestic producer, an export entrant, an export quitter, or a continuing exporter.

[35] We refer to the online Appendix for more details on the Slovenian data, and to De Loecker (2007). In the online Appendix we also list the variables we use in our empirical work and how they are measured. The unit of observation is an establishment (plant) level, but we refer to it as a firm.

Table 1 provides some summary statistics about the industrial dynamics in our sample. While the annual average exit rate is around 3 percent, entry rates are very high, especially at the beginning of the period. This reflects new opportunities that were exploited after transition started.

Our summary statistics show how labor productivity increased dramatically, consistent with the image of a Slovenian economy undergoing successful restructuring. At the same time, the number of exporters grew by 35 percent, taking up a larger share of total manufacturing both in total number of firms, as in total sales and total employment.

We study the relationship between exports and markups since exports have gained dramatic importance in Slovenian manufacturing. We observe a 42 percent increase in total exports of manufacturing products over the sample period 1994–2000. Furthermore, entry and exit has reshaped market structure in most industries. Both the entry of more productive firms and the increased export participation were responsible for significant productivity improvements in aggregate (measured) productivity (De Loecker and Konings 2006 and De Loecker 2007). Therefore, we want to analyze the impact of the increased participation in international markets on the firms' ability to charge prices above marginal cost using our proposed empirical framework.

V. Results

In this section, we use our empirical model to estimate markups for Slovenian manufacturing firms, and test whether exporters have, on average, different markups. In addition, we rely on substantial entry into foreign markets in our data to analyze how markups change with export entry and exit, and as such we are the first, to our knowledge, to provide robust econometric evidence of this relationship.

Applying our method to the case of exporting requires including a firm's export status, and any other factor that impacts optimal input demand, into the control function. To be precise, we include a firm's export status in all input demand equations (as an element of $\mathbf{z}_{it}$), and allow it to directly affect the law of motion of productivity.[36] We refer the reader to the online Appendix for the details of the estimation routine for this application.

After estimating the output elasticity of labor and materials, we can compute the implied markups from the FOCs as described above. We use our markup estimates to discuss several major findings. First, we compare our markup estimates to the literature (Hall 1986 and Klette 1999) and we consider a restricted version of our approach that revisits the Hall/Klette framework but relies on our proxy for productivity. Second, we look at the relationship between markups and firm-level export status in both the cross-section and the time series. Third, we briefly discuss the relationship between markups and other economic variables. This analysis cannot be done using previous methods where a common markup across a set of producers is estimated.[37] Finally, we discuss an important aggregate implication using our results.

[36] In addition, when we consider extensions where markups are allowed to be different across different export destinations, we include destination dummies in the control function as well. One could potentially include other market characteristics but they need to be firm-specific. Otherwise they will be subsumed in the time subscript of $\phi_t(\cdot)$.

[37] An exception is Klette (1999), who estimates the covariance of time-averaged markups and productivity, $cov(\mu_i, \omega_i)$, while relying on additional assumptions. We discuss those in detail and compare it to our framework.

A. *Firm-Level Markups*

We obtain an estimate of each firm's markup and compare the average or median with the Hall/Klette approach. Although our focus is not so much on the exact level of the markup, we want to highlight that the markup estimates are comparable to those obtained with different methodologies, but are different in an important way.

Our procedure generates industry-specific production function coefficients which in turn deliver firm-specific output elasticity of variable inputs. The latter are plugged in the FOC of input demand together with data on input expenditure to compute markups. We list the median markup using a wide set of specifications to highlight our results. We first present results using the standard methods in the literature, using Hall (1986) and Klette (1999). We present our results using both value added and gross output production functions (for value added production functions, we rely on the output elasticity of labor to compute markups and compare them with markups obtained from the output elasticity of materials under a gross output production function), allowing for endogenous productivity processes, under a translog and CD technology. We also consider a specification where we include the export dummy as an input.[38] Finally, we estimate a few restricted versions of our model where we impose a common markup by industry, and take first differences while controlling for productivity using our proxy method.

Empirical Specifications.—More specifically, we run the following specifications for each industry separately: **I**: Value Added under CD; **II**: **I** + endogenous productivity process, where past exporting can impact current productivity as given by $\omega_{it} = g(\omega_{it-1}, e_{it-1}) + \xi_{it}$; **III**: **I** + impose both moments on capital, $E(\xi_{it}(\beta)k_{it-s}) = 0$ for $s = \{0,1\}$, and rely on a weighing matrix in the GMM procedure; **IV**: Value Added under Translog; **V**: **II** and include an export dummy as an additional input. These specifications allow us to see how sensitive the markup estimates are to restricting the output elasticity of any input to be common across firms (under CD) and by assuming a fixed proportion technology (under a value added specification). Moreover, we verify whether relaxing the role of export status in the underlying model of production matters for the markup estimates. In particular, specification **II** allows future productivity to depend on past export behavior directly, and **V** directly allows for exporters to produce under a different technology by including a firm's export status as an input in the production function. In specification **VI** we consider a gross output production function where we can rely on two first-order conditions, labor and materials, to compute markups and compare.

Finally, we also consider two specifications where we directly impose a common markup across producers in an industry. In specification **VII** we consider **I** and impose a constant markup and directly impose the FOC in the production function. More specifically, we obtain the following estimating equation $y_{it} = \mu l^*_{it} + \beta_k k_{it} + \omega_{it} + \epsilon_{it}$, where $l^*_{it} = l_{it}\alpha^L_{it}$. Note that we do not impose the FOC on capital. Relying on our empirical framework and using $h_t(m_{it}, k_{it}, \mathbf{z}_{it})$ to control for productivity we

[38] Some literature has followed this approach to generate the result that exporters produce under different technologies. This specification does not sit well with the CD framework, however, which implies that a firm can substitute any other input for exporting.

Table 2—Estimated Markups

	Methodology	Markup
	Hall[a]	1.03 (0.004)
	Klette[a]	1.12 (0.020)
Specification		
	I (Cobb-Douglas)	1.17
	II (I w/ endog. productivity)	1.10
	III (I w/ additional moments)	1.23
	IV (Translog)	1.28
	V (II w/ export input)	1.23
	VI (Gross Output: labor)	1.26
	VI (Gross Output: materials)	1.22
	VII[a] (**I** w/ single markup)	1.16 (0.006)
	VIII[a] (First difference)	1.11 (0.007)

[a] Markups are estimated jointly with the production function (as discussed in Section III), and we report the standard errors in parentheses. The standard deviation around the markups in specifications **I–VI** is about 0.5.

directly obtain an estimate for the markup.[39] In specification **VIII** we estimate **VII** in first differences, which allows us to directly compare our estimate of the markup to the traditional Hall approach and verify the importance of controlling for unobserved productivity shocks using our proxy approach.

Estimated Markups.—Table 2 presents the median markup of the various specifications. We will exploit the heterogeneity in markups in the next section by relating markups to firm-level characteristics.

Our estimates of the markup are consistently higher compared to the Hall and Klette approach. The markup estimate under Hall is obtained by regressing output growth on an index of input growth where each input is weighted by their expenditure share, and we find a markup of 1.03. In the second row, we estimated a higher markup of 1.12 using Klette's algorithm.[40] Both these models are estimated in first differences, and it is well known to lead to a downward bias of the estimates, here the markup, by exacerbating measurement error.[41]

We obtain markups in the range of 1.17–1.28 and our various specifications give very similar results. Note that the markups obtained using specifications **I–VI** are medians over the underlying distribution, and in all cases the standard deviations are substantial as expected (around 0.5), and indicates a substantial variation in markups across all firms of the manufacturing sector, as expected.[42] We explore the

[39] The steps of the estimation procedure are as before and we obtain an estimate of the markup by relying on the same moments.

[40] Instead of using Arellano and Bond (1991), we use the more efficient method of Arellano and Bover (1995) and Blundell and Bond (1998). Also see Blundell and Bond (2000) for an application to production functions. We only use employment and capital (as in Klette), lagged from $t-2$ onward as instruments (this corresponds to model V in Klette), following the discussion in Section II

[41] In the traditional Hall model, a Taylor expansion of the production function gives rise to estimating the model in first differences. This implicitly restricts the underlying demand system, however, whereby markups do not change between two time periods. Klette (1999) first considers deviations from the median output/input firm before taking first differences in order to eliminate productivity shocks, which are assumed to be a fixed effect.

[42] We recover the distribution of markups for each two-digit manufacturing industry. We do not include those results and focus instead on the difference across various techniques. For example, for the 17 producers of basic

variation across firms in the next section when we relate markups to various economic variables, with a focus on export status.

As mentioned before, our methodology requires the availability of a variable input of production without adjustment costs, in order to rely on the FOC. We compare our markups obtained using cost minimization conditions on the labor input (**I–V**), with markups obtained using materials, **VI**, by running a gross output production function and our results are very similar.[43]

It is worth noting that the markups obtained imposing a static FOC on capital, which clearly goes against the evidence of important adjustment costs in capital, are considerably higher. The latter is as expected since the wedge between the output elasticity of capital and the revenue share contains current markups as well as capital adjustment costs, and should therefore be higher. We find a median markup of around 1.5–1.6 across the various specifications using this approach.

It is interesting to note that when relying on our methodology while imposing a common markup, **VII**, we obtain an estimate of 1.16, which is below our other estimates but still much higher than the standard Hall estimate. This estimate of the markup is obtained directly within our estimation routine by imposing the FOCs on the variable inputs in the production function. This approach is similar to the original Hall approach, except that the regression is estimated in levels and productivity shocks are explicitly controlled for using input demand. To further demonstrate the importance of controlling for unobserved productivity shocks, we consider a first difference version of our approach, **VIII**, while keeping the markup constant and we obtain an estimate of 1.11, which is higher than the standard Hall approach and closer to our preferred estimates.[44] More specifically, comparing the first and the last rows shows the importance of controlling for unobserved productivity shocks when estimating markups. As expected, our level approach, **VII**, leads to a higher estimate of the average markup of 1.16 compared to 1.11 under the (corrected) first difference approach. These restricted versions, **VII–VIII**, of our model highlight the additional assumptions and restrictions of previous approaches in the literature. We run these specifications to highlight the set of assumptions we relax in our approach, and how it impacts the results. In particular, relaxing the constant markup assumption across firms and allowing for time-varying productivity shocks leads to substantially higher markups, ranging up to 12 percent higher.

B. *Markups and Exporting*

We can now turn to the main focus of our application, whether exporters on average have higher markups and whether markups change when firms enter export markets. We first discuss the cross-sectional results, before turning to the time series dimension of our data and verifying whether markups change when firms enter

chemical products, we recover the distribution of markups that lies between 0.95 and 2.5, with a mode at 1.25.

[43] We obtain two separate measures for the markup using the gross output production function. It is feasible to use both estimates to learn about potential frictions in labor demand. This lies beyond the scope of this paper.

[44] We estimate equation (19) and use materials to proxy for productivity and identify the markup in a second stage. Alternatively, when we rely on investment to proxy for productivity, we can estimate the markup in a first stage when relying on additional assumptions as discussed in ACF.

export markets. Finally, we also show how our method allows us to shed light on the correlation of markups and other economic variables such as productivity.

The framework introduced in Section II was not explicit about firms selling in multiple markets. In light of our application we want to stress that our measure of markups for exporters is a share-weighted average markup across the two markets, where the weight by market is the share of an input's expenditure used in production sold in that market. We can correctly compare markups across producers and time without requiring additional information on input allocation across production destined for different markets. To compare markups across markets within a firm, we do require either more data or more theoretical structure to pin down the input allocation by final market. [45]

Do Exporters Have Different Markups?—Given that we have firm-specific markups, we can simply relate a firm's markup to its export status in a regression framework. As noted before, we are not interested in the level of the markup per se, and we therefore estimate the percentage difference in markups between exporters and domestic producers. We do convert these percentages into absolute markup differences in order to compare our results to those obtained using the Hall approach. The specification we take to the data is given by

$$\ln \mu_{it} = \delta_0 + \delta_1 e_{it} + \mathbf{b}'_{it}\sigma + \nu_{it}, \tag{20}$$

where e_{it} is an export dummy and δ_1 measures the percentage markup premium for exporters.[46] We control for labor and capital use in order to capture differences in size and factor intensity, as well as full year-industry interactions to take out industry-specific aggregate trends in markups. We collect all the controls in a vector $\mathbf{b}_{it}$ with σ the corresponding coefficients. We stress that we are not interpreting δ_1 as a causal parameter and we rely on our approach to test whether, on average, exporters have different markups. The latter, to our knowledge, has not been documented and we see this as a first important set of results. We are not interested in the coefficients on the various control variables, but later we will revisit the separate correlations of markups and other economic variables. We estimate this regression at the manufacturing level and include a full interaction of year and industry dummies.[47] Once we have estimated δ_1, we can compute the level markup difference by applying the percentage difference to the constant term, which captures the domestic markup average. We denote this markup difference by μ_E and we compute it by applying $\mu_E = \delta_1 \exp(\delta_0)$ after estimating the relevant parameters. Table 3 presents our results.

[45] Consider the FOC for labor by market, which gives equation (5) for each market s, $\mu^s_{it} = \theta^L_{it}(\rho^s_{it}[w_{it}L_{it}]/[P_{it}Q_{it}]_s)^{-1}$ where ρ^s_{it} is the share of the wage bill used on production sold on market s. Rewriting this expression to $\sum_s \rho^s_{it}\mu^s_{it}$ gives rise to the weighted average markup we rely on in our analysis, and is equal to $\theta^L_{it}(w_{it}L_{it}/P_{it}Q_{it})^{-1}$. We defer a more detailed discussion to Section VIB of the paper.

[46] We consider logged markups since the variation in firm-level markups is quite substantial and therefore rely on ordinary least squares (OLS) to minimize proportional deviations, rather than absolute deviations. We discuss an additional advantage of estimating this relationship in logs in Section VI. Since we rely on estimates of the production function to compute markups, we checked whether the standard errors of the OLS regressions were affected, and hence the statistical significance of the main parameters of interest.

[47] We have also run this by industry and the magnitude varies across the different industries, as expected.

TABLE 3—MARKUPS AND EXPORT STATUS I: CROSS-SECTION

	Methodology	Export Premium
	Hall	0.0155 (0.010)
	Klette	0.0500 (0.090)
Specification		
	I (Cobb-Douglas)	0.1633 (0.017)
	II (I w/ endog. productivity)	0.1608 (0.017)
	IV (Translog)	0.1304 (0.014)
	V (II w/ export input)	0.1829 (0.017)
	VIII (First difference)	0.1263 (0.013)

Notes: Estimates are obtained after running equation (21) where the different specifications refer to the different markup estimates, and we convert the percentage markup difference into levels as discussed above. The standard errors under specifications **I–V** are obtained from a nonlinear combination of the relevant parameter estimates. All regressions include labor, capital, and full year and industry dummies as controls. Standard errors are in parentheses.

We run the regression for the various estimates of the markups as described above. The parameter δ_1 is estimated very precisely in all specifications (**I–V**) and is around 0.078.[48] As expected, all the results relying on a CD technology are very similar because the variation in markups is almost identical across the various specifications.[49] Only the level of the markup differs due to different β_l estimates, which is captured by the constant term. The results using a translog production function, **IV**, rely on firm-specific output elasticities and we get a somewhat lower estimated μ_E of 0.1304. One important message that comes from this table is that no significant markup differences are detected when relying on the Hall or the Klette approach. In order to check whether restricting the markup to be constant across firms is important for this result, we consider a restricted version of our approach (**VIII**). The markup premium is estimated to be 0.1263, which is similar to the results under the more general framework. These results highlight the importance of controlling for unobserved productivity shocks when estimating markups directly.

An important advantage of considering log markups is that our results are unchanged even if all the variable inputs we considered to compute markups are subject to adjustment costs. As long as exporting firms are not more (or less) subject to these adjustment costs, our results are not affected.[50]

These results are consistent with recent models of international trade such as the model of Bernard et al. (2003), where exporters charge, on average, higher markups simply because they are more productive and can therefore undercut their rivals. This prediction is supported by comparing the average markup of exporters to non-exporters in the cross-section. In their model, however, firms of the same productivity will charge the same markup, making productivity differences the only source

[48] We no longer report the results using specification III because our markup estimates are not affected at all by adding lagged capital as an additional instrument when estimating the capital coefficients.

[49] Almost identical because the estimate of ϵ_{it} is potentially different across the various Cobb-Douglas specifications.

[50] We can write the first-order condition with adjustment costs in general as follows, $\theta_{it}^X = \mu_{it}(\alpha_{it}^X)^{-1}(1+\tau_{it}^X)$, where the term $(1+\tau_{it}^X)$ contains the additional wedge between the input's marginal product and the input price coming from the adjustment cost. We thus require $E(\ln(1+\tau_{it}^X)e_{it}) = 0$ in order to obtain consistent estimates of the percentage difference in markups, while controlling for l_{it} and k_{it} which further control for potential differences in adjustment costs related to the size of the firm.

for markup differences. Our procedure generates estimates for both markups and productivity and we can shed light on this by including both. When including both a firm's export status and productivity, the coefficient on export δ_1, expressed in percentages, goes down from 0.076 to 0.021, as expected. Once we control for productivity, we control for differences in marginal cost and the coefficient on export status picks up the variation in average prices between exporters and domestic firms. To see this, note that we are actually running

$$(\ln P_{it} - \ln C_{it}) = \delta_0 + \delta_1 e_{it} + \delta_2 \omega_{it} + \mathbf{b}'_{it}\sigma + \nu_{it}, \tag{21}$$

which shows clearly that δ_1 will measure the average price difference (in percentages) if ω_{it} picks up $\ln C_{it}$ fully. As discussed in Katayama, Lu, and Tybout (2009) and De Loecker (2011), we know that ω_{it} potentially picks up price differences and therefore we expect δ_2 to pick up additional variation across producers related to market power and demand conditions. An important point to take away from this is that the export effect is still present even after controlling for productivity differences. In fact, the export dummy still explains around 30 percent of the markup difference, while controlling for productivity. The latter implies that other factors, which are reflected in price differences, play an important role in explaining markup differences between exporters and domestic producers. Our results are therefore consistent with a recent literature emphasizing differences in product and input quality between exporters and domestic producers. Simple differences in demand elasticities and income across markets can equally explain price differences, however. Given our data constraints, we cannot further discriminate between those various mechanisms.

Taking stock of the results described above has potentially important policy implications. The well-documented productivity premium of exporters could, at least partly, be reflecting markup differences. Recent models of international trade with heterogeneous firms emphasize the reallocation of market share from less efficient producers to more efficient exporters. This mechanism relies on exporters being more productive, because they can cover the fixed cost of entering foreign markets. A growing list of empirical studies has documented (measured) productivity premia for exporters, and furthermore recent work has found evidence on further improvements in (measured) productivity post–export entry (learning by exporting). Our results, however, require a more cautious interpretation of the exporter productivity premium and how exporting contributes to aggregate productivity growth. More specifically, given that measured productivity is a residual of a sales-generating production function, it is well known that it contains unobserved quality differences in both inputs and output, as well as market power effects, broadly defined.[51] Our results therefore provide additional information in explaining the measured productivity premium, and emphasize the importance of studying the export-productivity relationship jointly with market power in an integrated framework. We further investigate the markup trajectory as a function of export status in the next section.

[51] In fact, the markup differences between exporters and domestic producers only fully reflect cost (productivity) differences if both domestic producers and exporters set the same output prices.

Export Entry and Markup Dynamics.—So far, we have just estimated differences in average markups for exporters and domestic producers. Our dataset also allows us to test whether markups differ significantly *within the group* of exporters. It is especially of interest to see whether there is a specific pattern of markups for firms that enter export markets; i.e., before and after they become exporters. This will help us to better interpret the results from a large body of empirical work documenting productivity gains for new exporters. These results are used to confirm theories of self-selection of more productive firms into export markets as in Melitz (2003) or learning by exporting. We now turn our attention to the various categories of exporters that we are able to identify in our sample: starters, quitters, and firms that export throughout the sample period.

We run the following regressions on the data where we simply compare markups before and after export entry (and exit), while also estimating the markup differential for firms who continuously export in our sample.[52]

$$(22) \qquad \ln \mu_{it} = \gamma_0 + \gamma_1 Entry_{it} + \gamma_2 Exit_{it} + \gamma_3 Always_i + \mathbf{b}_{it}'\sigma + \nu_{it},$$

where $Entry_{it} = 1$ if a firm becomes an exporter and zero otherwise, and $Exit_{it} = 1$ if a firm stops exporting.[53] The constant term captures the average log markup for domestic producers, including firms that become export entrants or already stopped serving export markets. The interest lies in the coefficient γ_1, which measures the markup percentage difference, for starters, between the post– and pre–export entry periods. The other coefficient γ_2 measures a similar effect but for export exit. Finally, γ_3 measures the markup difference for firms exporting throughout, and we expect this coefficient to be positive. There is little guidance from theory on the coefficient γ_1, given that almost all models are static in nature as discussed before. We therefore see our results as providing new evidence on markup dynamics and export status.

We compute the implied markup-level effects from export entry as before, $\mu_{st} = \gamma_1 \exp(\gamma_0)$, and report them for our various specifications in Table 4.[54]

We find that export entry is associated with substantially higher markups, ranging around 4 percent while controlling for aggregate markup changes. The other coefficients are also as expected. Interestingly, we can include productivity (as before) and still find a significant positive effect for export entry. The latter suggests again that price changes are associated with export entry, which can come from differences in demand conditions (elasticities, etc.) and quality differences, as discussed before. Table 4 lists both the percentage and the level estimates, and our estimates suggest that export entry is associated with a significant increase in markups of around 4 to 5 percent, or between 0.079 and 0.099 in levels. We compare our results to the restricted common markup model in a first difference setting and we obtain a similar export entry effect of 0.07 in the level of the markup. The estimates across the various rows demonstrate that our results are robust with respect to various production technologies and assumptions on the underlying productivity process.

[52] We eliminate the very small fraction of firms that enters or exits export markets more than once in our sample.

[53] Note that this specification estimates an average markup for domestic firms including firms that eventually become exporters, or those who exported in the past. We considered different averages for before and after export entry/exit and the results are similar.

[54] The Table A1 in the online Appendix lists all estimated coefficients. We focus only on the export entry effect.

Table 4—Markups and Export Status II: Export Entry Effect

Method output elasticity	Export entry effect	
	Percentage (γ_1)	Level (μ_{st})
I (Cobb-Douglas)	0.0467	0.0939
	(0.0127)	(0.0260)
II (I w/ endog. productivity)	0.0467	0.0925
	(0.0127)	(0.0250)
IV (Translog)	0.0481	0.0797
	(0.0128)	(0.021)
V (II w/ export input)	0.0497	0.0994
	(0.0127)	(0.0260)
VIII (First difference)	NA	0.0700
		(0.022)

Notes: The standard errors under **I**–**V** are obtained from a nonlinear combination of the relevant parameter estimates. We drop the estimates from specifications **III** and **VI** since they are identical to the ones reported in this table. The latter is as expected since the estimate of the capital coefficient does not impact the markup estimates for instance. Specification **VIII** delivers an immediate estimate of the level impact on markups. All regressions include labor, capital, and full year and industry fixed effects as controls.

When relying on the same regression framework and allowing the markup effect to depend on export intensity, by interacting the export dummies with the share of export sales in total sales, the coefficient on the export entry effect is larger, 0.097, and allows us to compute the export entry markup trajectory as obtained by tracing the share of export sales in total sales over time.

It is important to note that we do not find the markup-export relationships when relying on standard methods. When we rely on our approach, we find significantly higher markups for exporters in the cross-section, and find that markups increase with export entry.

Interpreting Our Results.—In sum, we report two major findings: (i) in the cross-section we find that exporters have higher markups than their domestic counterparts in the same industry, and (ii) in the time series we find that markups increase when firms enter export markets, while controlling for aggregate demand and supply effects through year dummies. How can we explain our results?

A few recent models (Bernard et al. 2003; Melitz and Ottaviano 2008) provide a theoretical analysis of the relationship between firm export status and (market-specific) markups. Under various hypotheses regarding the nature of competition, more efficient producers are more likely to have more efficient rivals, to charge lower prices, to sell more on the domestic market, and to beat rivals on export markets. They benefit from a cost advantage over their competitors, set higher markups (under certain conditions regarding the relative efficiency between firms on the domestic and the export market, in the case of the Melitz and Ottaviano model), and have higher levels of measured productivity. An alternative explanation could be that the elasticity of demand is different on the export market, or that consumers have different valuation for the good. The exact mechanism underlying these results is not testable given the data at hand. For instance, we do not have firm-specific information on prices that could allow us to separate out the markup difference into a cost and price effect. We did show that controlling for cost differences, exporters on average still have higher markups, which suggests additional factors impacting prices are important, and is consistent with recent work by Manova and Zhang (2012) and Hallak and Sivadasan (2009).

Finally, at a broader level, our evidence suggests that the gap between the notion of (physical) productivity in theoretical models of international trade with heterogeneous producers and the empirical measurement of productivity is an important one given that markups are different for exporters and that they change significantly, both economically and statistically, when firms enter export markets.

C. *Markups and Other Economic Variables*

Although not the focus of our analysis, we further rely on our estimates of firm-level markups and relate them to other economic variables of interest, such as productivity. Our procedure generates estimates for both markups and productivity. To be precise, after we have estimated the production function coefficients, we directly obtain an estimate for productivity from

$$\hat{\omega}_{it} = \hat{\phi}_{it} - f(x_{it}, k_{it}; \hat{\beta}), \tag{23}$$

where $f(x_{it}, k_{it}; \hat{\beta})$ is the predicted output using variable inputs and capital using the estimated coefficients $\hat{\beta}$.

A large class of models in industrial organization predict that firms with lower marginal cost (higher productivity) will be able to charge higher markups, all things the same. For example, in a model of Cournot competition, more productive firms will have a higher market share and hence have higher markups. Recent models of international trade with heterogeneous firms also predict this positive correlation. We run the same regression as before and replace the export status by productivity. We obtain a highly significant and positive estimate of 0.3 for the coefficient on productivity. Our results are therefore consistent with a wide range of theory models, and confirm that more productive firms have higher markups. We do not pursue any further analysis given that productivity measures potentially contain price/demand variation as well, and might be poor measures of marginal cost as discussed by Katayama, Lu, and Tybout (2009) and De Loecker (2011).

Our framework could potentially shed light on the separate role of productivity and markups in export entry/exit behavior. We see this as an important next step in this research program, but this lies beyond the scope of this paper.[55]

[55] For completeness, we do like to mention that both markups and productivity enter highly significantly in a set of export entry and exit regressions (using probit analysis) while controlling for industry and year effects. This is at least suggestive of the separate roles both variables play in shaping export entry behavior.

D. *Aggregate Implications*

The Hall framework was initially set out to obtain estimates for productivity growth while appropriately controlling for imperfect competition. We briefly revisit this by considering the Hall version of our framework and use it to back out estimates for productivity growth after estimating markups. Note that our methodology generates estimates for productivity and markups for each firm. We could compute productivity growth directly after estimating the production function. Here we revisit the literature using a restricted version of our model to highlight the importance of correctly estimating markups. We rely on our estimates of the markup $\hat{\mu}$ and the capital coefficient β_k to compute productivity growth as follows:

$$\Delta y_{it} - \hat{\mu}\Delta \tilde{x}_{it} - \hat{\beta}_k \Delta k_{it} = \Delta \omega_{it}. \tag{24}$$

In addition to a different estimate for the markup, as presented in Table 2, our approach does not impose any restrictions on returns to scale. It is clear that using standard techniques will lead to biased estimates for productivity growth since they are based on downward-biased markup estimates. Within the context of sorting out markup differences between exporters and domestic producers, the uncorrected approach would actually predict no differences in productivity growth, conditional on input use, between the two, which is clearly in contradiction with empirical evidence.

It is clear that productivity growth is overestimated without controlling for the endogeneity of inputs and markup differences. This bias further increases when we allow for markups to change when firms switch export status. Although our method is not intended to provide estimates for productivity growth directly, we see this as an important cross-validation of the estimated markup parameters. Our estimates suggest average annual productivity growth rates for Slovenian manufacturing between 3 and 1.5 percent.

Our results have some important implications for aggregate productivity. It is immediately clear that when relying on the standard framework, markups are underestimated for domestic producers and even more so for exporters. It first of all implies that we will overestimate aggregate manufacturing productivity growth, which is obtained by a weighted average of firm-level productivity growth, even when ignoring differences in markups between exporters and domestic producers. When analyzing productivity growth of sectors or countries during a period where export participation increased substantially, however, an additional bias kicks in. Based on our estimates, it is straightforward to show how aggregate productivity growth is overestimated when not controlling for different markups across domestic producers and exporters. In the case of Slovenia, the bias in aggregate productivity growth becomes larger as resources were reallocated toward exporters and therefore accounting for a growing share in aggregate output as the number of exporters quadrupled and export sales grew substantially. These results therefore suggest that the estimated aggregate productivity gains from increased export participation are biased upward when ignoring that exporters charge, on average, higher markups. The wedge between *measured* and *actual* aggregate productivity growth increases as a larger share of manufacturing firms are becoming exporters and are accounting for a larger share of total output.

VI. Robustness and Final Remarks

We discuss two robustness checks below. In turn, we discuss the use of deflated sales to proxy for output and we discuss differences in markups for exporters in the foreign and the domestic market.

A. *Unobserved Prices and Revenue Data*

Implicitly we have treated deflated sales as a measure of physical quantity when estimating output elasticities, and therefore our approach is potentially subject to the omitted price variable bias discussed in Klette and Griliches (1996). In our context, however, we are not concerned with obtaining correct productivity estimates. As discussed by De Loecker (2011), not controlling for unobserved prices is particularly problematic for obtaining reliable estimates for productivity. In our setting, unobserved prices are expected, if anything, to bias the output elasticities downward. The correlation between inputs and prices is expected to be negative, as mentioned in the original work by Klette and Griliches (1996), under quite general demand and cost specifications; i.e., all things equal, more inputs will lead to higher output and push prices down. This implies that, if anything, we are underestimating markups. Unobserved prices will only affect our estimates of the *level* of the markup, however, and will not impact our results on the relationship of markups and export status. We do correct markups from the bias coming from price variation correlated with variables in our proxy $\big(h(\cdot)\big)$.

The use of the proxy for productivity does help against not observing prices as well. Price variation that is correlated with variation in productivity will be controlled for and will therefore not bias the estimates of the production function. Price variation due to demand shocks not correlated with $\phi_t(\cdot)$ can still bias the estimates of the input coefficients, however. The latter will potentially bias the output elasticity estimates but will not impact our main results because in all of our empirical work we correlate log markups to export status. Given our framework, this implies that we ran

$$\big(\ln \theta_{it}^X - \ln \alpha_{it}^X\big) = \theta_0 + \theta_1 e_{it} + \nu_{it} \tag{25}$$

on the data. Under a CD technology, the output elasticity θ_{it}^X reduces to a constant, β_l in the case of using labor, and therefore the bias induced by unobserved prices impacts only the estimate of the constant term θ_0. In other words, we obtain the correct percentage difference in markups between exporters and domestic producers, and if anything underestimate the difference in levels. When considering a more flexible production technology, like the translog, we face a trade-off between allowing for variation in output elasticities and potentially introducing a bias through unobserved prices. Our estimates of the average percentage difference in markups are consistent as long as the difference $\big(\ln \hat{\theta}_{it}^X - \ln \theta_{it}^X\big)$ is not correlated with the firm's export status e_{it}, controlling for differences in input use. When relying on a translog production function, we always include inputs as control in the markup regressions.[56]

The estimated percentage differences presented in the online Appendix show that the results using CD (**I**,**II**,**V**) and Translog (**IV**) are very similar, and we see those in support of the fact that unobserved prices are not impacting our main estimates. The estimated markup-level differences are somewhat lower under the translog production function. This is consistent with a potential downward bias in the production function coefficients, which leads to a lower average output elasticity and hence a lower θ_0 used to compute markup levels.[57] Variation in output elasticities, however, also impacts the point estimate of the constant term.

B. *Exporting and Markups: Digging Deeper*

We documented that exporters have, on average, higher markups, and that markups increase after export entry. Exporters sell products on different markets, however, and our estimate of the markup contains different market-specific markups. We rely on firm-specific export destination information and check whether we can detect differences in markups across destination markets. We revisit the effect of export entry on markups and include the intensity of exporting to shed light on the separate effect of export entry on domestic and foreign markups.

Export Destinations and Markups.—We rely on firm-level export destination information to check whether markups are different across various export destination markets.[58] For the case of Slovenia, exporting includes shipping products to regions formerly part of the Yugoslavian Republic prior to Slovenia's independence in 1991, as well as high-income regions such as the United States and western Europe.

As mentioned above, recent work has documented that exporters produce and ship higher-quality products while controlling for a host of firm-level characteristics including size, where quality is measured indirectly by either unit prices or whether a firm has an International Organization for Standardization 9000 certification.[59] In order to see whether markups are higher for exporters sending their products to high-income regions such as western Europe, we simply include interaction terms with

[56] It is easy to show that $\hat{\theta}^X_{it} = \theta^X_{it} + \rho(l_{it}, k_{it})$, where $\rho(\cdot)$ is equal to the bias in the output elasticity $(\hat{\theta}_{it} - \theta_{it})$, when inputs are correlated with unobserved firm-level price deviations away from the price index. Working through this case suggests running the markup regression and including $\rho(l_{it}, k_{it})$ which will pick up the potentially biased coefficients of the production function. We follow this strategy throughout all our analysis.

[57] If unobserved prices are correlated negatively with inputs, all production function coefficients estimates $\hat{\beta}$ are biased downward. This in turn implies that the estimated output elasticities $\hat{\theta}^X_{it}$ and hence the markups $\hat{\mu}_{it}$ are downward-biased as well. Consequently, the (log) average of the markups are estimated lower, and result in lower estimates of the constant term. The table in the online Appendix demonstrates this potential effect.

[58] As mentioned in De Loecker (2007), the destination information is not available at each point in time in our sample. We therefore return to our cross-sectional comparison of exporters and domestic producers. In addition, we face the trade-off of including the destination dummies in the control function to control appropriately for input demand differences, hereby reducing the sample over which we can estimate the output elasticities. We experimented with relying on both the restricted and entire sample and found no differences in the markup differences across markets.

[59] For instance, Kugler and Verhoogen (2008) document this for Colombia, and Hallak and Sivadasan (2009) provide evidence for manufacturing establishments in India, the United States, Chile, and Colombia.

the various export destination regions to the estimating equation (20). We obtain a 0.045 higher markup (in levels) for firms exporting to western Europe, but estimated less precise than expected given the remaining degree of heterogeneity within the region of western Europe. This implies that exporters shipping to this region, on average, charge a higher markup compared to the average exporter shipping to other regions. Our results are consistent with the quality hypothesis, given that it is expected that quality standards are higher in western European markets than in the Slovenian domestic market. Given the data constraints, we cannot measure quality at the firm level and therefore leave this for future research.

Decomposing Export Entry Markup Effect.—So far, we have shown that markups increase when firms enter export markets. For exporting firms, however, we rely on a markup across the domestic and foreign market. In principle, our methodology can generate markup estimates by market. Applying the first-order condition of labor by market s, where $s = \{d(Domestic), e(Export)\}$, we can compute the markup as before. In our data, however, we do not observe hours worked or number of employees used in production by destination market. We observe only total number of workers in production and this is a standard restriction in plant-level data. Using equation (6), and relying explicitly on the assumption that an exporting plant produces with a given technology in a given location where it faces a given wage rate, implies that we can write

$$\mu_{it}^{s} = \theta_{it}^{L}\left(\frac{\rho_{it}^{s}[w_{it}L_{it}]}{[P_{it}Q_{it}]_{s}}\right)^{-1}, \tag{26}$$

where ρ_{it}^{s} measures the share of the wage bill used in production sold in market s. Total export sales, $[P_{it}Q_{it}]_s$, and the total wage bill are directly observed in our data. Therefore, in order to compute the domestic markup for an exporter and compare it with the average markup across all destination markets, we can compare ρ_{it}^{e} to ρ_{it}^{d} by plant. We adopt the following strategy to verify whether the domestic markup of export entrants changes with export entry. We run the same procedure as in equation (22), but we rely on the share of export sales in total sales, and interact this with the $Entry_{it}$ dummy. This specification allows us to inspect whether the increase in the firm's average markup (across domestic and foreign markets) due to export entry depends on the intensity of exporting. We can look at firms with a very small fraction of sales coming from exporting, say less than 1 percent, when they enter the export market, which can be informative about what happens to their domestic markup. We obtain a significant coefficient of 0.097 for γ_1 and this implies a level estimate of 0.16, which is substantially higher than the estimates reported before. To get the total effect of export entry, however, we need to multiply this estimate with the relevant export share ρ_{it}^{s}, and this implies that the markup entry effect is very small for firms selling a small share of their production abroad. For exporters selling less than 1 percent on foreign markets, markups only increase with 0.001 percent, suggesting that domestic markups do not change. This approach is clearly not without problems as the export share increases over time and the separation between domestic and export markups becomes harder to make. In addition, this approach does not necessarily use the optimal weight, which will depend on how we aggregate inputs across production by destination within a firm. The export sales weight assumes implicitly that inputs are used in proportion to final sales. The latter is an assumption maintained throughout most empirical work (see Foster, Haltiwanger and Syverson 2008, for example). Given the data constraints, we leave the discussion of the optimal weight for future research.

Finally, we want to stress that our methodology can, in principle, deliver markup estimates by market for each firm. The data at hand might restrict the analysis, however. Input use is often not broken down by the final market on which products are sold. Even in this case, our approach is informative about the markup differences between exporters and domestic producers, and whether export entry is related to a change in markups. Observing only total input expenditures at the firm level does restrict our ability to compare markups across markets within a firm without making additional assumptions on how inputs are allocated. In fact, when we rely explicitly on the share-weighted average markup expression, we can write the change in a markup before and after export entry as follows:

$$\Delta \mu_{it} = \Delta\left(\rho_{it}^{d} \mu_{it}^{d}\right) + \rho_{it}^{e} \mu_{it}^{e}. \tag{27}$$

Our results indicate that, for export entrants, the effect is on average positive, and estimated about 7 percent. We can rewrite the change in the average markup using the fact that at $t - 1$ export entrants only sold on the domestic market, or $\rho_{it-1}^{d} = 1$, as

$$\rho_{it}^{d} \mu_{it}^{d} + \rho_{it}^{e} \mu_{it}^{e} > \mu_{it-1}^{d}. \tag{28}$$

Using this decomposition, our results suggest that for firms with very small export sales, markups do not change, suggesting that the domestic markups are unaffected, as it is safe to assume that the input cost share ρ_{it}^{e} will be small as well. In order to obtain market specific estimates of markups by firm, we could introduce a specific demand system for each market, coupled with an assumption on the cost function. Note that our approach is based specifically on not having to specify these at all. We can still compare markups across producers, and how markups change over time with export entry without decomposing how market-specific markups are different across markets within a firm.

VII. Conclusion

This paper investigates the link between markups and exporting behavior. In order to analyze this relationship, we propose a simple and flexible methodology to estimate markups building on the seminal paper by Hall (1986) and the work by Olley and Pakes (1996). The advantages of our method are that we can accommodate a large class of price-setting models while recovering firm-specific markups and do not need to rely on the assumption of constant returns to scale and measuring the user cost of capital.

We use data on Slovenia to test whether (i) exporters, on average, charge higher markups, and (ii) whether markups change for firms entering and exiting export markets. Slovenia is a particularly interesting emerging economy to study as it has transformed successfully from a socially planned economy to a market economy in less than a decade, reaching a level of GDP per capita over 65 percent of the EU average by the year 2000. More specifically, the sample period that we consider is characterized by considerable productivity growth and relatively high turnover. Our methodology is therefore expected to find significantly different markups as we explicitly control for unobserved productivity shocks. Our results confirm the importance of these controls.

Our method delivers higher estimates of firm-level markups compared to standard techniques that cannot control directly for unobserved productivity shocks. Our estimates are robust to various price-setting models and specifications of the production function. We find that markups differ dramatically between exporters and nonexporters, and find significant and robust higher markups for exporting firms. The latter is consistent with the findings of productivity premium for exporters, but at the same time requires a better understanding of what these (revenue-based) productivity differences measure exactly. We provide one important reason for finding higher measured revenue productivity: higher markups. Furthermore, we provide new econometric evidence that markups increase when firms enter export markets.

Our evidence suggests that the gap between the notion of (physical) productivity in theoretical models of international trade with heterogeneous producers and the empirical measurement of productivity is an important one; i.e., markups are different for exporters and change significantly, both economically and statistically, when firms enter export markets. We see these results as a first step in opening up the productivity-export black box, and provide a potential explanation for the big measured productivity gains that go hand in hand with becoming an exporter.

REFERENCES

Ackerberg, Dan, Lanier Benkard, Steve Berry, and Ariel Pakes. 2007. "Econometric Tools for Analyzing Market Outcomes." In *Handbook of Econometrics. Volume 6*, edited by James Heckman and Edward Leamer, 4173–4276. Amsterdam: North Holland.

Ackerberg, Dan, K. Caves, and Garth Frazer. 2006. "Structural Identification of Production Functions." Unpublished.

Arellano, Manuel, and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies* 58 (2): 277–97.

Arellano, Manuel, and Olympia Bover. 1995. "Another Look at the Instrumental Variable Estimation of Error-Components Models." *Journal of Econometrics* 68 (1): 29–51.

Aw, Bee Yan, Mark Roberts, and Yi Xu. 2011. "R&D Investment, Exporting, and Productivity Dynamics." *American Economic Review* 101 (4): 1312–44.

Basu, Susanto, and John G. Fernald. 1997. "Returns to Scale in U.S. Production: Estimates and Implications." *Journal of Political Economy* 105 (2): 249–83.

Basu, Susanto, and John G. Fernald. 2002. "Aggregate Productivity and Aggregate Technology." *European Economic Review* 46 (6): 963–91.

Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum. 2003. "Plants and Productivity in International Trade." *American Economic Review* 93 (4): 1268–90.

Berry, Steve, James A. Levinsohn, and Ariel Pakes. 2004. "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market." *Journal of Political Economy* 112 (1): 68–105.

Bils, Mark, and Peter J. Klenow. 2004. "Some Evidence on the Importance of Sticky Prices." *Journal of Political Economy* 112 (5): 947–85.

Blundell, Richard, and Stephen Bond. 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel-Data Models." *Journal of Econometrics* 87 (1): 115–43.

Blundell, Richard, and Stephen Bond. 2000. "GMM Estimation with Persistent Panel Data: An Application to Production Functions." *Econometric Review* 19 (3): 321–40.

Clerides, Sofronis, Saul Lach, and James Tybout. 1998. "Is 'Learning-by-Exporting' Important? Micro-Dynamic Evidence from Colombia, Mexico, and Morocco." *Quarterly Journal of Economics* 454 (3): 903–47.

Cosar, Kerem, Nezih Guner, and James Tybout. 2009. "Firm Dynamics, Job Turnover, and Wage Distributions in an Open Economy." Unpublished.

De Loecker, Jan. 2007. "Do Exports Generate Higher Productivity? Evidence from Slovenia." *Journal of International Economics* 73 (1): 69–98.

D Loecker, Jan. 2010. "A Note on Detecting Learning by Exporting." National Bureau of Economic Research Working Paper 16548.

De Loecker, Jan. 2011. "Product Differentiation, Multi-Product Firms and Estimating the Impact of Trade Liberalization on Productivity." *Econometrica* 79 (5): 1407–51.

De Loecker, Jan, and Jozef Konings. 2006. "Job Reallocation and Productivity Growth in Post-Socialist Economy: Evidence from Slovenian Manufacturing." *European Journal of Political Economy* 22 (2006): 388–408.

De Loecker, Jan, and Frederic Warzynski. 2012. "Markups and Firm-Level Export Status: Dataset." *American Economic Review*. http://dx.doi.org/10.1257/aer.102.6.2437.

Domowitz, Ian, R. Glenn Hubbard, and Bruce C. Petersen. 1988. "Market Structure and Cyclical Fluctuations in U.S. Manufacturing." *Review of Economics and Statistics* 70 (1): 55–66.

Foster, Lucia S., John Haltiwanger, and Chad Syverson. 2008. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review* 98 (1): 394–425.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2010."The Slow Growth of New Plants: Learning about Demand." Unpublished.

Goldberg, Pinelopi K. 1995. "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry." *Econometrica* 63 (4): 891–951.

Goldberg, Pinelopi K., Amit K. Khandelwal, Nina Pavcnik, and Petia Topalova. 2010. "Imported Intermediate Inputs and Domestic Product Growth: Evidence from India." *Quarterly Journal of Economics* 125 (4): 1727–67.

Hall, Robert E. 1986. "Market Structure and Macroeconomic Fluctuations." *Brookings Papers on Economic Activity* 2 (1986): 285–322.

Hall, Robert E. 1988. "The Relation between Price and Marginal Cost in U.S. Industry." *Journal of Political Economy* 96 (5): 921–47.

Hall, Robert E. 1990. "Invariance Properties of Solow's Productivity Residual." In *Growth/Productivity/Unemployment: Essays to Celebrate Bob Solow's Birthday*, edited by Peter Diamond, 71–112. Cambridge, MA: MIT Press.

Hallak, Juan Carlos, and Jagadeesh Sivadasan. 2009. "Exporting Behavior under Quality Constraints." National Bureau of Economic Research Working Paper 14928.

Harrison, Ann E. 1994. "Productivity, Imperfect Competition, and Trade Reform: Theory and Evidence." *Journal of International Economics* 36 (1–2): 53–73.

Katayama, Hajime, Shihua Lu, and James Tybout. 2009. "Why Plant-Level Productivity Studies Are Often Misleading, and an Alternative Approach to Inference." *International Journal of Industrial Organization* 27 (3): 403–13.

Klette, Tor J. 1999. "Market Power, Scale Economies and Productivity: Estimates from a Panel of Establishment Data." *Journal of Industrial Economics* 47 (4): 451–76.

Klette, Tor J., and Zvi Griliches. 1996. "The Inconsistency of Common Scale Estimators when Output Prices Are Unobserved and Endogenous." *Journal of Applied Econometrics* 11 (4): 343–61.

Konings, Jozef, and Hylke Vandenbussche. 2005. "Antidumping Protection and Markups of Domestic Firms." *Journal of International Economics* 65 (1): 151–65.

Kugler, Maurice, and Eric Verhoogen. 2008. "The Quality-Complementarity Hypothesis: Theory and Evidence from Colombia." National Bureau of Economic Research Working Paper 14418.

Levinsohn, James A. 1993. "Testing the Imports-as-Market-Discipline Hypothesis." *Journal of International Economics* 35 (1–2): 1–22.

Levinsohn, James A., and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *Review of Economic Studies* 70 (2): 317–40.

Manova, Kalina, and Zhiwei Zhang. 2012. "Export Prices and Heterogeneous Firm Models." *Quarterly Journal of Economics* 127 (1): 379–436.

Melitz, Marc. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71 (6): 1695–1725.

Melitz, Marc, and James A. Levinsohn. 2006. "Productivity in a Differentiated Products Market Equilibrium." Unpublished.

Melitz, Marc, and Gianmarco Ottaviano. 2008. "Market Size, Trade, and Productivity." *Review of Economic Studies* 75 (1): 295–316.

Morrison, Catherine J. 1992. "Unraveling the Productivity Growth Slowdown in the United States, Canada and Japan: The Effects of Subequilibrium, Scale Economies and Markups." *Review of Economics and Statistics* 74 (3): 381–93.

Norrbin, Stefan C. 1993. "The Relation between Price and Marginal Cost in U.S. Industry: A Contradiction." *Journal of Political Economy* 101 (6): 1149–64.

Olley, Steve G., and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica* 64 (6): 1263–97.

Petrin, Amil, and Jagadeesh Sivadasan. 2010. "Estimating Lost Output from Allocative Inefficiency, with Application to Chile and Firing Costs." Unpublished.

Roeger, Werner. 1995. "Can Imperfect Competition Explain the Difference between Primal and Dual Productivity Measures? Estimates for U.S. Manufacturing." *Journal of Political Economy* 103 (2): 316–30.

Slovenian Agency for Public Legal Records and Related Services (AJPES). 1994–2000. Firm-level Accounting Data.

Van Biesebroeck, Johannes. 2005. "Exporting Raises Productivity in Sub-Saharan African Manufacturing Firms." *Journal of International Economics* 67 (2): 373–91.

Waldmann, Robert J. 1991. "Implausible Results or Implausible Data? Anomalies in the Construction of Value-Added Data and Implications for Estimates of Price-Cost Markups." *Journal of Political Economy* 99 (6): 1315–28.

Wooldridge, Jeffrey M. 2009. "On Estimating Firm-Level Production Functions using Proxy Variables to Control for Unobservables." *Economics Letters* 104 (3): 112–14.