Novel methodologies for markup estimation with multiproduct firms

Frederic Warzynski[1]

[1]Aarhus University

2023 Firm and Industry Dynamics Workshop
HKU, June 26-27, 2023

Motivation

- marginal cost key variable in many micro economics models
- IO economists typically believe we can't observe MC and therefore have to estimate it (with very rare exceptions)
- various techniques designed in EIO using either demand, production or cost function estimation
- extension to multi product firms not obvious due to allocation of inputs: De Loecker et al. (2016)
- today: provide two additional approaches to estimate marginal costs with multiproduct firms using duality theory

New methodologies from 2 papers

- describe methodologies from 2 papers (focus on first one)
  - Deregulation and Investment Spillovers in Multi-Product Production Settings (Dhyne, Petrin and Warzynski, 2021)
  - Theory for Extending Single-Product Production Function Estimation to Multi-Product Settings (Dhyne et al., 2022)

Paper #1: Deregulation and Investment Spillovers in Multi-Product Production Settings: Motivation/Big picture

- huge interest in empirical IO for the measurement of productivity and current debate on the proper method to use (OP, LP, ACF, Wooldridge, GNR,...)
- recent literature not very clear about how to deal with multiproduct firms
- old literature in production theory discussing this issue (Diewert, Lau,...)
- our approach: combine the old and the new, provide a modern and relatively simple methodology to estimate productivity and markups at the firm-product level adapting the Diewert-Lau theoretical framework
- economic question: analysis of deregulation in the bread and cake industry in Belgium

Contribution

- provide a novel methodology that combines a joint estimation of a demand function, a production function and a cost function
- apply this methodology using a rich Belgian production survey containing information about value and quantities (hence unit value) for a long time period (1997-2019) at the quarterly level
- show the method yields sensible results in line with theory
- find evidence that deregulation was associated with large welfare gains

Related literature

- original theoretical work by Diewert, Lau, Mundlak,... in the 70s
- empirical approach using US Census data: Dunne and Roberts (1992), Roberts and Supina (2000), Foster, Haltiwanger and Syverson (2008)
- More recently: De Loecker et al. (2016), Valmari (2016), Orr (forthcoming)

Data (I)

- Belgian PRODCOM survey (been used also by Bernard et al., 2019 CAT paper and Amiti et al., 2019, 2022 pass through papers)
- production survey of all manufacturing firms with at least 10 employees run by the Belgian Statistical Office (designed to cover at least 90% of production value in each NACE 4-digit industry)
- provides quarterly 8-digit firm-product observations on values and quantities sold from 1997–2007 (note: extended to 2019 in new version)
- identify all firms producing either bread or cake (categories 15.81.11.00 and 15.81.12.00)
- in this specific market, most firms are producing both goods and nothing else; or are single product and produce only bread or cake

Examples of products

- 15.81.11.00 Fresh bread containing by weight in the dry matter state $\leq$ 5% of sugars and $\leq$ 5% of fat (excluding with added honey; eggs; cheese or fruit)
- 15.81.12.00 Cake and pastry products; other baker's wares with added sweetening matter
- 15.96.10.00 Beer made from malt (excluding non-alcoholic beer, beer containing $\leq$ 0.5% by volume of alcohol, alcohol duty)
- 15.84.22.45 Chocolates (excluding those containing alcohol, in blocks; slabs or bars)

Data (II)

- match this firm-product dataset with standard accounting dataset provided by the VAT statistics, the Social Security records and Central Balance Sheet Office:
  - from VAT statistics: quarterly sales revenues, input purchases and investment in capital (purchases of durable goods)
  - from the National Social Security declarations: quarterly level of employment and total wage bill
  - from Central Balance Sheet Office: total fixed assets as starting capital stock (then use PIM to build a measure of capital)

Methodology

- 3 steps
  - demand function estimation (recover demand shock)
  - production function estimation (recover supply shock, TFPQ)
  - cost estimation and marginal cost computation

Demand estimation

- first step of the analysis: estimate a demand function for differentiated products and with heterogeneity in consumer tastes following the seminal work by Berry (1994) and followers
- use the simplest example of the logit form for illustration:

$$\ln(s_j) - \ln(s_0) = \delta_j = \beta_0 + \alpha p_j + \xi_j$$

- deal with price endogeneity by using classical Hausman-style IVs and input prices
- also run a nested logit specification
- only care here about recovering the demand shock ξ_j

Production function estimation (I)

- builds on Dhyne et al. (2022)
- demand shocks might contaminate the physical productivity measure when production functions are estimated using deflated revenue as measure of production (see e.g. the discussion in Klette and Griliches, 1995)
- estimate the production functions using physical quantity instead, as reported in our surveys for single and multi-product firms separately to recover estimates of firm-product level productivity
- model the joint production of products building on Diewert (1973), who shows that under mild regularity conditions there will exist a multi-product transformation function that relates the output of any good j to all the other goods a firm produces and to aggregate input use

Production function estimation (II)

- for a single product firms: recovering the firm-level productivity shock is standard.
- write production as:

$$q_{it} = \beta_0 + \beta l_{it} + \beta_k k_{it} + \beta_m m_{it} + \omega_{it} + \eta_{it}$$

  where in logs physical quantity is q_{it}, labor is l_{it}, k_{it} is capital, m_{it} is materials, the productivity term ω_{it} is assumed to be first-order Markov, and η_{it} is an i.i.d. shock to production. $\beta = (\beta_l, \beta_k, \beta_m)$ are the elasticities of output of good j with respect to the inputs
- use the Wooldridge (2009) versions of Levinsohn and Petrin (2003) and Olley and Pakes (1996) estimators to deal with the endogeneity of input choice

Production function estimation (III)

- for a multi product firms, write production as:

$$ln\, q_{ipt} = \beta_0 + \beta l_{it} + \beta_k k_{it} + \beta_m m_{it} + \gamma_C \;\, ln\, q_{i(-p)t} + \omega_{ipt} + \eta_{ipt}$$

  where q_{ipt} and $q_{i(-p)t}$ denote the output quantities of the specific good we consider and all the other goods produced by the firm (as a vector or an aggregate) respectively
- the production parameters $\beta = (\beta_l, \beta_k, \beta_m)$ now have the interpretation as the percentage change in output of j due to a percent change in each of the total input levels respectively while holding the production of the other goods

Production function estimation (IV)

- γ_C is the change in output of j that results from increasing the output of the other goods by one percent holding overall input use constant
- the function is only well-defined when $\beta > 0$ and $\gamma_c < 0$
- proxy methods for the estimation of production function parameters straightforward to adapt to the transformation function setting.; use the same methods as in the single product case, but with one additional variable to instrument for, $q_{i(-p)t}$ ans use similar Wooldridge GMM setting with control function approach
- in both the single product and multi-product cases: retrieve the productivity shock ω; important to note that these estimates are not contaminated by demand shocks and are measuring the physical efficiency at the firm-product level

Cost function estimation (I)

- main contribution of this paper
- to measure marginal cost, one approach is to estimate directly a variable cost function including physical quantity produced and input prices, since marginal cost is by definition a function of the first derivative of cost with respect to quantity:

$$MC_{ijt} = \frac{C_{ijt}}{q_{ijt}} \frac{\partial lnC_{ijt}}{\partial lnq_{ijt}}$$

Cost function estimation (II)

- use a variable cost function that allows for fixed input adjustment costs for single-product production functions from Lau (1976) and Berndt and Morrison (1981) by including the level of investment in capital in addition to output quantities and input prices
- As well known from duality theory, productivity is also a component of the error term in the cost function (dual Solow residual); so we face a similar endogeneity problem as in the production function
- two solutions to solve this problem:
  - find an instrumental variable, as proposed by Roberts and Supina (2000)
  - our approach: use the technical efficiency term recovered from the production function estimation and include it directly in the cost estimation, therefore eliminating the unobserved productivity shock problem

Cost function estimation (III)

- for a single product firm:

$$lnVC_{it} = \rho_0 + \rho_{P_m} P_{mt} + \rho_{P_l} P_{lt} + \rho_k k_{it} + \rho_{\Delta K} \Delta k_{it} + \rho_q q_{it} + \rho_\omega \omega_{it} + \epsilon_{it}$$

- testable implications of the theory are that the input price and quantity coefficients should all be positive and the coefficient on capital and productivity should be negative, while quality should have a positive coefficient

Cost function estimation (IV)

- for a multi product firm:

$$lnVC_{ipt} = \rho_0 + \rho_{P_m} P_{mt} + \rho_{P_l} P_{lt} + \rho_k k_{it} + \rho_{\Delta K} \Delta k_{it} + \rho_p q_{ipt}$$

$$+\rho_{-p} q_{i(-p)t} + \rho_{\omega_p} \omega_{ipt} + \rho_{\omega_{(-p)}} \omega_{i(-p)t} + \epsilon_{ipt}$$

- once we get our coefficients, it is straightforward to compute a measure of marginal costs by using the definition shown above.

Results

- Show the results in a sequential way:
  - demand function
  - production function
  - cost function
  - implications for markups and deregulation

Demand function estimation

A. Bread

Variables	Logit OLS	Logit IV[1]	Nested Logit IV[1]
p	-0.796***	-3.253***	-2.229***
	(0.022)	(0.186)	(0.183)
$\ln\left(\frac{s}{1-s_0}\right)$	-	-	0.580***
			(0.080)
Avg. price elasticity	-1.181	-4.829	-7.802
Med. price elasticity	-1.054	-4.308	-6.993
N	14,586	13,892	13,892
F-stat first stage			
p	-	14.97***	14.97***
$\ln\left(\frac{s}{1-s_0}\right)$	-	-	35.08***

B. Cake

Variables	Logit OLS	Logit IV[1]	Nested Logit IV[1]
p	-0.138***	-0.675***	-0.662***
	(0.003)	(0.051)	(0.061)
$\ln\left(\frac{s}{1-s_0}\right)$	-	-	0.284*
			(0.158)
Avg. price elasticity	-0.722	-3.532	-4.833
Med. price elasticity	-0.533	-2.610	-3.539
N	14,558	13,542	13,497
F-stat first stage			
p	-	12.69***	7.57***
$\ln\left(\frac{s}{1-s_0}\right)$	-	-	36.41***

[1] Instrument set includes average wage and prices for 6 inputs.
Standard errors in parentheses.
All specification include NUTS3 dummies and quarter dummies.

Interpretation

- negative coefficient for price, OLS biased
- IV estimates appear to properly correct for the bias related to endogeneity
- nested logit model: estimates of α and σ are in a reasonable range and translate into an average price elasticity between -5 and -8
- from these estimations, we generate the distribution of unobserved quality (ξ_{jt}) that we will plug into the cost function

Production function estimation

Table 4 - Production functions - Estimation results

A. Two products firms

Variables	Bread WOPLP	Bread WOP	Bread WLP	Cake WOPLP	Cake WOP	Cake WLP
ln l	**0.106*****	0.105***	0.124***	**0.503*****	0.487***	0.501***
	(0.024)	(0.024)	(0.022)	**(0.034)**	(0.034)	(0.031)
ln k	**0.029**	0.032	0.049	**0.265***	0.224	0.268**
	(0.102)	(0.099)	(0.091)	**(0.147)**	(0.143)	(0.129)
ln m	**0.860*****	0.869***	0.860***	**0.320*****	0.459***	0.298***
	(0.085)	(0.018)	(0.074)	**(0.126)**	(0.032)	(0.109)
ln q_{other}	**-0.048*****	-0.047***	-0.046***	**-0.108*****	-0.111***	-0.102***
	(0.010)	(0.010)	(0.009)	**(0.022)**	(0.022)	(0.020)
N	**5,353**	5,586	6,522	**5,355**	5,588	6,526

B. Single product firms

Variables	Bread WOP	Bread WLP	Cake WOP	Cake WLP
ln l	**0.108*****	0.085*	**0.517*****	0.415***
	(0.054)	(0.052)	**(0.061)**	(0.055)
ln k	**0.329**	0.109	**0.123**	0.135
	(0.269)	(0.247)	**(0.305)**	(0.270)
ln m	**0.714*****	0.871***	**0.785*****	1.147***
	(0.032)	(0.053)	**(0.056)**	(0.175)
N	**588**	663	**521**	590

All specifications are estimated using the Wooldridge estimation procedure.
Preferred specifications in bold.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Interpretation

- most coefficients are in line with previous studies: for bread, coefficient of material is relatively high, coefficient of labor relatively low, but the coefficient of capital is similar to what is usually found in the literature
- for cake, labor appears to be contributing more to output; more skills required?
- as expected, the output of the other good (cake or bread) reduces the quantity produced of bread once controlling for input
- as result of the estimation, get time varying firm-product level estimates of productivity that we plug in the cost function estimation

Cost function estimation

Table 5 - Cost functions - Estimation results controlling for firm x product specific technological efficiency and product quality

	Two products firms	Single product firms Bread only	Cake only
$\ln \frac{K}{L}$	-0.122***	-0.136***	-0.046***
	(0.007)	(0.016)	(0.017)
$(i > 0)\, i$	5.023***	2.466***	2.554
	(0.383)	(0.463)	(1.634)
$\ln w$	0.319***	0.419***	0.749***
	(0.014)	(0.033)	(0.057)
$\ln q_{Bread}$	0.483***	0.743***	-
	(0.006)	(0.015)	
$\ln q_{Cake}$	0.202***	-	0.623***
	(0.005)		(0.020)
tfp_{Bread}	-0.843***	-0.449***	-
	(0.049)	(0.058)	
tfp_{Cake}	-0.765***	-	-0.309**
	(0.054)		(0.151)
ξ_{Bread}	0.095***	0.164***	-
	(0.003)	(0.006)	
ξ_{Cake}	0.059***	-	0.097***
	(0.002)		(0.007)
N	5,611	594	493

All specifications include time dummies.
The product specific TFP are computed using the preferred specifications of Table 4. Standard errors in parentheses

Interpretation

- input prices are positively and significantly related to variable cost
- output of both products is also positively correlated with cost
- asymmetry between single product and multi product (coefficients of output much lower for multi product firms - sign of economies of scope)
- tfp and demand shock going in expected direction as well
- negative coefficient of the capital per employee ratio, but a positive relationship with positive investment: evidence of investment being positively correlated with input quality

Marginal cost

year	p_{Bread}	mc_{Bread}	μ_{Bread}	ξ_{Bread}	p_{Cake}	mc_{Cake}	μ_{Cake}	ξ_{Cake}
1999	1.356	1.082	1.345	1.771	4.115	2.517	2.276	0.656
2000	1.317	1.042	1.376	1.828	4.148	2.540	2.113	0.846
2001	1.287	1.021	1.415	1.796	3.896	2.631	2.053	0.778
2002	1.235	0.998	1.403	1.822	3.845	2.844	1.979	0.886
2003	1.202	0.963	1.786	1.891	3.837	2.897	1.982	0.967
2004	1.189	0.943	2.980	1.992	3.733	2.917	1.870	1.098
2005	1.157	0.947	1.482	1.959	3.599	2.986	1.828	1.475
2006	1.140	0.890	1.488	2.099	3.480	2.919	1.725	1.500
2007	1.152	0.962	1.291	2.130	3.510	2.786	2.029	1.634

Basic correlations

	p_{Bread}	mc_{Bread}	μ_{Bread}	ξ_{Bread}	p_{Cake}	mc_{Cake}	μ_{Cake}	ξ_{Cake}
p_{Bread}	1.000							
mc_{Bread}	0.945	1.000						
μ_{Bread}	-0.017	-0.096	1.000					
ξ_{Bread}	0.944	0.863	0.025	1.000				
p_{Cake}	0.215	0.209	-0.033	0.148	1.000			
mc_{Cake}	0.193	0.171	0.026	0.147	0.730	1.000		
μ_{Cake}	-0.015	0.036	-0.068	-0.044	0.100	-0.214	1.000	
ξ_{Cake}	0.185	0.205	-0.052	0.207	0.888	0.607	0.126	1.000

Marginal costs (mc) and mark-ups (μ) are computed using the cost function in Table 5.
The ξ are computed using our preferred specification of the demand equation in Table 3.

Conclusion

- provide a simple methodology to estimate marginal cost, quality and productivity at the firm-product level when firms are multi-product
- apply our methodology on a rich firm-product level dataset covering a large subsample of Belgian manufacturing
- analyze the effect of price deregulation in the bread and cake industry in the Summer of 2004 on firms' product quality and efficiency
- find that both quality and efficiency increased substantially after price deregulation took place, generating considerable gains in consumer and producer surpluses

Paper #2: Identification of marginal costs

Shephard (1970), McFadden (1978)

Minimization of the variable cost function given the desired output vector of $q^* = (q_1^*, q_2^*, \ldots, q_M^*)$ is given by

$$Min_x \;\; P * x \;\; s.t. \;\; f_j(q_{-j}^*, x, K, \omega) - q_j^* \geq 0$$

where $P = [P_1 \cdots P_{N_1}]'$ denotes the input prices for the variable inputs. The Lagrangian is

$$L = \;\; P * x - \lambda_j (f_j(q_{-j}^*, x, K, \omega) - q_j^*),$$

which yields the first-order conditions of which optimal input choice x^* is the solution:

$$P_i = \lambda_j \frac{\partial f_j(q_{-j}^*, x^*, K, \omega)}{\partial x_i} \quad i = 1, \ldots, N_1$$

Identification of marginal costs (2)

With the marginal cost is given by λ_j and β_i^j denotes the elasticity of the output of good j with respect to input i:

$$\frac{\partial L}{\partial q_j} = \lambda_j = \frac{P_i}{\frac{\partial f_j(q^*_{-j}, x^*, K, \omega)}{\partial x_i}} = \frac{P_i x_i^*}{\beta_i^j * q_j^*} \quad i = 1, \ldots, N_1.$$

and

$$\lambda_l = -\lambda_j * \frac{\partial f_j(q^*_{-j}, x^*, K, \omega)}{\partial q_l} \quad l \neq j$$

or

$$\frac{\partial L}{\partial q_l} = \lambda_l = -\lambda_j * \gamma_l \frac{q_l^*}{q_j^*}, \quad l \neq j$$

for the Cobb-Douglas log-linear approximation. Thus we have N_1 estimates for the marginal cost for each output good, one for each freely variable input.

Hall (1986, 1988), De Loecker and Warzynski (2012)

Value of observing p and q

In the single-product case it simplifies down to

$$\lambda = \frac{P_i x_i^*}{\beta_i * q^*},$$

Letting p_{q^*} denote the price of output and multiplying this formula through by $\frac{1}{p_{q^*}}$ and inverting we have the markup given as

$$\mu = \frac{p_{q^*}}{\lambda} = \frac{\beta_i}{\frac{P_i x_i^*}{p_{q^*} q^*}}$$

Marginal cost not separately identified.

Table 1: Summary statistics on marginal costs and markups using Cobb Douglas estimates and elasticities. Bread and cake producers, Belgium

	Marginal cost		**Price**		**Markup (P/MC)**	
	Bread	Cake	Bread	Cake	Bread	Cake
mean	1.616	3.004	1.474	5.032	1.099	2.128
10%	0.755	1.064	1.103	3.15	0.518	0.808
25%	0.968	1.459	1.149	3.431	0.731	1.255
50%	1.377	2.295	1.325	3.802	1.051	1.976
75%	1.985	3.791	1.715	5.127	1.411	2.825
90%	2.817	6.156	1.971	8.894	1.746	3.653
std dev	0.90	2.153	0.422	3.109	0.471	1.104

Table 2: Summary statistics on marginal costs and markups using translog estimates and elasticities. Bread and cake producers, Belgium

	Marginal cost		**Markup**	
	Bread	Cake	Bread	Cake
mean	1.362	2.665	1.452	2.543
10%	0.570	0.970	0.580	0.837
25%	0.743	1.312	0.912	1.351
50%	1.087	2.014	1.344	2.281
75%	1.644	3.424	1.920	3.422
90%	2.568	5.412	2.437	4.566
std dev	0.894	1.899	0.713	1.526

Table 3: Comparing marginal costs between single and multi product firms, OLS and FE results. 6-digit Prodcom level, Belgian data

		OLS		FE		# obs.
		MP	*logq*	*MP*	*logq*	
107111	Bread	-0.535***	-0.065***	-0.480***	-1.237***	8,973
		(0.020	(0.004))	(0.008)	(0.021)	
107112	Cake	-0.778***	-0.191***	-0.540***	-1.214***	8,684
		(0.024)	(0.006)	(0.008)	(0.032)	
221314	Doors of plastic	-0.058	-0.537***	-0.898***	-0.306***	2,062
		(0.036)	(0.010)	(0.010)	(0.034)	
251210	Doors of metal	-0.743***	-0.601***	-0.775***	-0.941***	3,355
		(0.029)	(0.009)	(0.010)	(0.062)	
108222	Chocolate	0.214***	-0.178***	-0.465***	0.095***	2,515
		(0.021)	(0.007)	(0.012)	(0.029)	
108223	Sugar confectionery	-0.233***	-0.189***	-0.440***	-1.431***	1,664
		(0.051)	(0.013)	(0.015)	(0.068)	
251123	Structures	0.383***	-0.539***	-0.716***	0.015	8,582
		(0.028)	(0.005)	(0.007)	(0.044)	
236111	Bricks	-0.937***	-0.263***	-0.487***	-1.380***	1,994
		(0.023)	(0.012)	(0.014)	(0.255)	
236112	Prefabricated	-0.860***	-0.335***	-0.569***	-1.321***	3,915
		(0.020)	(0.008)	(0.013)	(0.052)	

Table 4: Comparing marginal costs between single and multi product firms, FE and IV-FE results. 6-digit Prodcom level, Belgian data

Dependent variable: *logMC*		FE		IV-FE			# obs.
		MP	*logq*	*MP*	*logq*	F-stat	
107111	Bread	-1.237***	-0.480***	-1.288***	-0.091***	427.07	8,973
		(0.021)	(0.008)	(0.024)	(0.029)		
107112	Cake	-1.214***	-0.540***	-1.171***	-0.407***	349.84	8,684
		(0.032)	(0.008)	(0.034)	(0.028)		
221314	Doors of plastic	-0.306***	-0.898***	-0.359***	-1.059***	339.87	2,062
		(0.034)	(0.010)	(0.037)	(0.021)		
251210	Doors of metal	-0.941***	-0.775***	-1.095***	-0.922***	610.6	3,355
		(0.062)	(0.010)	(0.066)	(0.019)		
108222	Chocolate	0.095***	-0.465***	0.034	-0.407***	137.54	2,515
		(0.029)	(0.012)	(0.028)	(0.036)		
108223	Sugar confectionery	-1.431***	-0.440***	-1.304***	-0.194**	23.23	1,664
		(0.068)	(0.015)	(0.088)	(0.095)		
251123	Structures	0.015	-0.716***	0.083*	-0.553***	416.43	8,582
		(0.044)	(0.007)	(0.046)	(0.023)		
236111	Bricks	-1.380***	-0.487***	-1.374***	-0.461***	1143.52	1,994
		(0.255)	(0.014)	(0.256)	(0.019)		
236112	Prefabricated	-1.321***	-0.569***	-1.308***	-0.400***	1296.21	3,915
		(0.052)	(0.013)	(0.053)	(0.021)		

Note: This table provides the results of a regression of the log of marginal costs (*logMC*) over the log of firm size (*loq*) and a multi product dummy (*MP*). We show results for the fixed effect specification and the IV-FE specification using Hausmanian instruments (average size of the other firms in the industry and average share of multi product firms for the other firms in the industry). Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Conclusion

- introduced 2 novel methodologies to estimate markups for multi product firms
- using duality theory: both methods work to recover the dual cost function from the production side
- delivers reasonable results
- currently working on extending the approach to firms producing more than 2 products